ON THE SPATIAL DISTRIBUTION OF SOLUTIONS OF DECOMPOSABLE FORM EQUATIONS

G. Everest, I. Gaál, K. Györy, C. Röttger

School of Mathematics, UEA Norwich
and
Institute of Mathematics, Debrecen

Abstract. We study the distribution in space of the integral solutions to an integral decomposable form equation, by considering the images of these solutions under central projection onto a unit ball. If we think of the solutions as stars in the night sky, we ask what constellations are visible from the earth (the unit ball). Answers are given for a large class of examples which are then illustrated using the software packages KANT and Maple. These pictures highlight the accuracy of our predictions and arouse interest in cases not covered by our results. Within the range of applicability of our results lie solutions to norm form equations and units in abelian group rings. Thus our theory has a lot to say about where these interesting objects can be found and what they look like.

Introduction

Let $F(\underline{x}) = F(x_1, \ldots, x_n) \in \mathbf{Z}[x_1, \ldots, x_n]$ denote a *decomposable form.* This is a homogeneous polynomial with coefficients in $\mathbf{Z}$ which factorises over $\mathbf{C}$ as a product of linear forms. We assume throughout this paper that F contains n linearly independent linear forms among its factors. It is known that there are $q \in \mathbf{Q}^*$, finite extension fields $M_1, \ldots, M_t$ of $\mathbf{Q}$ and linear forms $\phi_i(\underline{x})$ with coefficients in $M_i, i = 1, \ldots, t$ such that

$$F(\underline{x}) = q \prod_{i=1}^{t} N_{M_i|\mathbf{Q}}(\phi_i(\underline{x})). \tag{0.1}$$

In (0.1), $N_{M_i|\mathbf{Q}} : M_i \to \mathbf{Q}, i = 1, \ldots, t$, denotes the field norm. Given a non-zero $a \in \mathbf{Z}$, the *decomposable form equation*

$$F(\underline{x}) = a, \quad \underline{x} \in \mathbf{Z}^n, \tag{0.2}$$

is a very general equation with many important examples.

1991 *Mathematics Subject Classification.* 11D57,11Y50.

Röttger's research was supported by a PhD grant from the UEA. Györy thanks the LMS for a scheme 2 grant at an early stage of this research. Györy and Gaál were supported by the Hungarian Academy of Sciences and by grants 16975, 25157 and 29330 from the Hungarian National Foundation for Scientific Research.

Examples.

(1) If $d > 0$ is a non-square integer then the equation

$$x_1^2 - dx_2^2 = 1,$$

is a decomposable form equation, with $t = 1$ and $M_1 = \mathbf{Q}(\sqrt{d})$, known as *Pell's equation.* This has been studied extensively because of its inter-relation with many other branches of number theory.

(2) A more general case of Pell's equation, also with $t = 1$, is the *norm form equation.* Here $q = 1, \phi_1(\underline{x}) = \sum_i a_i x_i$ and the a_i lie in the ring of integers of M_1. See [Sc1], [Sc2] and [Sc3] for background to this equation. Schmidt made some fundamental breakthroughs in the study of the norm form equation, using powerful techniques from diophantine approximation. In this case, when $a = \pm 1$ and the a_i form a $\mathbf{Z}$-basis for the ring of integers, the solutions correspond to units of the number field M_1. Our results are new, even in this special case.

(3) A less well known example of a decomposable form equation arises with the study of units in abelian group rings. Let Γ denote a finite abelian group with $\mathbf{Z}\Gamma$ denoting the integral group ring. This is the set of all expressions

$$\sum_{\gamma \in \Gamma} x_\gamma \gamma, \quad x_\gamma \in \mathbf{Z}.$$

This set forms a ring with component-wise addition, and with multiplication respecting both the operation in Γ and the distributive law. There is considerable interest in the group of units of this ring; see [K] and [Se] for details. In [EG], we showed that the group of units can be identified with the integral solutions to two decomposable form equations. Our methods give refined information about where the units in group rings lie and what they look like; see especially section 2.

This paper is about the internal structure of the set of solutions of the decomposable form equation. We deal with the case when equation (0.2) has infinitely many solutions. Our results are trivial if the number of solutions is finite. Given any positive real number T, let $F(a, T)$ denote the set

$$F(a, T) = \{\underline{x} \in \mathbf{Z}^n : \underline{x} \text{ satisfies (0.2) and } |\underline{x}| < T\}. \tag{0.3}$$

In (0.3), $|\underline{x}|$ denotes the 'max'-norm defined by $|\underline{x}| = \max_{1 \le i \le n}\{|x_i|\}$. Since it is clear that $F(a, T)$ is a finite set, the two questions that follow are natural:

Q1 What is the size of $F(a, T)$ for large T? In other words, what is the asymptotic behaviour of $|F(a, T)|$, the cardinality of the set?

Q2 For every $\underline{x} \in F(a, T)$, let $c(\underline{x}) = \underline{x}/|\underline{x}|$ denote the central projection of $\underline{x}$ onto the unit ball centred at $\underline{0}$. We ask what is the asymptotic distribution of the images of the elements of $F(a, T)$ under this projection? In other words, can this set of points be described when T is large. One imagines the solutions of the equation as corresponding to the stars in the night sky. Standing upon the earth (the unit ball) and looking up, what constellations would be visisble?

