ACTA ARITHMETICA
XLIII (1984)

Some upper bounds in the theory of irregularities of distribution

by

JÓZSEF BECK (Budapest)

1. Introduction. Let $\mathscr{U}^r$ be the r-dimensional unit cube $0 \leqslant x_1 < 1$, $0 \leqslant x_2 < 1, \ldots, 0 \leqslant x_r < 1$. The elements of the σ-algebra generated by the open sets in $\mathscr{U}^r$ are called the Borel sets in $\mathscr{U}^r$. Let μ be a nonnegative normed Borel measure in $\mathscr{U}^r$, that is, a nonnegative measure μ defined on the class of Borel sets with $\mu(\mathscr{U}^r) = 1$. A set $B \subseteq \mathscr{U}^r$ will be called a *box* if it is a Cartesian product $I_1 \times I_2 \times \ldots \times I_r$ of intervals $I_1, I_2, \ldots, I_r$. By definition the sides of B are parallel to the coordinate axes.

Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ be an infinite sequence of points in $\mathscr{U}^r$. Given a box B, write $Z(n, B) = Z(n, B; \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots)$ for the number of $i, 1 \leqslant i \leqslant n$ for which $\boldsymbol{x}_i \in B$. The sequence $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ is called *μ-uniformly distributed* if for every box B we have the asymptotic relation $Z(n, B)/n \to \mu(B)$.

Now we introduce a quantity that measures the deviation of the distribution of $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ from the measure μ.

Set

$$D(\mu, n, B) = |Z(n, B) - n \cdot \mu(B)|,$$

and

$$\Delta(\mu, n) = \sup_B D(\mu, n, B),$$

where the supremum is taken over all the boxes in $\mathscr{U}^r$. Here $\Delta(\mu, n) = \Delta(\mu, n; \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots)$ is called the *μ-discrepancy* function of the sequence $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$

We are interested in sequences which are very well μ-uniformly distributed, i.e., which have $\Delta(\mu, n) \ll f(n)$ where $f(n)/n$ tends to zero very rapidly.

If μ coincides with the r-dimensional normed Lebesgue measure λ_r, the problem above is classical. Let $p_1, p_2, \ldots, p_r$ be the first r primes. Write n in the scale of p_i,

$$n = \sum_{j=0}^{s_i} a_{ij}(p_i)^j, \quad \text{where} \quad 0 \leqslant a_{ij} < p_i.$$

Then we set

$$x_{ni} = \sum_{j=0}^{s_i} a_{ij}(p_i)^{-j-1} \quad \text{and} \quad x_n = (x_{n1}, x_{n2}, \dots, x_{nr}).$$

The sequence $x_1, x_2, \dots$ so constructed is called *van der Corput–Hammersley–Halton sequence* (C–H–H sequence). It is known (J. Halton [4], see also W. M. Schmidt [8])

THEOREM A. *The C–H–H sequence has*

$$\Delta(\lambda_r, n) \ll (\log n)^r$$

where the implicit constant depends only on r.

It is a well-known conjecture that *no* infinite sequence can have a λ_r-discrepancy function of smaller order of magnitude than the C–H–H sequence. For $r = 1$ this conjecture have been proved by W. M. Schmidt [7].

Our first aim is to prove a general result.

THEOREM 1. *Given any nonnegative normed Borel measure μ in $\mathscr{U}^r$, one can find an infinite sequence $x_1, x_2, \dots$ in $\mathscr{U}^r$ having*

$$\Delta(\mu, n) \ll (\log n)^{2r+2}$$

where the implicit constant is independent of n and μ.

The proof will be based on the following "integer-making" lemma (see Beck–Fiala [3]).

LEMMA 1. *Let the real numbers $a_1, \dots, a_s$ be given and a family $\mathscr{F}$ of subsets of the index set $\{1, 2, \dots, s\}$. Assume that each $i \in \{1, 2, \dots, s\}$ belongs to at most t elements of $\mathscr{F}$. Then there exist integers $a_1, \dots, a_s$ so that $|a_i - a_i| < 1$ and*

$$\Big|\sum_{i \in E} a_i - \sum_{i \in E} a_i\Big| \leqslant t-1 \quad \text{for all } E \in \mathscr{F}.$$

Secondly we shall investigate the μ-discrepancy of finite sequences with respect to tilted boxes and balls. Let $x_1, \dots, x_N$ be N points in $\mathscr{U}^r$. Given a measurable subset A of Euclidean r-space $\boldsymbol{R}^r$, write $Z(A) = Z(A; x_1, \dots, x_N)$ for the number of points $x_1, \dots, x_N$ in A. Set

$$D(\mu, A) = D(\mu, A; x_1, \dots, x_N) = |Z(A) - N \cdot \mu(A \cap \mathscr{U}^r)|.$$

The following beautiful results are due to W. M. Schmidt [6].

THEOREM B. *Let N points in $\mathscr{U}^r$ be given and let $0 < \varepsilon, \delta < 1$.*

(i) *Let $r = 2$ and $N\delta^2 > \varepsilon$. Then there exists a tilted rectangle R with diameter $\leqslant \delta$ and with $D(\lambda_2, R) \gg (N\delta^2)^{1/4-\varepsilon}$.*

(ii) *Suppose $N\delta^r > \varepsilon$. Then there exists a ball B with diameter $\leqslant \delta$ and with $D(\lambda_r, B) \gg (N\delta^r)^{\frac{1}{2}-\frac{1}{2r}-\varepsilon}$. The implicit constants depend only on r and ε.*

Note that the tilted rectangle and the ball in Theorem B are not necessarily contained in the unit cube, but should be interpreted as subsets of the torus.

We say that T is a *tilted box* if it can be obtained from a box parallel to the coordinate axes by an orthogonal transformation. The theorem below shows that the exponents in Theorem B are the best possible.

THEOREM 2. *Let μ be a nonnegative normed Borel measure in $\mathscr{U}^r$ such that μ is totally continuous with respect to λ_r (i.e. $\mu(A) \neq 0$ implies $\lambda_r(A) \neq 0$) and the Radon–Nikodym derivative $\frac{d\mu}{d\lambda_r}$ satisfies the inequality $\frac{1}{M} \leqslant \frac{d\mu}{d\lambda_r}(x) \leqslant M$ for almost all $x \in \mathscr{U}^r$.*

Then

(i) *there exists an N-element set in $\mathscr{U}^r$ such that, for every tilted box $T \subset \boldsymbol{R}^r$,*

$$D(\mu, T) \ll (N\delta^r)^{\frac{1}{2}-\frac{1}{2r}}(\log N)^{1/2},$$

where δ denotes the diameter of T;

(ii) *there exists an N-element set in $\mathscr{U}^r$ such that, for every ball $B \subset \boldsymbol{R}^r$,*

$$D(\mu, B) \ll (N\delta^r)^{\frac{1}{2}-\frac{1}{2r}}(\log N)^{1/2},$$

where δ denotes the diameter of B. The implicit constants in $\ll$ depend only on r and M.

