Astérisque

ALAIN PLAGNE

On the two-dimensional subset sum problem

Astérisque, tome 258 (1999), p. 375-409

<http://www.numdam.org/item?id=AST_1999__258__375_0>

Astérisque **258**, 1999, p. 375–409

ON THE TWO-DIMENSIONAL SUBSET SUM PROBLEM

by

Alain Plagne

Abstract. — We consider a system of two linear boolean equations. Using methods from analytic number theory, we obtain sufficient conditions ensuring the solvability of the system. This completes Freiman's work on the subject.

1. Introduction

In this paper, we are interested in considering the system of two linear equations

$$a_1x_1 + \cdots + a_mx_m = b, \tag{1}$$

where $a_i = (a_{i,1}, a_{i,2})$ and $b = (b_1, b_2)$ are in $\mathbb{Z}^2$ and the x_i's, the unknowns, restricted to be either 0 or 1: that is, we are only interested in the boolean system induced by (1). Our intention is to give sufficient conditions for the set of coefficients $A = \{a_1, \ldots, a_m\}$ and b to ensure the solvability of (1). Probabilistic considerations show that, if the a_i's are "well distributed" and if their number is large enough, we should have solutions for all b in the neighbourhood of $\sum_{i=1}^{m} a_i/2$ and, more precisely, that the distribution of the number of solutions must be Gaussian: in fact, we are expecting a central limit theorem. So that here we investigate conditions ensuring a "good" distribution and then deduce the general case, that is, we describe the structure of A^*, the set of all sums $a_1x_1 + \cdots + a_mx_m$ with boolean unknowns.

The corresponding one-dimensional problem has been much studied in the past recent years from this point of view (see for example **[F80, AF88, EF90, F93]** and **[C91b]** for a complete bibliography). It has been shown that A^* is a collection of arithmetical progressions with the same difference. Each of these papers uses methods coming from analytic number theory, in the vein introduced in the 80's by Freiman (in the first quoted paper), essentially the principle of the circle method.

1991 ***Mathematics Subject Classification.*** — 11P99, 11P55, 11H06.

Key words and phrases. — additive number theory, structure theorem, subset sum, two-dimensional subset sum problem, geometry of numbers, integer points, Farey dissection, convex set.

Freiman began to generalize these results in two dimensions **[F96]** but some details remained obscure (computations on page 143 for example). A little later, Chaimovich **[C91a]** tried to generalize this in higher dimensions but some algorithmical problems arose in these cases (see, for example, our counterexample in section 2.3 to the extension of Proposition 4 stated in **[C91a]**). Our goal here is to make clear the situation. We complete, correct and improve in some places Freiman's **[F96]**. In addition, the results given here are in an explicit form, because of the opportunity they offer to design algorithms. However the constants for which we prove the theorems are still far from being the best one could expect.

For the sake of completeness, the present paper is self-contained except for very classical tools (as, for example, Farey dissection) for which we refer as usual to **[HW]**.

In this paper we shall use the following notation: if u is in $\mathbb{R}^2$, we denote by u_1 and u_2 its coordinates with respect to the canonical basis (ϵ_1, ϵ_2) and by O the origin point. The e function is, as usual, defined by $e(t) = \exp(2\pi i t)$. For a real t, $||t||$ will denote the distance between t and $\mathbb{Z}$ and $[t]$ its integer part. The usual Euclidean scalar product is denoted simply with a point and the Lebesgue measure is denoted by μ. Finally, the volume of a fundamental parallelogram of any lattice Γ is denoted $\mathrm{Vol}\,\Gamma$.

When $k, l \geq 1$ (in order to deal with really two-dimensional problems), we denote $P_{k,l}$ the integer rectangle

$$P_{k,l} = ([-k,k] \times [-l,l]) \cap \mathbb{Z}^2$$

and v its "volume", $v = (2k+1)(2l+1)$, that is, the number of integer points of $P_{k,l}$. In the sequel, A will denote a set of $m = |A|$ different integer points, $A = \{a_1, \ldots, a_m\}$ and $J(b)$ the number of solutions of (1). We write $M = \sum_{i=1}^m a_i/2$ and

$$V = \begin{pmatrix} V_1^2 & V_{12} \\ V_{12} & V_2^2 \end{pmatrix},$$

where we have put $V_{12} = \sum_{j=1}^m a_{j,1}a_{j,2}$ and $V_i^2 = \sum_{j=1}^m a_{j,i}^2$ for $i = 1, 2$.

We denote by q_V the quadratic form naturally associated to this matrix $q_V(x) = \sum_{j=1}^m (a_j.x)^2$ $(x \in \mathbb{R}^2)$, and by $q_{V^{-1}}$ that one associated to V^{-1} that is $q_{V^{-1}}(x) = \frac{1}{\det V}\sum_{j=1}^m \det^2(a_j, x)$. Finally, we define the constants

$$k_1 = 25, \qquad k_2 = 6, \qquad k_4 = 189912,$$
$$k_5 = 100k_1 = 2500, \quad k_6 = 100k_2 = 600,$$
$$k_8 = \max(10k_6, k_5) = 6000, \qquad k_9 = \frac{9}{20},$$

and $k_3 = k_7$ being any constant $< 1/2$.