It was proved in [G] that the set of solutions of (0.2) is the union of finitely many families of solutions (to be defined in section 1). In [EG] this was combined with a new variant of the Hardy-Littlewood method to give a very accurate formula in answer to Q1. The diophantine input to this method is Schmidt's Subspace Theorem, which is a powerful generalisation of Roth's Theorem (see [Sc2], [Sc3]). In this paper, whose layout is now described, we are going to present theoretical and computational answers to Q2. In section 1, the results from [EG] will be recalled then recast, as Theorem 1, in geometric terms. We suggest that section 1 is read in tandem with section 4 at the end of the paper. Here computational results are presented in the form of pictures, which give a convincing account of the phenomena in [EG]. Examples are also included of cases not covered by our results. Besides their aesthetic appeal, we found these pictures inspired both our curiosity and our understanding. Theorem 2 (see (1.4) and (1.5)), which is stated in section 1 (see (1.11)), is an explanation of some of the pictures on view in section 4. Section 2 considers more deeply the geometric clustering phenomenon from section 1 and proves Theorem 1. Section 3 proves Theorem 2, using the theory of uniform distribution.

§1 Statement of Results

Write $P(T)$ for the cardinality of $F(a,T)$. In [EG], we showed there is a two-term asymptotic formula for $P(T)$ and we specified a large class of examples where a three-term asymptotic formula holds. This class of examples will now be defined.

We say F is of *CM type* if the M_i in (0.1) are totally real fields or totally imaginary quadratic extensions of totally real fields and none of them has a (not necessarily proper) subfield of unit rank 1. If $n = \sum_{i=1}^{t}[M_i : \mathbf{Q}]$ then the condition on the subfields of $M_1, \dots, M_t$ can be omitted. This condition is very slightly broader than the one in Theorem 2 of [EG]. There we insisted that the unit ranks all be greater than 1 but this is not necessary. It is only rank equal to 1 that we wish to avoid.

Theorem A. *([EG]) Suppose that (0.2) has infinitely many solutions. With $P(T)$ as above:*
(i) (See also [EvG].) There is a positive integer r, defined by (1.9), and a constant $\rho_1 > 0$ depending on F and a such that

$$P(T) = \rho_1(\log T)^r + O((\log T)^{r-1}), \quad T \to \infty. \tag{1.1}$$

(ii) If F is of CM type and if r, defined by (1.9), is greater than 1, then there are constants $\rho_1 > 0$, ρ_2 depending on F and a such that

$$P(T) = \rho_1(\log T)^r + \rho_2(\log T)^{r-1} + o((\log T)^{r-1}), \quad T \to \infty. \tag{1.2}$$

Thus Q1 is answered fairly successfully. Of particular note is the three-term formula (1.2) in the CM case. Question Q2 was posed to try to understand better the implications of this three-term formula in the CM case. Also because our curiosity was aroused as to what can happen in the non-CM case. The method of proof of Theorem A has already implicit within it statements about distribution of the kind in Q2. We are now going to bring these to the fore in Theorem 1.

Let Y denote a finite dimensional real vector space, and let N denote an arbitrary Euclidean norm on Y. This is a function $N : Y \to \mathbf{R}_{\geq 0}$ which satisfies the following three properties:
(i) $N(\underline{y}) = 0$ if and only if $\underline{y} = \underline{0}$,
(ii) $N(\lambda \underline{y}) = |\lambda| N(\underline{y})$ for all $\lambda \in \mathbf{R}$ and all $\underline{y} \in Y$,
(iii) $N(\underline{y}_1 + \underline{y}_2) \leq N(\underline{y}_1) + N(\underline{y}_2)$, for all $\underline{y}_1, \underline{y}_2 \in Y$.
Any two Euclidean norms N_1 and N_2 on a given Y are *commensurate* in the sense that

$$N_1(\underline{y}) << N_2(\underline{y}) << N_1(\underline{y}), \text{ for all } \underline{y} \in Y, \tag{1.3}$$

where, in (1.3), the constants implied by the notation are uniform.

Given a fixed N on a fixed space Y, let S_N denote the surface of the unit ball with respect to N; $S_N = \{\underline{y} \in Y : N(\underline{y}) = 1\}$. Given any $\underline{0} \neq \underline{y} \in Y$, let $c_N(\underline{y})$ denote the central projection with respect to N; $c_N(\underline{y}) = \underline{y}/N(\underline{y})$. If $R \subset S_N$, write

$$P_R(T) = \#\{\underline{x} \in F(a,T) : c_N(\underline{x}) \in R\}.$$

Given any point Q on S_N and $\epsilon > 0$, let $Q(\epsilon)$ denote the ϵ-neighbourhood of Q on S_N. If $R \subset S_N$, let $R(\epsilon)$ denote the union of the $Q(\epsilon)$ for $Q \in R$.

Before we state Theorem 1, we wish to be clear that solutions of (0.2) are being counted with respect to the max norm. They are projected onto the unit ball S_N with respect to a fixed Euclidean norm N, whose shape will obviously depend upon N. In section 4, we always choose N to be the max norm so S_N is a cube (in whatever dimension).