A particular case of Theorem 2 was treated in [2]. Though the proof of Theorem 2 goes on the same line as that of this particular case, for the sake of completeness we include here the simple proof. Almost certainly the theorem remains true without any additional condition on μ, but in the general case our method does not work. In the proof we shall apply probabilistic arguments, namely the so-called Bernstein–Chernoff inequality of large deviation type.

Finally we mention a result on irregularities of integer sequences with respect to arithmetic progressions. Our starting points is the following remarkable theorem of K. F. Roth [5].

First some notation. Let

$$\mathrm{AP}(h, q, m) = \{1 \leqslant a \leqslant m\colon\ a \equiv h \pmod q\},$$

that is, $\mathrm{AP}(h, q, m)$ denotes the intersection of the congruence class $h \pmod q$ with the interval $[1, m]$. Denote by $|S|$ the cardinality of the set S. Given $S \subseteq [1, N]$ and positive integers h, q and m, let $D(h, q, m;\ S)$ denote the "discrepancy" of S with respect to $\mathrm{AP}(h, q, m)$, i.e.,

$$D(h, q, m;\ S) = \left| |S \cap \mathrm{AP}(h, q, m)| - \frac{|S|}{N} |\mathrm{AP}(h, q, m)| \right|.$$

THEOREM C (K. F. Roth). *Let N be any natural number, S a subset of $[1, N]$. Set*

$$V(q, m;\ S) = \sum_{h=1}^{q} D^2(h, q, m;\ S).$$

Then for any positive integer Q,

$$\sum_{q=1}^{Q} q^{-1} \sum_{m=1}^{N} V(q, m;\ S) + Q \sum_{q=1}^{Q} V(q, N;\ S) \geqslant \frac{|S|}{N}\left(1 - \frac{|S|}{N}\right) Q^2 N,$$

where the implicit constant is absolute.

Perhaps the most interesting consequence of Theorem C is as follows: Coloring the integers from 1 to N red and blue in any fashion, there always exists an arithmetic progression such that the difference of the numbers of red and blue terms in it has absolute value $\geqslant N^{1/4}$. Here the exponent $1/4$ of N is best possible, see [1].

Let $S \subseteq [1, N]$ and introduce

$$\Delta(S) = \max D(h, q, m;\ S)$$

where the maximum is taken over all positive triplets h, q, m for which $\mathrm{AP}(h, q, m) \subseteq [1, N]$. Further interesting consequence of Theorem C can be obtained by choosing $\min\{|S|, N - |S|\}/N^{1/2} \to \infty$. In this case Theorem C yields $\Delta(S) \to \infty$. Now we shall present a result in the other direction.

THEOREM 3. *Given $\varepsilon > 0$ and a natural number N, there exists $S \subset [1, N]$ such that*

$$\min\{|S|, N - |S|\} > (\log N)^{1-\varepsilon} \quad \text{and} \quad \Delta(S) \ll 1,$$

where the implicit constant depends only on ε.

Similarly as in the proof of Theorem 2 here we shall also use probabilistic arguments. There is a huge gap between the upper and lower bounds. It would be worth improving both of them.

2. Proof of Theorem 1. Throughout this section the implicit constants in $\ll$ depend only on the dimension. In the first step we shall reduce Theorem 1 to the following assertion concerning finite sequences.

PROPOSITION 1. *Given any positive integer N and an arbitrary non-negative normed Borel measure ν in $\mathscr{U}^d$, there exists an N-element set in $\mathscr{U}^d$ such that, for every box $B \subseteq \mathscr{U}^d$, $D(\nu, B) \ll (\log N)^{2d}$.*

The reduction goes as follows. Let $d = r + 1$ and $\nu = \mu \times \lambda$, where $\lambda = \lambda_1$ denotes the one-dimensional normed Lebesgue measure in $[0, 1]$. Furthermore, let $N = N_i = 2^{2^{i+1}} - 2^{2^i}$. Proposition 1 yields the existence of an N_i-element set $S_i \subset \mathscr{U}^d$ such that, for every box $B \subseteq \mathscr{U}^d$,

$$D(\nu, B) = D(\nu, B;\ S_i) \ll (\log N_i)^{2d} \ll 2^{2id}. \tag{1}$$

Rearrange the elements of S_i in increasing order by their last coordinates,

$$S_i = \{\boldsymbol{y}_1^{(i)}, \boldsymbol{y}_2^{(i)}, \ldots, \boldsymbol{y}_{N_i}^{(i)}\},$$

where

$$\boldsymbol{y}_j^{(i)} = (y_{j1}^{(i)}, y_{j2}^{(i)}, \ldots, y_{jd}^{(i)}) \quad \text{and} \quad y_{1d}^{(i)} \leqslant y_{2d}^{(i)} \leqslant \ldots \leqslant y_{N_i d}^{(i)}.$$

Simple approximation procedure shows that we can assume $y_{1d}^{(i)} < y_{2d}^{(i)} < \ldots < y_{N_i d}^{(i)}$. Now here is the definition of the desired very well μ-uniformly distributed sequence $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ in $\mathscr{U}^r$: If $n = 2^{2^i} + j$ with $1 \leqslant j \leqslant 2^{2^{i+1}} - 2^{2^i}$, let

$$\boldsymbol{x}_n = (y_{j1}^{(i)}, y_{j2}^{(i)}, \ldots, y_{j(d-1)}^{(i)}) \in \mathscr{U}^{d-1} = \mathscr{U}^r,$$

that is, $\boldsymbol{x}_n$ can be obtained from $\boldsymbol{y}_j^{(i)}$ by omitting its last coordinate. Now let $B \subseteq \mathscr{U}^r$ be any r-dimensional box and let $n = 2^{2^i} + j$ with $1 \leqslant j \leqslant 2^{2^{i+1}} - 2^{2^i} = N_i$. We have

$$\begin{aligned} D(\mu, n, B;\ \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots) &= |Z(n, B;\ \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots) - n \cdot \mu(B)| \\ &= \Big| \sum_{l=0}^{i-1} \{Z(B \times [0, 1);\ S_l) - N_l \cdot \nu(B \times [0, 1))\} + \\ &\qquad + Z(B \times [0, y_{jd}^{(i)}];\ S_i) - (n - 2^{2^i}) \cdot \mu(B) \Big| \\ &\leqslant \sum_{l=0}^{i-1} D(\nu, B \times [0, 1);\ S_l) + |Z(B \times [0, y_{jd}^{(i)}];\ S_i) - N_i \cdot \nu(B \times [0, y_{jd}^{(i)}])| + \\ &\qquad + |N_i \cdot \nu(B \times [0, y_{jd}^{(i)}]) - (n - 2^{2^i}) \mu(B)|. \end{aligned} \tag{2}$$