Our aim is to prove the following three Theorems:

Theorem 1. — *Let $A \subset P_{l_1,l_2}$ and $v = (2l_1+1)(2l_2+1)$. Assume*

$$|A| \geq k_1 v^{2/3} \log^{1/3} v \tag{2}$$

and that for each integer lattice Γ different from $\mathbb{Z}^2$ we have

$$|A \setminus A \cap \Gamma| \geq k_2 v^{2/3} \log^{1/3} v, \tag{3}$$

then we have the following asymptotic equivalent (when $v \to +\infty$)

$$J(b) \sim \frac{2^{m+1}}{\pi\sqrt{\det V}} \exp\{-2q_{V^{-1}}(M-b)\}, \tag{4}$$

provided that $q_{V^{-1}}(M-b) \leq k_3 \log\log v - 4$.

Notice first that the density hypothesis (2) implies

$$\frac{v}{\log v} \geq k_1^3,$$

that implies

$$v \geq k_4. \tag{5}$$

The previous Theorem is slightly better than Freiman's Theorem 1 of **[F96]**, the main difference being that the size of domain of validity of (4) is increased by a factor $\log\log v$ tending to infinity with v. This result is the heart of this work, but this is not entirely satisfying because dealing only with rectangle cases. That is why it is generalized in the following form.

Theorem 2. — *Let C be a compact convex set in $\mathbb{R}^2$ containing O, E be its integer points, and A be a subset of E. Assume*

$$|A| \geq k_5 |E|^{2/3} \log^{1/3} |E|$$

and that for each integer lattice Γ different from $\mathbb{Z}^2$, we have

$$|A \setminus A \cap \Gamma| \geq k_6 |E|^{2/3} \log^{1/3} |E|, \tag{6}$$

then we have the following asymptotic equivalent (when $|E| \to +\infty$)

$$J(b) \sim \frac{2^{m+1}}{\pi\sqrt{\det V}} \exp\{-2q_{V^{-1}}(M-b)\},$$

provided that $q_{V^{-1}}(M-b) \leq k_7 \log\log |E| - 4$.

Once again, it is not completely satisfying because it deals only with "good" cases: those where the elements of A are "well distributed". The conclusion of this paper will be the following general result.

Theorem 3. — *Let C be a compact convex set in $\mathbb{R}^2$ containing O, E be its integer points, and A be a subset of E. Assume $|A| \geq k_8 |E|^{2/3} \log^{1/3} |E|$ and that for each line D such that $O \in D$, one has*

$$|A \cap D| < k_9 |A|. \tag{7}$$

Then there exists a lattice Λ_0 such that, if A' stands for $A \setminus A \cap \Lambda_0$, one has $|A'| \leq |A \cap \Lambda_0|$ and

$$A'^* + (\Lambda_0 \cap F) \subset A^*,$$

where $F = \{x \in \mathbb{Z}^2, q_{W^{-1}}(M'-x) \leq k_7 \log\log(|A|/2) - 4\}$ and $W(x) = \sum_{a \in A'} (a.x)^2$, $M' = \sum_{a \in A'} a/2$.

This is a structural theorem because it describes how the set A^* is made, at least locally. It is a powerful result in order to design algorithms, as it has already been done in the one-dimensional case (see for example **[CFG89]**).

We notice that hypothesis (7) is in fact not very restrictive: it ensures that our set A is an essentially two-dimensional set. If that condition is not fulfilled, we have the possibility to treat our problem as a one-dimensional one, by forgetting some points and this is even much simpler.

Acknowledgments. — Many thanks go to G.A. Freiman for his kind help during the preparation of this paper. I would like also to thank J.-M. Deshouillers for his advice.

2. Preliminary lemmas

We begin this section by quoting some inequalities (whose validity can easily be seen by using, for instance, some Taylor-Lagrange's inequalities). For any real t, if $0 \leq |t| \leq 1/2$, we have

$$|1 + e(t)| \leq 2\exp(-\pi^2 t^2/2), \tag{8}$$

and if $|t| \leq \pi/2$,

$$0 \leq 1 - \exp(t^2/2)\cos t \leq (2t/\pi)^4. \tag{9}$$

Finally, for reals $(\epsilon_i)_{1\leq i\leq n}$ between 0 and 1, we have, with a trivial induction argument,

$$\prod_{i=1}^{n}(1-\epsilon_i) \geq 1 - \sum_{i=1}^{n}\epsilon_i. \tag{10}$$

Now we present several propositions that we shall need in the sequel.

2.1. Arithmetical lemmas. — Here, we give two results concerning the number of solutions of a Diophantine inequality.

Lemma 1. — *Let a, b, ϵ be real numbers and k, n be integers such that $0 < |a|k < 1$ and $\epsilon < (1 - k|a|)/2$. Then we have*

$$|\{x \in \mathbb{N}, n \leq x \leq n + k : ||ax + b|| \leq \epsilon\}| \leq 1 + [2\epsilon/|a|].$$

Proof. — Without loss of generality we may assume $a > 0$ and write $u_s = as+b$. This is a strictly increasing sequence. Let s_1 be the smallest integer, with $n \leq s_1 \leq n + k$, such that $||u_{s_1}|| \leq \epsilon$ (if s_1 does not exist, then the cardinality studied is zero); we thus have $|u_{s_1} - e| \leq \epsilon$ for some integer e. Let s_2 be the largest integer satisfying $|u_{s_2} - e| \leq \epsilon$. We claim that $s_2 < t \leq n + k$ implies $||u_t|| > \epsilon$; indeed $|u_t - e| > \epsilon$ is clear by definition of s_2 and

$$u_t = u_{s_1} + (t - s_1)a \leq u_{s_1} + ka \leq e + \epsilon + ka < e + 1 - \epsilon.$$

Since $s_2 - s_1 = (u_{s_2} - u_{s_1})/a \leq 2\epsilon/a$, we get, for the cardinality studied, the desired upper bound. □