Theorem 1. *Suppose that equation (0.2) has infinitely many solutions. Assume the CM case, and let N denote any Euclidean norm on $\mathbf{R}^n$.*
(i) There is a set of points $V = \{Q_1, \ldots, Q_m\} \subset S_N$ such that for any $\epsilon > 0$, with ρ_1 and r as in (1.1),

$$P_{V(\epsilon)}(T) = \rho_1 (\log T)^r + O((\log T)^{r-1}), \quad T \to \infty. \tag{1.4}$$

(ii) Suppose $r > 1$ and let W denote the union of the projections to S_N of the straight lines joining the $Q_i, i = 1, \ldots, m$. Then for any $\epsilon > 0$, with ρ_2 as in (1.2),

$$P_{W(\epsilon)}(T) = \rho_1 (\log T)^r + \rho_2 (\log T)^{r-1} + o((\log T)^{r-1}), \quad T \to \infty. \tag{1.5}$$

Formulae (1.4) and (1.5) say that ‘most’ of the images of the solutions cluster around the lines comprising W and ‘most’ of these cluster more densely around the points in V. In astronomical terms, the formulae posit the existence of finitely many ‘Milky Ways’ which contain finitely many brighter clusters of stars. We refer to V and W as the set of *cluster points* and *cluster lines* respectively. Obviously there is some laxity in the definitions because we can add arbitrary points to V and lines to W without changing the formulae in (1.4) and (1.5). If a point belongs to the set V but the formulae do not change when it is removed, we say it is a *virtual cluster point*, otherwise an *actual cluster point*. Similar definitions hold for cluster lines. Group rings provide natural examples of virtual cluster points.

Of course all of these definitions apply in the CM case up to now. The non-CM case throws open some fascinating possibilities. Formula (1.1) always holds and formula (1.4) holds for certain subsets $V \subset S_N$. One can ask what is the actual subset V, that is, the smallest subset - assuming it exists - of S_N for which (1.4) holds. Potentially there will be examples where formula (1.2) holds. It is a challenge to prove formula (1.5) for these examples and describe the actual cluster regions V and W. We know that (1.2) does not always hold; for example, given non-square, positive integers d_1 and d_2, consider the equation

$$(x_1^2 - d_1 x_2^2)(x_3^2 - d_2 x_4^2) = 1. \tag{1.6}$$

Example 5 in section 4 also does not satisfy (1.2). In examples like these, probably the best formula is one of the shape

$$P(T) = P_{W(\epsilon)}(T) + o((\log T)^{r-1}), \quad T \to \infty. \tag{1.7}$$

One can always take W to be the convex hull of V then it is a challenge to describe the actual region W. The reader can verify that for equation (1.6), the actual V is a finite set of points. The actual W is finite set of lines unless d_1/d_2 is a rational square. In the latter case, the actual W is an infinite set of points lying on finitely many lines with only a finite set of limit points, and the set of limit points is the actual V. See examples 6 and 7 in section 4 for pictures in both cases.

The formulae in Theorem 1 are already implicit in [EG] due to the style of proof of Theorems 1 and 2 in that paper. They will be proved in section 2. We will now describe the distribution of points outside $V(\epsilon)$ in the CM case. Turning to figures 1, 2, 4A and 4B in section 4 arouses suspicion that the distribution is not uniform around the lines joining the cluster points. Theorem 2 below gives the distribution for a particular choice of Euclidean norm in terms of a simple function.

In order to state this, we will need to go into the background to the proof of Theorem A. The solutions of (0.2) lie in a finite number of classes which are orbits of unit groups. The technical term for a class is *family of solutions* and we begin by defining this term.

Let A denote the algebra

$$A = M_1 \oplus \cdots \oplus M_t.$$

This is the $\mathbf{Q}$-algebra direct sum of the number fields $M_1, \dots, M_t$ formed with componentwise operations. Thus, $1_A = (1, \dots, 1)$ is the unity of A and A^*, the multiplicative group of invertible elements of A is $\{(\alpha_1, \dots, \alpha_t) \in A : \alpha_1 \dots \alpha_t \neq 0\}$. The norm $N_{A|\mathbf{Q}}(\alpha)$ of $\alpha = (\alpha_1, \dots, \alpha_t) \in A$ is defined to be the usual algebra norm, i.e. the determinant of the $\mathbf{Q}$-linear map $x \mapsto \alpha x$ from A to itself. The norm is multiplicative and

$$N_{A|\mathbf{Q}}(\alpha) = \prod_{i=1}^{t} N_{M_i|\mathbf{Q}}(\alpha_i).$$

Therefore re-write equation (0.2) as

$$qN_{A|\mathbf{Q}}(c) = a, \qquad c \in \mathfrak{M}, \tag{1.8}$$

where $\mathfrak{M}$ is defined to be $\mathfrak{M} = \{c = (\phi_1(\underline{x}), \ldots, \phi_t(\underline{x})) \in A : \underline{x} \in \mathbf{Z}^n\}$. Now $\mathfrak{M}$ is a finitely generated $\mathbf{Z}$-module. Let $V = \mathbf{Q}\mathfrak{M}$ denote the $\mathbf{Q}$-vector space generated by $\mathfrak{M}$. For any subalgebra B of A with $1_A \in B$, denote by O_B the integral closure of $\mathbf{Z}$ in B and by O_B^* the multiplicative group of invertible elements of O_B. Let

$$V^B = \{v \in V : vB \subseteq V\} \quad \text{and} \quad \mathfrak{M}^B = V^B \cap \mathfrak{M}.$$

Obviously V^B is closed under multiplication by elements of B. Now define

$$U_{\mathfrak{M},B} = \{u \in O_B^* : u\mathfrak{M}^B = \mathfrak{M}^B, \quad N_{A|\mathbf{Q}}(u) = 1\}.$$

This is a subgroup of finite index in O_B^*. If $c \in \mathfrak{M}^B$ is a solution of (1.8) so is every element of $cU_{\mathfrak{M},B}$. Such a coset is called an $(\mathfrak{M}, B)$-*family of solutions* of (1.8), and hence of (0.2) as well. It is a fundamental result in this subject (see [G]) that the set of solutions of (1.8) is a union of finitely many families of solutions.