By definition,

$$|Z(B \times [0, y_{jd}^{(i)}];\ S_i) - N_i \cdot \nu(B \times [0, y_{jd}^{(i)}])| = D(\nu, B \times [0, y_{jd}^{(i)}];\ S_i). \tag{3}$$

On the other hand,

$$\begin{aligned} |N_i \cdot \nu(B \times [0, y_{jd}^{(i)}]) - (n - 2^{2^i}) \mu(B)| &= \mu(B) \cdot |N_i \cdot y_{jd}^{(i)} - j| \leqslant |N_i \cdot y_{jd}^{(i)} - j| \\ &= |N_i \cdot \nu(\mathscr{U}^r \times [0, y_{jd}^{(i)}]) - Z(\mathscr{U}^r \times [0, y_{jd}^{(i)}];\ S_i)| \\ &= D(\nu, \mathscr{U}^r \times [0, y_{jd}^{(i)}];\ S_i). \end{aligned} \tag{4}$$

Thus, by (1), (2), (3) and (4)

$$D(\mu, n, B; x_1, x_2, \ldots)$$
$$\leqslant \sum_{l=0}^{i-1} D(\nu, B\times[0,1);\ S_l)+D(\nu, B\times[0, y_{jd}^{(i)}];\ S_i)+D(\nu, \mathscr{U}^r\times[0, y_{jd}^{(i)}];\ S_i)$$
$$\leqslant \sum_{l=0}^{i-1} 2^{2ld}+2^{2id}+2^{2id} \leqslant 2^{2id} \leqslant (\log n)^{2d} = (\log n)^{2r+2},$$

which completes the deduction of Theorem 1 from Proposition 1.

To verify Proposition 1 we observe that ν can be approximated by discrete measures. Hence Proposition 1 is essentially equivalent with

PROPOSITION 2. *Given any (not necessarily distinct) K points $z_1, \ldots, z_K$ in $\mathscr{U}^d$ and a natural number $N \leqslant \sqrt{K}$ there exists an N-element subset $\{y_1, \ldots, y_N\}$ of $\{z_1, \ldots, z_K\}$ such that for every box $B \subseteq \mathscr{U}^d$,*

$$\left|\sum_{y_i\in B} 1 - \frac{N}{K}\sum_{z_i\in B} 1\right| \leqslant (\log N)^{2d}.$$

For convenience in the proof of Proposition 2 we shall restrict ourselves to the case $d = 2$. Rearrange the points $z_1, \ldots, z_K \in \mathscr{U}^2$ in two ways as follows:

$$\{z_1, \ldots, z_K\} = \{z_1', \ldots, z_K'\} = \{z_1'', \ldots, z_K''\}$$

where

$$z_i' = (z_{i1}', z_{i2}'), \quad z_{11}' \leqslant z_{21}' \leqslant \ldots \leqslant z_{K1}'$$

and

$$z_i'' = (z_{i1}'', z_{i2}''), \quad z_{12}'' \leqslant z_{22}'' \leqslant \ldots \leqslant z_{K2}''.$$

Let $k_i = \left\lfloor i\frac{K}{N}\right\rfloor$ (integral part), $0 \leqslant i \leqslant N$. Set

$$F_i = \{z_j':\ k_{i-1} < j \leqslant k_i\}, \quad G_i = \{z_j'':\ k_{i-1} < j \leqslant k_i\},$$

$$H_{ij} = F_i \cap G_j \quad \text{and} \quad \beta_{ij} = \frac{N}{K+N}\sum_{z_l\in H_{ij}} 1, \quad 1 \leqslant i, j \leqslant N.$$

Clearly, $0 \leqslant \beta_{ij} \leqslant 1$. In order to complete the proof of Proposition 2 we need the following purely combinatorial statement.

PROPOSITION 3. *Let be given an N by N matrix $[a_{ij}]$ with $0 \leqslant a_{ij} \leqslant 1$. Then there exists an N by N 01-matrix $[a_{ij}]$ so that $a_{ij} = 0$ whenever $a_{ij} = 0$ and*

$$\left|\sum_{i=1}^{r}\sum_{j=1}^{s}(a_{ij}-a_{ij})\right| \leqslant (\log N)^4 \quad \text{for all } 1 \leqslant r, s \leqslant N.$$

By Proposition 3 there exists an N by N 01-matrix $[b_{ij}]$ so that

$$\left|\sum_{i=1}^{r}\sum_{j=1}^{s}(b_{ij}-\beta_{ij})\right| \leqslant (\log N)^4 \quad \text{for all } 1 \leqslant r, s \leqslant N.$$

We can assume $b_{ij} = 0$ if $\beta_{ij} = 0$. Now select from each H_{ij} a point y_{ij} arbitrarily whenever $b_{ij} = 1$, and set $Y = \{y_{ij}:\ b_{ij} = 1\}$. Let $B_{r,s} = \bigcup_{i=1}^{r}\bigcup_{j=1}^{s} H_{ij}$. The point-set Y has the property

$$\left|\sum_{z_l\in Y\cap B_{r,s}} 1 - \frac{N}{K}\sum_{z_l\in B_{r,s}} 1\right|$$
$$\leqslant \left|\sum_{z_l\in Y\cap B_{r,s}} 1 - \frac{N}{K+N}\sum_{z_l\in B_{r,s}} 1\right| + \left|\left(\frac{N}{K}-\frac{N}{K+N}\right)\sum_{z_l\in B_{r,s}} 1\right|$$
$$= \left|\sum_{i=1}^{r}\sum_{j=1}^{s}(b_{ij}-\beta_{ij})\right| + O(1) = O\left((\log N)^4\right)+O(1) = O\left((\log N)^4\right)$$

for all $1 \leqslant r, s \leqslant N$ since $N \leqslant \sqrt{K}$. Choosing $r = s = N$ we conclude that the cardinality of Y differs from N by $\leqslant (\log N)^4$.

Let $\triangle$ denote the symmetric difference, i.e., $A \triangle B = (A \setminus B)\cup(B \setminus A)$.