The group O_B^* is finitely generated, let r_B denote the torsion-free rank. Use r to denote the maximum of the r_B,

$$r = \max_B \{r_B\}, \tag{1.9}$$

taken over all $\mathbf{Q}$-algebras B of A with $1_A \in B$ for which (1.8) has an $(\mathfrak{M}, B)$-family of solutions. Our results are non-trivial only if $r > 0$ and we will assume this is always the case. Any $(\mathfrak{M}, B)$-family with $r = r_B$ is called a *maximal family.* In the CM case, we may replace $U_{\mathfrak{M},B}$ by a subgroup $\overline{U}_{\mathfrak{M},B}$ of finite index consisting of elements $(u_1, \ldots, u_t)$ with totally real and totally positive units $u_1, \ldots, u_t$ from $M_1, \ldots, M_t$ respectively. When this is done, we refer to *real families* and *maximal real families* with the obvious abuse of language. (A real family does not necessarily consist of real numbers; rather, of numbers which are the orbit of a group consisting of real numbers.) Note, in the CM case, that the solutions of (0.2) are contained in a union of finitely many real families. It is therefore sufficient to do any counting within a fixed, maximal real family of solutions which we denote $\mathfrak{F}$. Write

$$\mathfrak{F}(T) = \mathfrak{F} \cap F(a, T).$$

It will be easier to state Theorem 2 assuming the condition

$$\sum_{i=1}^{t} [M_i : \mathbf{Q}] = n. \tag{1.10}$$

The condition (1.10) holds in all three examples in the introduction and in the computations in section 4. For any $R \subset S_N$, write R' for $S_N - R$, the set-theoretic complement. Also, for a fixed choice of norm N, write

$$P_R(\mathfrak{F}, T) = \#\{\underline{x} \in \mathfrak{F}(T) : c_N(\underline{x}) \in R\}.$$

Theorem 2. *Assume the CM case with $r > 1$ and (1.10). For every maximal real family $\mathfrak{F}$, there is a Euclidean norm and a constant ρ_3 which both depend upon $\mathfrak{F}$ only such that for all $0 < \epsilon < 1$,*

$$P_{V(\epsilon)'}(\mathfrak{F}, T) = \rho_3 \log\left(\frac{1}{\epsilon}\right)(\log T)^{r-1} + o((\log T)^{r-1}), \quad T \to \infty. \tag{1.11}$$

This shows a kind of logarithmic distribution, with sinks at the cluster points. Obviously the distribution function will vary with the choice of Euclidean norm but we could calculate this for any reasonable norm from the formula above. This is because the unit of distance with respect to one norm is a continuous function of the unit distance with respect to another norm. Certainly, it looks clear that the norm in Theorem 2 is canonical in the sense that it gives such a simple distribution for the images of the points $c_N(\underline{x})$.

§2 Cluster Points and Cluster Lines

We are going to show how the cluster points and lines arise enabling a re-statement of Theorems 1 and 2 in terms more amenable to proof. Let $\mathfrak{F}$ denote a fixed maximal real family as in section 1. Write U for the associated $\overline{U}_{\mathfrak{M},B}$. For each $M_i, i = 1, \dots, t$, let $\sigma_{ij} : M_i \to \mathbf{C}$, $j = 1, \dots, [M_i : \mathbf{Q}]$ denote the distinct embeddings into $\mathbf{C}$. Write $\phi_{ij}(\underline{x})$ for the conjugates of the forms $\phi_i(\underline{x})$, $i = 1, \dots, t$. We are assuming that these $\sum_{i=1}^{t}[M_i : \mathbf{Q}]$ forms contain n linearly independent forms. By (1.8) and the definition of the algebra norm, there are algebraic numbers b_{ij} such that for all $x \in \mathfrak{F}$

$$\phi_{ij}(\underline{x}) = b_{ij}u_{ij}, \quad i = 1, \dots, t;\ j = 1, \dots, [M_i : \mathbf{Q}]. \tag{2.1}$$

with algebraic units $u_{ij} = \sigma_{ij}(u_i)$. Write $(u_{ij}) = (u_k)_{1 \le k \le m}$ for the vector of the u_{ij}, $i = 1, \dots, t$, $j = 1, \dots, [M_i : \mathbf{Q}]$, where $m = \sum_{i=1}^{t}[M_i : \mathbf{Q}]$, and similarly for $(b_{ij}) = (b_k)$. Then we obtain a system of m linear equations

$$\Phi\underline{x} = (b_k u_k), \tag{2.2}$$

where the coefficients of the $m \times n$ matrix Φ are those of the linear forms ϕ_{ij}. The system in (2.2) is (left) invertible by the assumption being made about the linear factors of F. Writing Ψ for the left inverse of Φ gives

$$\underline{x} = \Psi(b_k u_k). \tag{2.3}$$

If $u = (u_1, \dots, u_t) \in U$ then define

$$H(u) = \max_{i,j}\{\sigma_{ij}(u_i)\}, \tag{2.4}$$

the largest value of any conjugate of any $u_i, i = 1, \dots, t$. Write $H^*(u)$ for the second largest element of the set in (2.4), where complex conjugate embeddings are identified. It follows from (2.2), (2.3) and the triangle inequality that $|\underline{x}|$ and $H(u)$ are commensurate in the sense of (1.3). Formulae (1.1) and (1.4) come about by exploiting that fact, enabling the counting of solutions of (0.2) in a particular family to be effected by counting elements $u \in U$ with respect to H.

Proof of Theorem 1.

The concepts of cluster point and cluster line are actually indpendent of the choice of Euclidean norm. Once they have been determined for one choice of norm, they can be re-scaled to any other norm. Therefore, it is sufficient to prove existence without reference to any particular norm.

Fix indices (i,j) with $H(u) = u_{ij}$. Using (2.3), there is a vector $\underline{c}$ depending on $\mathfrak{F}$ and the (i,j) only (via Ψ and $\underline{b}$) such that

$$\underline{x} = \underline{c}H(u) + O(H^*(u)). \tag{2.5}$$

A fundamental result from [EG] (see Lemma 6(i)) is that in the CM case, for all $0 < \epsilon < 1$, asymptotically all $u \in U$ have $H^*(u)/H(u) < \epsilon$. In other words,

$$|\{u \in U : \epsilon \leq H^*(u)/H(u), H(u) < T\}| = O((\log T)^{r-1}). \tag{2.6}$$

Formula (1.4) in Theorem 1 follows from (2.5) and (2.6), by varying the indices (i,j) and the maximal real family $\mathfrak{F}$. Each vector $\underline{c}$ in (2.5) gives rise to a cluster point and all cluster points arise in this way. Note that $\underline{c}$ is guaranteed to be real.

Formulae (1.2) and (1.5) come about by refining (2.5) and (2.6) together with a more delicate inter-play between H and $|.|$. Write $H^{**}(u)$ for the third largest element of $\{\sigma_{ij}(u_i)\}$, where complex conjugate embeddings are identified. Fix indices (i,j) and (k,l) with $H(u) = u_{ij}$ and $H^*(u) = u_{kl}$. There is a vector $\underline{c}^*$ with

$$\underline{x} = \underline{c}H(u) + \underline{c}^* H^*(u) + O(H^{**}(u)). \tag{2.7}$$

In [EG] (see Lemma 6(ii)), we proved

$$|\{u \in U : \epsilon \leq H^{**}(u)/H(u), H(u) < T\}| = O((\log T)^{r-2}). \tag{2.8}$$

Formula (1.5) in Theorem 1 follows from (2.7) and (2.8), by varying the indices $(i,j), (k,l)$ (and hence the vectors $\underline{c}$, $\underline{c}^*$) and the maximal real family $\mathfrak{F}$. Each pair of vectors $\underline{c}$ and $\underline{c}^*$ in (2.7) give rise to a cluster line and all cluster lines arise in this way. ■

In the proof of Theorem 1, the quantity $H(u)$ is behaving as though it is a Euclidean norm. By choosing an appropriate basis, we will see that $H(u)$ is indeed a Euclidean norm. Then it is clear that the distribution along the cluster lines depends upon the relative sizes of $H(u)$ and $H^*(u)$. Theorem 2 will be re-stated in these terms (see Theorem 2.1 below) then proved in section 3.

Under condition (1.10), for each maximal real family, there are at most n cluster points. These cluster points are linearly independent, because they are linear combinations of distinct subsets of columns of an invertible matrix. Let $V_{\mathfrak{F}} = \{P_1, \dots, P_l\}$, for a fixed maximal real family, denote the set of linearly independent cluster points. Then the set of solutions $\underline{x}$ in that family lie in the space generated by the $P_i, i = 1, \dots, l$. Write $Y_{\mathfrak{F}} =< P_1, \dots, P_l >= \oplus_{i=1}^l P_i\mathbf{R}$. Let $|.|_{\mathfrak{F}}$ be the Euclidean norm on $Y_{\mathfrak{F}}$ defined by

$$|\underline{y}|_{\mathfrak{F}} = \left|\sum_{i=1}^{l} P_i y_i\right|_{\mathfrak{F}} = \max_{1\le i\le l}\{|y_i|\}. \tag{2.9}$$

Theorem 2.1. *Let $C > 0$ denote a constant and define*

$$U_C(T) = |\{u \in U : e^{-C} < H^*(u)/H(u), H(u) < T\}|. \tag{2.10}$$

There is a positive constant ρ, which depends upon U only, such that $U_C(T)$ satisfies the following asymptotic formula

$$U_C(T) = C\rho(\log T)^{r-1} + o((\log T)^{r-1}), \qquad \text{as } T \to \infty. \tag{2.11}$$

Theorem 2 in §1 follows from Theorem 2.1, with the Euclidean norm in the statement of Theorem 2 taken to be that in (2.9). The solutions of (0.2) are counted with respect to H. This makes no difference to the statement of (1.11) since $H(u)$ is commensurate with $|\underline{x}|$ when $u \in U$ is associated with $\underline{x} \in \mathbf{Z}^n$ (see the remark after (2.4)). Notice that Theorem 2.1 is a much more refined statement than (2.6) above. Theorem 2.1 will be re-stated as Proposition 3.1 in section 3.