It follows that there exists an N-element subset Y_1 of $\{z_1, \ldots, z_K\}$ such that $|Y_1 \triangle Y| \leqslant (\log N)^4$ and

$$\left|\sum_{z_l\in Y_1\cap B_{r,s}} 1 - \frac{N}{K}\sum_{z_l\in B_{r,s}} 1\right| \leqslant (\log N)^4 \quad \text{for all } 1 \leqslant r, s \leqslant N. \tag{5}$$

We claim that Y_1 is the desired N-element set. For let $B = [0, x]\times[0, y]$. Clearly one can find indices r and s, $0 \leqslant r, s \leqslant N-1$ such that

$$B_{r,s} \subset \{z_l:\ z_l \in B\} \subseteq B_{r+1,s+1}.$$

Since $|B_{r+1,s+1} \setminus B_{r,s}| \leqslant 2K/N$, we conclude that

$$|\{z_l:\ z_l \in B\} \setminus B_{r,s}| \leqslant 2K/N, \tag{6}$$

and by (5),

$$\left|\sum_{z_l\in Y_1\cap B} 1 - \sum_{z_l\in Y_1\cap B_{r,s}} 1\right| \tag{7}$$
$$\leqslant \left|\sum_{z_l\in Y_1\cap B_{r+1,s+1}} 1 - \sum_{z_l\in Y_1\cap B_{r,s}} 1\right| \leqslant \left|\frac{N}{K}\sum_{z_l\in B_{r+1,s+1}} 1 - \frac{N}{K}\sum_{z_l\in B_{r,s}} 1\right| + O\left((\log N)^4\right)$$
$$= O(1)+O\left((\log N)^4\right) = O\left((\log N)^4\right).$$

Therefore, by (5), (6) and (7)

$$\Big|\sum_{x_l\in Y_1\cap B}1-\frac{N}{K}\sum_{x_l\in B}1\Big|\leqslant\Big|\sum_{x_l\in Y_1\cap B_{r,s}}1-\frac{N}{K}\sum_{x_l\in B_{r,s}}1\Big|+$$

$$+\Big|\sum_{x_l\in Y_1\cap B}1-\sum_{x_l\in Y_1\cap B_{r,s}}1\Big|+\Big|\frac{N}{K}\sum_{x_l\in B}1-\frac{N}{K}\sum_{x_l\in B_{r,s}}1\Big|$$

$$=O\big((\log N)^4\big)+O\big((\log N)^4\big)+\frac{N}{K}\sum_{x_l\in B\setminus B_{r,s}}1=O\big((\log N)^4\big),$$

which completes the proof of Proposition 2.

The proof of Proposition 3 will be based on Lemma 1 (see Section **1**). We may assume $N=2^l$. For $0\leqslant p,q\leqslant l$ we partition the matrix $[a_{ij}]$ into 2^{p+q} submatrices, splitting the horizontal side of the matrix into 2^p equal pieces and the vertical side of the matrix into 2^q equal pieces. There are $(l+1)^2\sim(\log N)^2$ such submatrices. Let us call a submatrix *special* if it occurs in one of these partitions. Lemma 1 yields the existence of an N by N 01-matrix $[a_{ij}]$ so that the absolute value of the sum $\sum(a_{ij}-a_{ij})$ in each of the special submatrices is at most $(l+1)^2$. But any submatrix containing the lower left corner is the union of at most l^2 special submatrices. Formally,

$$[1,r]\times[1,s]=\bigcup_{i=1}^{k}\bigcup_{j=1}^{n}\Big[1+\sum_{h\leqslant i-1}2^{u_h},\sum_{h\leqslant i}2^{u_h}\Big]\times\Big[1+\sum_{h\leqslant j-1}2^{v_h},\sum_{h\leqslant j}2^{v_h}\Big],$$

where

$$r=2^{u_1}+2^{u_2}+\ldots+2^{u_k},\quad u_1>u_2>\ldots>u_k\geqslant 0$$

and

$$s=2^{v_1}+2^{v_2}+\ldots+2^{v_n},\quad v_1>v_2>\ldots>v_n\geqslant 0.$$

Proposition 3 follows.

Finally, for the sake of completeness we include the proof of Lemma 1. The construction of the integers $a_1,\ldots,a_s$ will be based on a repeated application of the following almost trivial fact from linear algebra: If a linear system of equations has more variables than equations, then there exists a nontrivial solution. We can assume $0<\alpha_i<1$. We shall define a sequence $\boldsymbol{a}^0,\boldsymbol{a}^1,\ldots,\boldsymbol{a}^p$ of s-dimensional vectors $\boldsymbol{a}^j=(\alpha_1^j,\alpha_2^j,\ldots,\alpha_s^j)$ and a sequence X^j of subsets of $\{1,2,\ldots,s\}$ with the following properties:

(8) $$\alpha_i^0=\alpha_i\quad\text{for}\quad 1\leqslant i\leqslant s;$$

(9) $$0\leqslant\alpha_i^j\leqslant 1\quad\text{for}\quad 1\leqslant j\leqslant p,\ 1\leqslant i\leqslant s;$$

(10) X^j is the set of indices i for which α_i^j is not 0 or 1;

(11) $$X^0\supsetneqq X^1\supsetneqq X^2\supsetneqq\ldots\supsetneqq X^p=\emptyset;$$

(12) $\alpha_i^j=\alpha_i^k$ for $k=j+1,j+2,\ldots,p$ whenever α_i^j is 0 or 1;

(13) $$\sum_{i\in E}\alpha_i^j=\sum_{i\in E}\alpha_i^{j+1}\quad\text{for all } E\in\mathscr{F}\text{ with }|E\cap X^j|>t;$$

(14) if $|E\cap X^j|=t$ and $\sum_{i\in E}\alpha_i^j\neq\sum_{i\in E}\alpha_i^{j+1}$ then

$$E\cap X^{j+1}=\emptyset\quad\text{and}\quad\Big|\sum_{i\in E}\alpha_i^{j+1}-\sum_{i\in E}\alpha_i\Big|\leqslant t/2.$$

According to (10) these sequences terminate, if the final vector $\boldsymbol{a}^p$ has only 0, 1 coordinates. Choosing $a_i=\alpha_i^p$, $1\leqslant i\leqslant s$ it follows from (14), (8) and (13) that for all $E\in\mathscr{F}$

$$\text{either }\Big|\sum_{i\in E}a_i-\sum_{i\in E}\alpha_i\Big|\leqslant t/2\quad\text{or}\quad\Big|\sum_{i\in E}a_i-\sum_{i\in E}\alpha_i\Big|<t-1,$$

and this is a bit more than Lemma 1.

We construct the sequence $\boldsymbol{a}^j$ of vectors by induction. Let $\boldsymbol{a}^0=(\alpha_1,\ldots,\alpha_s)$. Now suppose that $\boldsymbol{a}^j$ is defined and X^j is non-empty. Introduce

$$\mathscr{F}_j=\{E\in\mathscr{F}\colon|E\cap X^j|\geqslant t\}.$$

If $\mathscr{F}_j$ is empty, then set $p=j+1$, $\alpha_i^{j+1}=\alpha_i^j$ for $i\notin X^j$ and $\alpha_i^{j+1}=0$ or 1 arbitrarily for $i\in X^j$. If $\mathscr{F}_j$ is not empty, then by the hypothesis of the lemma there are only two possibilities:

Case 1. $|\mathscr{F}_j|=|X^j|$.

Case 2. $|\mathscr{F}_j|<|X^j|$.

Again by the hypothesis of the lemma in Case 1 every intersection $E\cap X^j$ has exactly t elements whenever $E\cap X^j\neq\emptyset$. Set $p=j+1$, let α_i^{j+1} be the integer closest to α_i for all $i\in X^j$. Hence $|\alpha_i^{j+1}-\alpha_i|\leqslant 1/2$ and

$$\Big|\sum_{i\in E}\alpha_i^{j+1}-\sum_{i\in E}\alpha_i\Big|\leqslant\frac{|E\cap X^j|}{2}\leqslant\frac{t}{2}\quad\text{for each } E\in\mathscr{F}.$$

In the second case let us associate a real variable x_i with each i, $1\leqslant i\leqslant s$ and consider the following linear system of equations:

$$\sum_{i\in E\cap X^j}x_i=0\quad\text{for each } E\in\mathscr{F}_j,\quad\text{and } x_i=0\quad\text{for each } i\notin X^j.$$

A nontrivial solution $\{x_i\}_{i=1}^s$ exists, because there are more variables than equations. Now let y_0 be the greatest positive value for which the

inequalities $0 \leqslant a_i^j + y_0 x_i \leqslant 1$, $i \in X^j$ hold, and set

$$a_i^{j+1} = a_i^j + y_0 x_i, \quad 1 \leqslant i \leqslant s.$$

By the maximality of y_0, $X^{j+1} \subsetneq X^j$. It is easily seen that

$$\sum_{i \in E} a_i^{j+1} = \sum_{i \in E} a_i^j \quad \text{for all } E \in \mathscr{F}_j.$$

One can easily verify that in each case the relations (8)–(14) hold, and this completes the proof of Lemma 1.

3. Proof of Theorem 2. We say that R is an r-dimensional *generalized polytope* if it is representable as the intersection of r-dimensional halfspaces and balls. Theorem 2 easily follows from the following result.

THEOREM 2*. *Under the hypothesis of Theorem 2 there exists an N-element set in $\mathscr{U}^r$ such that for every r-dimensional generalized polytope $R \subseteq \mathscr{U}^r$,*

$$D(\mu, R) \ll (N\delta^r)^{\frac{1}{2}-\frac{1}{2r}} (\log N)^{1/2},$$

where δ denotes the diameter of R and the implicit constant in $\ll$ depend only on r, M and on the number of sides of R.

For the proof of Theorem 2* we note that there is a measurable partition of $\mathscr{U}^r$ into disjoint subsets $Q_1, \ldots, Q_N$, each with $\mu(Q_i) = 1/N$ and with diameter δ_i where $c_1(r, M)N^{-1/r} \leqslant \delta_i \leqslant c_2(r, M)N^{-1/r}$, $1 \leqslant i \leqslant N$. Let us associate with each Q_i a "random point" $\xi_i \in Q_i$ as follows:

$$\text{Prob}(\xi_i \in A) = \frac{\mu(A)}{\mu(Q_i)} = N\mu(A),$$

where A is a measurable subset of Q_i. Furthermore, assume that the random variables $\xi_1, \ldots, \xi_N$ are independent of each other (the existence of $\xi_1, \ldots, \xi_N$ is guaranteed by a basic theorem of Kolmogorov, see any textbook on probability theory).

Let $s(R)$ denote the number of sides (faces) of the generalized polytope R. Our goal is to prove that the random N-element set $\{\xi_1, \ldots, \xi_N\}$ defined above has

$$D(\mu, R;\ \xi_1, \ldots, \xi_N) \leqslant c_3(r, M, s(R))(N\delta^r)^{\frac{1}{2}-\frac{1}{2r}} (\log N)^{1/2}$$

for every generalized polytope R with diameter δ, $0 < \delta < 1$ with probability $\geqslant 1/2$. Clearly this will complete the proof.

Now consider an arbitrary generalized polytope $R \subseteq \mathscr{U}^r$ with diameter δ and with $s(R) \leqslant S$. It is easily seen that the sides of R intersect $\ll (N\delta^r)^{1-1/r}$ Q_i's. Here and in what follows the implicit constant in $\ll$ depends only on r, M and S. Therefore, R is representable as the disjoint union of Q_i's entirely contained by R and the union of $\ll (N\delta^r)^{1-1/r}$ "pieces" which are the intersections of some Q_i's and R, i.e.,

$$R = \bigcup_{i \in I} Q_i \cup \bigcup_{j \in J} (Q_j \cap R),$$

where the index-set J has cardinality $c\lfloor (N\delta^r)^{1-1/r} \rfloor$. Since for every Q_i, $\mu(Q_i) = 1/N$ and Q_i contains exactly one element of $\{\xi_1, \ldots, \xi_N\}$, the μ-discrepancy of $\bigcup_{i \in I} Q_i$ is zero. Therefore, it remains to investigate the μ-discrepancy of $\bigcup_{j \in J} (Q_j \cap R)$.

For notational convenience let

$$T = \bigcup_{j \in J} (Q_j \cap R) \quad \text{and} \quad J = \{1, 2, \ldots, l\}.$$

Let us define the random variables χ_j, $1 \leqslant j \leqslant l$ as follows: Let $\chi_j = 1$ if $\xi_j \in Q_j \cap R$, otherwise let $\chi_j = 0$. By definition

$$D(\mu, R;\ \xi_1, \ldots, \xi_N) = D(\mu, T;\ \xi_1, \ldots, \xi_N) = \Big| \sum_{j=1}^{l} \chi_j - N\Big(\sum_{j=1}^{l} \mu(Q_j \cap R) \Big) \Big|. \tag{15}$$

Since $\text{Prob}(\chi_j = 1) = \mu(Q_j \cap R)/\mu(Q_j) = N \cdot \mu(Q_j \cap R)$, we have

$$\text{E}\chi_j = N \cdot \mu(Q_j \cap R), \tag{16}$$

where $\text{E}(\cdot)$ denotes the expected value.