It looks as though (2.11) could be strengthened by allowing a 'shrinking target'. The following is probably true; for each $C, D > 0$, as $T \to \infty$,

$$\left|\left\{u \in U : \frac{e^{-C}}{H(u)^D} < \frac{H^*(u)}{H(u)}, H(u) < T\right\}\right| = \nu(\log T)^{r-1}(D\log\log T + C) + o((\log T)^{r-1}).$$

We have chosen not to pursue this because it is not clear how it translates into geometric concepts. Also, we admit, the formula looks a beast to prove.

This section closes with an example where the set of cluster points can be written down explicitly. Studying the units in the integral group ring of a finite abelian group, one can assign to the group ring terms like 'cluster point', which are borrowed from the theory of the corresponding decomposable form equations. Group rings provide natural examples of virtual cluster points. In [EG], we showed that the units of $\mathbf{Z}\Gamma$ yield a CM equation if and only if the following property holds:

$$\text{no quotient of } \Gamma \text{ is cyclic of order 5, 8 or 12.} \tag{2.12}$$

Thus, formulae (1.1) and (1.4) hold always but (1.2) and (1.5) hold only under condition (2.12).

Lemma 2.2. *Suppose Γ is a finite abelian group. Let $\chi \in \hat{\Gamma}$ denote a character and define*

$$e_\chi = \frac{1}{|\Gamma|} \sum_{\gamma \in \Gamma} \overline{\chi}(\gamma)\gamma \in \mathbf{C}\Gamma.$$

(i) The e_χ for $\chi \in \hat{\Gamma}$ form a system of $n = |\Gamma|$ independent, orthogonal idempotents for $\mathbf{C}\Gamma$.
(ii) The elements $e_\chi \chi(\gamma) + e_{\overline{\chi}} \overline{\chi}(\gamma) \in \mathbf{R}\Gamma$ for $\chi \in \hat{\Gamma}, \gamma \in \Gamma$ give the cluster points for $\mathbf{Z}\Gamma^*$.
(iii) The cluster points in (ii) are virtual when $\mathbf{Q}(\chi)$, the field generated over $\mathbf{Q}$ by the values of χ, is $\mathbf{Q}$ or an imaginary quadratic extension of $\mathbf{Q}$.

Note that for any finite abelian group, there will always be virtual cluster points. If we take the trivial character $\chi_0(\gamma) = 1$, for all $\gamma \in \Gamma$ then $\mathbf{Q}(\chi_0) = \mathbf{Q}$. Similarly, if $|\Gamma|$ is even then Γ will have a character of order 2. The values of this character will be ± 1 so the field generated by its values over $\mathbf{Q}$ will be $\mathbf{Q}$.

Proof of Lemma 2.2. We appeal to the results in [E1] and [E2] which are phrased in the language of Dirichlet Series. ■

§3 Counting Units

With the notation in section §2, note that U is a free abelian group of rank r. Taking logarithms of the $\sigma_{ij}(u_i) = u_{ij}$ gives rise to a family of linear forms $L_1, \ldots, L_u$ on U. Each form corresponds to the logarithm of a conjugate of some component of $\underline{u}$. After choosing a basis of U, we may regard the L_i, $i = 1, \ldots, u$ as linear forms on $\mathbf{Z}^r$. Assuming that forms are not counted if they are identically zero, the CM condition (in particular, the prohibition of the rank 1 case) guarantees that at least two of the coefficients of each L_i are linearly independent over $\mathbf{Q}$. The following relation is satisfied by this family of forms,

$$L_1(\underline{x}) + \cdots + L_u(\underline{x}) = 0, \qquad \text{for all } \underline{x} \in \mathbf{Z}^r. \tag{3.1}$$

This comes from the fact that the underlying quantity is a unit so the product of all the conjugates of all the components is equal to 1. Taking logarithms gives the relation in (3.1). Clearly each of the forms extends to $\mathbf{R}^r$ and the same relation (3.1) holds. Counting heights of elements $u \in U$ with $H(u) < T$ is equivalent to counting lattice points $\underline{x} \in \mathbf{Z}^r$ satisfying $L(\underline{x}) = \max_i\{L_i(\underline{x})\} < X = \log T$.

Let $L^*(\underline{x})$ denote the second largest component of the vector $(L_i(\underline{x}))_{1 \le i \le u}$. The inequalities defining $U_C(T)$, in (2.10), become $(X = \log T)$,

$$L(\underline{x}) < X, \quad -C < L^*(\underline{x}) - L(\underline{x}). \tag{3.2}$$

Define the following counting function

$$A_C(X) = |\{\underline{x} \in \mathbf{Z}^r : L(\underline{x}) < X, \quad L^*(\underline{x}) < L(\underline{x}) < L^*(\underline{x}) + C\}|. \tag{3.3}$$

Theorem 2.1 is a direct consequence of Proposition 3.1 following.

Proposition 3.1. *There is a positive constant ρ which depends only upon the linear forms $L_1, \dots, L_u$ such that the following asymptotic formula holds*

$$A_C(X) = C\rho X^{r-1} + o(X^{r-1}), \qquad \text{as } X \to \infty.$$

The proof of Proposition 3.1 will follow after Lemma 3.3. The best approach to proving Proposition 3.1 is the direct one of comparing the number of lattice points being counted with the volume of the region defined by the inequalities in (3.3). But note that the volume of the boundary of the region has the same order of magnitude as the main term of the asymptotic formula so it cannot be used as the error term. However, the boundary is of the type to allow a uniform distribution argument to estimate the error. Let $\mathfrak{S}_C(X)$ denote the region of $\mathbf{R}^r$ defined by the inequalities in (3.3) and let μ_r denote Lebesgue measure in $\mathbf{R}^r$.

Lemma 3.2. *There is a positive constant ρ such that*

$$\mu_r(\mathfrak{S}_C(X)) = C\rho X^{r-1} + O(X^{r-2}).$$

Proof. This is obtained by multiple integration as follows. The region of integration subdivides according to the possible orderings on the forms. After re-labelling, it is sufficient to consider the region $T(X)$ defined by $L(\underline{y}) = L_1(\underline{y}) \leq X$ and

$$L_2(\underline{y}) + C \geq L_1(\underline{y}) \geq L_2(\underline{y}) \geq \cdots \geq L_u(\underline{y}).$$

To avoid discussing trivial cases, let us assume that $T(X)$ has positive volume. For $0 < \gamma \leq C$, let $T_\gamma(X)$ be the region defined by $L(\underline{y}) = L_1(\underline{y}) \leq X$ and

$$L_2(\underline{y}) + \gamma = L_1(\underline{y}) \geq L_2(\underline{y}) \geq \cdots \geq L_u(\underline{y}).$$

A special property of the linear forms L_i is that any two of them are linearly dependent if and only if they are equal (disregarding forms which are identically zero). Hence we may assume that L_1 and L_2 are linearly independent, and this guarantees that $T_\gamma(X)$ is a $(r-1)$-dimensional polytope. The inequalities defining $T_\gamma(X)$ are such that

$$\mu_{r-1}\left(T_\gamma(X)\right) = X^{r-1}\mu_{r-1}\left(T_{\frac{\gamma}{X}}(1)\right) \tag{3.4}$$

with μ_{r-1} denoting $(r-1)$-dimensional Lebesgue measure. Therefore we want to calculate the $(r-1)$-dimensional volume of $T_\gamma(1)$ for small γ. Again because L_1 and L_2 are independent, this volume is a differentiable function of γ in the neighbourhood of $\gamma = 0$:

$$\mu_{r-1}\left(T_\gamma(1)\right) = \mu_{r-1}\left(T_0(1)\right) + O(\gamma). \tag{3.5}$$

Now substitute γ/X for γ in (3.5) and put this into (3.4). Integrating γ over $0 < \gamma \leq C$ gives the required estimate. ∎

Lemma 3.3. *For every choice of $L(\underline{x})$ and $L^*(\underline{x})$, the linear form $L - L^*$ has the property that at least two of its coefficients are linearly independent over* $\mathbf{Q}$.

Proof. Firstly, do the case where $t = 1$. The linear forms L and L^* correspond to embeddings σ and σ^* of the field M_1. The identification of complex conjugate embeddings make it sufficient to assume σ and σ^* differ on M_1^+, the maximal real subfield of M_1. If the allegation in Lemma 3.3 is false then $L - L^*$ is a real multiple of an integral linear form whose integral zeros are a lattice of rank $r-1$. There are finitely many lattices coming from units belonging to proper subfields of M_1^+ and the rank of each one is bounded by $\frac{r+1}{2} - 1$. This is strictly less than $r - 1$ because $1 < r$. There exists an integer vector $\underline{x}$ with $L(\underline{x}) = L^*(\underline{x})$ that does not belong to any of these lattices. This vector $\underline{x}$ corresponds to a unit of M_1^+ which does not lie in any proper subfield of M_1^+. Thus σ and σ^* agree on this unit and hence on M_1^+, a contradiction. The general case is entirely similar. Now σ and σ^* correspond to vectors of embeddings. Assuming they differ on one component, we can use the equation $L(\underline{x}) = L^*(\underline{x})$ to find a unit $u \in U$ upon which σ and σ^* agree on every component, a contradiction. ■

Proof of Proposition 3.1. The region $\mathfrak{S}_C(X)$ is defined by inequalities involving finitely many linear forms. Thus the boundary consists of a finite union of hyperplanes. Write $\Delta\mathfrak{S}_C(X)$ for the boundary of the region. For lattice points $\underline{x} \in \mathbf{Z}^r$, write $C_{\underline{x}}$ for the unit ball centred at $\underline{x}$. Let $\mathbf{Z}_C(X)$ denote the lattice points $\underline{x} \in \mathbf{Z}^r$ such that $C_{\underline{x}}$ has non-empty intersection with $\Delta\mathfrak{S}_C(X)$. Write

$$S_1 = \sum_{C_{\underline{x}} \subset \mathfrak{S}_C(X)} 1, \tag{3.6}$$

$$S_2 = \sum_{\underline{x} \in Z_C(X) \cap \mathfrak{S}_C(X)} \mu_r(C_{\underline{x}} \cap \mathfrak{S}_C(X)), \tag{3.7}$$

$$S_3 = \sum_{\underline{x} \in Z_C(X) - \mathfrak{S}_C(X)} \mu_r(C_{\underline{x}} \cap \mathfrak{S}_C(X)). \tag{3.8}$$