By (15) and (16)

$$D(\mu, R;\ \xi_1, \ldots, \xi_N) = \Big| \sum_{j=1}^{l} (\chi_j - \text{E}\chi_j) \Big|. \tag{17}$$

Since the random variables χ_j, $1 \leqslant j \leqslant l$ are independent of each other, in order to estimate the sum $\sum (\chi_j - \text{E}\chi_j)$ we are able to apply the classical Bernstein–Chernoff inequality of large deviation type.

LEMMA 2 (Bernstein–Chernoff). *Let $\eta_1, \ldots, \eta_l$ be independent random variables with $\text{E}(\eta_i) = 0$ and $|\eta_i| \leqslant 1$, $1 \leqslant i \leqslant l$. Denote by σ_i^2 the variance of η_i, $\sigma_i^2 = \text{E}(\eta_i^2)$. Set $\beta = \big(\sum_{i=1}^{l} \sigma_i^2 \big)^{1/2}$. Then*

$$\text{Prob}\Big(\Big| \sum_{i=1}^{l} \eta_i \Big| \geqslant \gamma \Big) \leqslant \begin{cases} 2e^{-\gamma/4} & \text{if} \quad \gamma \geqslant \beta^2, \\ 2e^{-\gamma^2/4\beta^2} & \text{if} \quad \gamma \leqslant \beta^2. \end{cases}$$

Unfortunately there is no any wide-spread textbook containing this form of Bernstein–Chernoff's inequality, hence we shall present a proof at the end of this section.

Now let us return to (17). Let $\sigma_j^2 = \mathrm{E}(\chi_j - \mathrm{E}\chi_j)^2$ and set $\beta = \left(\sum_{j=1}^{l}\sigma_j^2\right)^{1/2}$. Choose $\gamma = c_4 l^{1/2}(\log N)^{1/2}$ where the constant $c_4 = c_4(r, M, S)$ will be specified later. Since $\beta^2 \leqslant l = |J| = c\lfloor (N\delta^r)^{1-1/r} \rfloor$, by Lemma 2 we obtain

$$\text{(18)}\quad \mathrm{Prob}\left(D(\mu, R;\ \xi_1, \dots, \xi_N) \geqslant (N\delta^r)^{\frac{1}{2}-\frac{1}{2r}}(\log N)^{1/2}\right) = \mathrm{Prob}\left(\left|\sum_{j=1}^{l}(\chi_j - \mathrm{E}\chi_j)\right| \geqslant \gamma\right) \leqslant N^{-c_5},$$

where $c_5 = c_5(r, M, S) \to \infty$ as $c_4 \to \infty$.

Though the class of generalized polytopes is uncountable, it suffices to consider a "small" subclass. A simple argument shows that there is a subclass $\mathscr{R}$ of cardinality $\leqslant N^{c_6(r,M,S)}$ such that given any generalized polytope R_0 with $s(R_0) \leqslant S$, there exist $R_1, R_2 \in \mathscr{R}$ having the properties $R_1 \subseteq R_0 \subset R_2$ and $\mu(R_2 \setminus R_1) \leqslant 1/N$. From this it follows

$$D(\mu, R_0;\ \xi_1, \dots, \xi_N) \leqslant \max\{1, \max_{i=1,2} D(\mu, R_i;\ \xi_1, \dots, \xi_N)\}.$$

This means that we can restrict ourselves to the elements of $\mathscr{R}$. Let $\delta(R)$ denote the diameter of R. By (18)

$$\mathrm{Prob}\left\{D(\mu, R;\ \xi_1, \dots, \xi_N) \geqslant \left(N\cdot\delta^r(R)\right)^{\frac{1}{2}-\frac{1}{2r}}(\log N)^{1/2} \quad \text{for some } R \in \mathscr{R}\right\} \leqslant |\mathscr{R}|\cdot N^{-c_5} \leqslant N^{c_6-c_5} \leqslant 1/2 \quad \text{if } c_5(r, M, S) > c_6(r, M, S).$$

Thus the proof of Theorem 2* is complete.

As we promised, now we shall give a proof of Lemma 2. Set $S_l = \sum_{i=1}^{l}\eta_i$. Clearly

$$\text{(19)}\quad \mathrm{Prob}(S_l \geqslant \gamma) = \mathrm{Prob}(e^{yS_l} \geqslant e^{y\gamma}) \leqslant \frac{\mathrm{E}e^{yS_l}}{e^{y\gamma}},$$

where the parameter y will be fixed later. Since S_l is the sum of independent random variables, we have

$$\mathrm{E}e^{yS_l} = \prod_{i=1}^{l}\mathrm{E}e^{y\eta_i}.$$

We give an upper bound on $\mathrm{E}e^{y\eta_i}$. Using $e^x = \sum_{n=0}^{\infty} x^n/n!$ we get after some easy calculation

$$\text{(20)}\quad \mathrm{E}(e^{y\eta_i}) = \sum_{n=0}^{\infty} y^n\frac{\mathrm{E}(\eta_i^n)}{n!} = 1 + 0 + \frac{y^2\sigma_i^2}{2} + \sum_{n=3}^{\infty} y^n\frac{\mathrm{E}(\eta_i^n)}{n!}$$

$$\leqslant 1 + \frac{y^2\sigma_i^2}{2} + \sum_{n=3}^{\infty} y^n\frac{\mathrm{E}(\eta_i^2)}{n!} \leqslant 1 + \frac{y^2\sigma_i^2}{2} + \frac{\sigma_i^2}{6}\sum_{n=3}^{\infty}\frac{y^n}{3^{n-3}} = 1 + \frac{y^2\sigma_i^2}{2} + \frac{\sigma_i^2}{6}\cdot\frac{y^3}{1-(y/3)}.$$

If we substitute (20) into (19) we obtain

$$\text{(21)}\quad \mathrm{Prob}(S_l \geqslant \gamma) \leqslant \exp\left\{\frac{y^2\beta^2}{2}\left(1 + \frac{y}{3-y}\right) - y\gamma\right\}.$$

We distinguish two cases. If $\gamma \geqslant \beta^2$, let $y = 1$. Then by (21),

$$\mathrm{Prob}(S_l \geqslant \gamma) \leqslant e^{-\gamma/4}.$$

If $\gamma \leqslant \beta^2$, let $y = \gamma/\beta^2$. Again by (21),

$$\mathrm{Prob}(S_l \geqslant \gamma) \leqslant e^{-\gamma^2/4\beta^2}.$$

Repeating the same calculation for $\mathrm{Prob}(S_l \leqslant -\gamma)$ we obtain the desired upper bounds. Lemma 2 follows.

4. Proof of Theorem 3. Let $K = \lfloor (\log N)^{1-\varepsilon} \rfloor$ (integral part) and $l_i = \left\lfloor i\frac{N}{K} \right\rfloor$, $0 \leqslant i \leqslant K$. We partition $[1, N]$ into K almost equal segments $J_i = (l_{i-1}, l_i]$, $1 \leqslant i \leqslant K$. We shall select from each interval J_i exactly one integer a_i such that for each positive triplet h, q and $k \leqslant K$,

$$\text{(22)}\quad \left|\left|\{a_i:\ 1 \leqslant i \leqslant k,\ a_i \equiv h \pmod q\}\right| - k/q\right| \leqslant c_7(\varepsilon).$$

Then we will be ready to finish the proof. Indeed, choosing $S = \{a_i:\ 1 \leqslant i \leqslant K\}$ we obtain

$$\text{(23)}\quad D(h, q, m;\ S) = \left||S \cap \mathrm{AP}(h, q, m)| - \frac{|S|}{N}|\mathrm{AP}(h, q, m)|\right| \leqslant \left||\{a_i:\ a_i \leqslant m,\ a_i \equiv h \pmod q\}| - \frac{k}{q}\right| + \left|\frac{k}{q} - \frac{|S|}{N}|\mathrm{AP}(h, q, m)|\right|,$$

where k is the greatest index i such that $a_i \leqslant m$. Clearly

$$\text{(24)}\quad \frac{|S|}{N}|\mathrm{AP}(h, q, m)| = \frac{K}{N}\left(\frac{m}{q} + O(1)\right)$$

and

$$\text{(25)}\quad k = \frac{mK}{N} + O(1).$$

Thus, by (22), (23), (24) and (25)

$$D(h, q, m;\ S) \leqslant c_7(\varepsilon) + O(1)$$

for all h, q, m with $1 \leqslant m \leqslant N$, $1 \leqslant q \leqslant N$, $1 \leqslant h \leqslant q$, and Theorem 3 follows.

In order to construct the desired sequence $a_i, 1 \leqslant i \leqslant K$ let $f(i) = \lfloor i^{1+\varepsilon/2} \rfloor$ and consider the set B_i of all integers $b_{ij} = j \cdot (f(i)!) + i$ with $b_{ij} \in J_i$. Simple calculation shows that $K \cdot (f(K)!) < N/K$, thus $|B_i| \geqslant K$ for each $1 \leqslant i \leqslant K$. Let ξ_i be a "random element" of B_i, i.e.,

$$\text{Prob}(\xi_i = b_{ij}) = 1/|B_i| \quad \text{for all } 1 \leqslant j \leqslant |B_i|.$$

Moreover, assume that the random variables $\xi_i, 1 \leqslant i \leqslant K$ are independent of each other. We claim that the random K-element subset $\{\xi_1, \ldots, \xi_K\} \subset [1, N]$ satisfies the property (22) with probability $\geqslant 1/2$, that is, the probability of the event "for all h, q, k with $1 \leqslant k \leqslant K$, $1 \leqslant q \leqslant K$, $1 \leqslant h \leqslant q$ the difference of $|\{\xi_i\colon 1 \leqslant i \leqslant k,\ \xi_i \equiv h \pmod q\}|$ and k/q has absolute value less than $c_7(\varepsilon)$" is at least $1/2$. Clearly this will complete the proof of Theorem 3.

Now fix a positive triplet $h, q, k \leqslant K$. We shall estimate

$$\text{Prob}\left\{\left||\{\xi_i\colon 1 \leqslant i \leqslant k,\ \xi_i \equiv h \pmod q\}| - k/q\right| > t\right\}$$

from above. Let $i_0 = i_0(q)$ be the smallest integer i such that $f(i+1) \geqslant q$. From the definition of the numbers b_{ij} it follows that $\xi_i \equiv i \pmod q$ whenever $i > i_0$, hence

$$\left||\{\xi_i\colon i_0 < i \leqslant k,\ \xi_i \equiv h \pmod q\}| - (k - i_0)/q\right| \leqslant 1. \tag{26}$$

Therefore it remains to investigate ξ_i, $1 \leqslant i \leqslant i_0$. Denote by (a, b) the largest common divisor of the natural numbers a and b. Now let j_1 denote the smallest integer such that $(f(j_1)!, q) = d_1 > 1$ and $j_1 \equiv h \pmod{d_1}$. Fixing any integer x, among the numbers

$$l \cdot (f(j_1)!) + j_1, \quad l \in \left[x, x + \frac{q}{d_1} - 1\right]$$

there exists exactly one which is $\equiv h \pmod q$. From this it follows that the probability of the event "$\xi_{j_1} \equiv h \pmod q$" equals $\dfrac{d_1}{q} + O\left(\dfrac{1}{|B_{j_1}|}\right) = \dfrac{d_1}{q} + O\left(\dfrac{1}{q}\right)$, where the implicit constant is absolute (we recall that for each i, $|B_i| \geqslant K \geqslant q$). Since q is divisible by d_1, $\xi_i \equiv i \pmod{d_1}$ for all $i \geqslant j_1$. Hence in the interval $[j_1+1, j_1+d_1-1]$ there is no integer i such that $\xi_i \equiv h \pmod q$. Let $\eta_i = 1$ if $\xi_i \equiv h \pmod q$, 0 otherwise. Using this notation we have obtained

$$\sum_{i=j_1}^{j_1+d_1-1} \text{Prob}(\eta_i = 1) = d_1/q + O(1/q).$$

Now let j_2 be the second integer such that

$$(f(j_2)!, q) = d_2 > 1 \quad \text{and} \quad j_2 \equiv h \pmod{d_2}.$$

Obviously $j_2 \geqslant j_1 + d_1$. Repeating the previous argument we get

$$\sum_{i=j_2}^{j_2+d_2-1} \text{Prob}(\eta_i = 1) = d_2/q + O(1/q),$$

and so on. Summarizing, we conclude that

$$\sum_{i=1}^{i_0} \text{Prob}(\eta_i = 1) \leqslant c_8 \cdot i_0/q, \tag{27}$$

where c_8 is a universal constant. We need the following simple probabilistic lemma.

LEMMA 3. *Let $\eta_1, \ldots, \eta_s$ be independent random variables having values only 0 and 1. Set*

$$p = \sum_{i=1}^{s} \text{Prob}(\eta_i = 1).$$

If $p < 1$, then

$$\text{Prob}\left\{\sum_{i=1}^{s} \eta_i \geqslant t\right\} \leqslant p^t/(1-p).$$

Its proof is a direct calculation. Set $p_i = \text{Prob}(\eta_i = 1)$. We have

$$\begin{aligned}\text{Prob}\left\{\sum_{i=1}^{s} \eta_i \geqslant t\right\} &= \sum_{j=t}^{s} \sum_{1 \leqslant i_1 < \ldots < i_j \leqslant s} p_{i_1} \cdot \ldots \cdot p_{i_j} \prod_{i \in [1,s] \setminus \{i_1, \ldots, i_j\}} (1 - p_i) \\ &\leqslant \sum_{j=t}^{s} \sum_{1 \leqslant i_1 < \ldots < i_j \leqslant s} p_{i_1} \cdot \ldots \cdot p_{i_j} \leqslant \sum_{j=t}^{s} \Big(\sum_{i=1}^{s} p_i\Big)^j \\ &= \sum_{j=t}^{s} p^j \leqslant \sum_{j=t}^{\infty} p^j = p^t/(1-p),\end{aligned}$$

and Lemma 3 is verified.

By (27) and Lemma 3 we obtain

$$\begin{aligned}\text{Prob}\left\{|\{\xi_i\colon 1 \leqslant i \leqslant i_0,\ \xi_i \equiv h \pmod q\}| \geqslant t\right\} \\ \leqslant \text{Prob}\left\{\sum_{i=1}^{i_0} \eta_i \geqslant t\right\} \leqslant (c_8 \cdot i_0/q)^t/(1 - c_8 \cdot i_0/q).\end{aligned} \tag{28}$$

We recall that $f(i) = \lfloor i^{1+\varepsilon/2} \rfloor$ and that $i_0 = i_0(q)$ is the smallest integer for which $f(i_0+1) \geqslant q$. A simple calculation shows $c_8 \cdot i_0/q \leqslant q^{-\varepsilon/4}$ for $q \geqslant c_9$. Choosing $t = \max\left\{c_9, \frac{12}{\lfloor \varepsilon \rfloor}+1\right\}$, by (26) and (28) we have

$$\begin{aligned}\text{Prob}\left\{\left|\left|\{\xi_i\colon 1 \leqslant i \leqslant k,\ \xi_i \equiv h \pmod q\}\right| - k/q\right| > t \text{ for some } h, q, k\right\}\\ \leqslant \sum_{q \geqslant c_9} \sum_{h=1}^{q} \text{Prob}\left\{\left|\{\xi_i\colon 1 \leqslant i \leqslant i_0(q),\ \xi_i \equiv h \pmod q\}\right| \geqslant t\right\}\\ \leqslant \sum_{q \geqslant c_9} \sum_{h=1}^{q} q^{-3} = \sum_{q \geqslant c_9} q^{-2} \leqslant 1/2.\end{aligned}$$

Thus the proof of Theorem 3 is complete.

References

[1] J. Beck, *Roth's estimate of the discrepancy of integer sequences is nearly sharp*, Combinatorica 1(4) (1981), pp. 319–325.

[2] — *Balanced two-colorings of finite sets in the square, I*, ibid. 1(4) (1981), pp. 327–335.

[3] J. Beck and T. Fiala, *"Integer-Making" theorems*, Discrete Applied Math. 3 (1981), pp. 1–8.

[4] J. H. Halton, *On the efficiency of certain quasirandom sequences of points in evaluating multi-dimensional integrals*, Num. Math. 2 (1960), pp. 84–90.

[5] K. F. Roth, *Remark concerning integer sequences*, Acta Arith. 9 (1964), pp. 257–260.

[6] W. M. Schmidt, *Irregularities of distribution, IV*, Inv. Math. 7 (1969), pp. 55–82.

[7] — *Irregularities of distribution, VII*, Acta Arith. 21 (1972), pp. 45–50.

[8] — *Lectures on irregularities of distribution*, Tata Institute of Fundamental Research, Lectures on Math. and Phys. 56 (1977).

MATHEMATICAL INSTITUTE
OF THE HUNGARIAN ACADEMY OF SCIENCES
H-1053 Budapest, Reáltanoda u. 13-15
Hungary

Received on 27. 11. 1981
and in revised form on 26. 7. 1982 (1280)

ACTA ARITHMETICA
XLIII (1984)

Polynômes de $F_q[X]$ ayant un diviseur de degré donné

par

MIREILLE CAR (Marseille)

1. Introduction. Soit F_q le corps fini à q éléments et $F_q[X]$ l'anneau des polynômes à une indéterminée sur le corps F_q.

Si -1 n'est pas carré dans le corps F_q, il y a une similitude parfaite entre l'étude des sommes de deux carrés dans $F_q[X]$ et l'étude des sommes de deux carrés dans l'anneau Z, comme le montrent la caractérisation des sommes de deux carrés donnée dans [7] ou l'estimation du nombre $A(n)$ de polynômes unitaires de degré n qui sont sommes de deux carrés obtenue dans [1].

Si -1 est carré dans le corps F_q, cette similitude disparait, et, le cas de la caractéristique 2 excepté, le problème des sommes de deux carrés devient trivial puisque tout polynôme de $F_q[X]$ est alors somme de deux carrés. On rend ce problème moins trivial en exigeant dans les sommes de deux carrés les conditions de degré les plus restrictives possibles, conditions qui sont automatiquement réalisées dans le cas où -1 n'est pas carré dans le corps F_q. Les polynômes de degré $2n$ ou $2n-1$ sommes de deux carrés

$$A^2 + B^2$$

où A et B sont des polynômes de degré au plus égal à n sont les polynômes de degré $2n$ ou $2n-1$ admettant un diviseur de degré n.

Par une méthode semblable à celle qu'ont utilisée Erdös [4] et Tenenbaum [9], on obtient dans [2] une estimation asymptotique du nombre $A(N)$ de polynômes unitaires de degré N de $F_q[X]$ ayant un facteur de degré égal à la partie entière de $N/2$, estimation donnée par le théorème suivant:

THÉORÈME. *Pour tout réel $\varepsilon > 0$, il existe un entier $N(q, \varepsilon)$ ne dépendant que de q et de ε, tel que, pour tout entier $N \geqslant N(q, \varepsilon)$ on ait*

$$\frac{q^N}{N^{a+\varepsilon}} \leqslant A(N) \ll \frac{q^N}{N^a}\,(\log N)^{-1/2},$$