The volume in Lemma 3.2 decomposes as follows;

$$\mu_r(\mathfrak{S}_C(X)) = S_1 + S_2 + S_3. \tag{3.9}$$

In S_2, the boundary conditions guarantee that the distances between the $\underline{x}$ and the boundary are uniformly distributed. We sum the values of a continuous function of those distances. The function clearly has integral $1/2$. Similar remarks hold for the sum S_3. From the theory of uniform distribution (see [KN]) and the symmetry, each of S_2 and S_3 is

$$\frac{1}{2}\left(\sum_{\underline{x} \in Z_C(X) \cap \mathfrak{S}_C(X)} 1\right) + o(X^{r-1}). \tag{3.10}$$

Thus, (3.6), (3.9) and (3.10) give

$$\mu_r(\mathfrak{S}_C(X)) = \sum_{\underline{x} \in \mathfrak{S}_C(X)} 1 + o(X^{r-1}),$$

which proves Proposition 3.1. ■

§4 Computational Results

In this section, we will present some pictures to illustrate the clustering phenomena in §1 for some decomposable form equations in a small number of variables. Examples in both the CM case and non-CM case are included. For generating the pictures, we used the software packages KANT from the TU Berlin ([Ka]) for the calculations of number field data and Maple for plotting. Note that the norm used for projecting solutions is always the max norm.

We begin with the simplest case of a norm form equation, Pell's equation from the introduction. Taking $d = 2$ gives the equation

$$x_1^2 - 2x_2^2 = 1. \tag{P_2}$$

The field is $K_1 = \mathbf{Q}(\sqrt{2})$ and the solutions (x_1, x_2) correspond to units of norm 1 in the ring $\mathbf{Z}[\sqrt{2}]$ via $(x_1, x_2) \longleftrightarrow x_1 + x_2\sqrt{2}$. Plotting the first 8 solutions $(x_1, x_2) \in \mathbf{Z}^2$ then projecting centrally gives figure 1. The solutions all lie on a hyperbola with asymptotes $x_1 = \pm\sqrt{2}x_2$ so the distribution is obvious. The solutions cluster densely around the four points $(\pm 1, \pm 1/\sqrt{2})$ which are shown on the unit square.

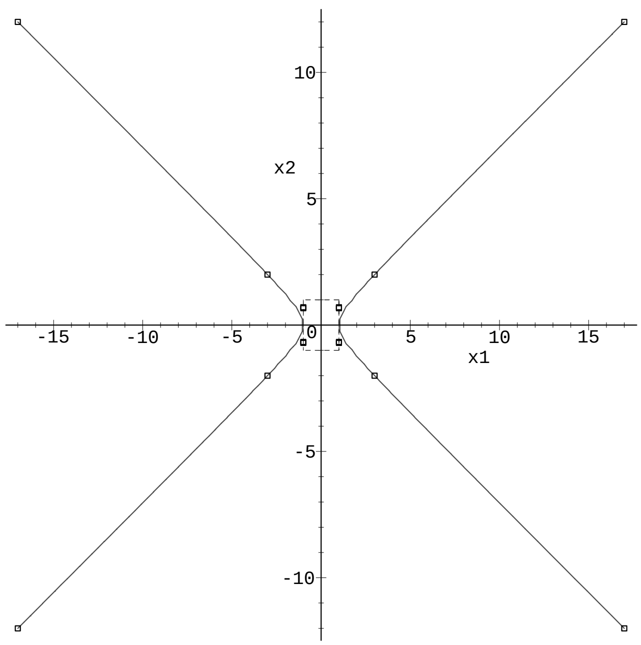

Figure 1. Projection of solutions of P_2

For cubics and quartics, we can still visualize the different types of behaviour. Consider the norm form equation for the totally real cubic K_2 of discriminant 49, which is the maximal real subfield of the cyclotomic number field generated by a primitive 7th root of unity ζ. As a field, K_2 is generated over $\mathbf{Q}$ by $\theta = \zeta + \zeta^6$, the minimal polynomial of θ is $x^3 + x^2 - 2x - 1$, and the ring of integers is $\mathbf{Z}[\theta]$. Choosing $1, \theta, \theta^2$ as the basis of $\mathbf{Z}[\theta]$, every unit of norm 1

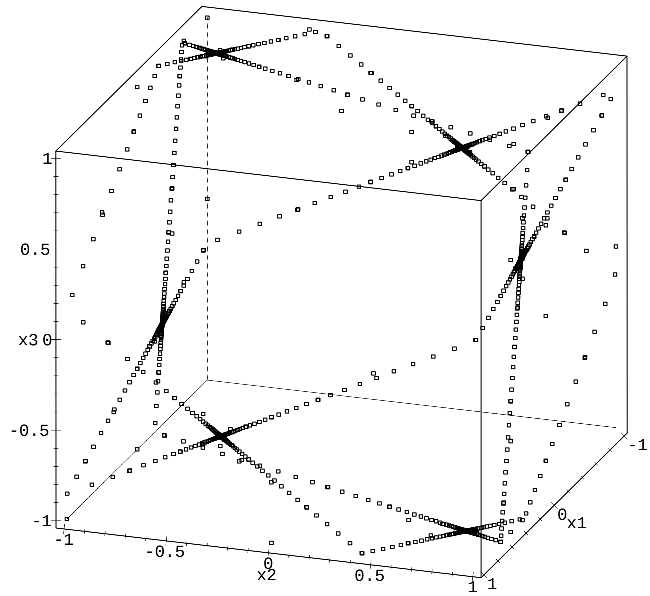

FIGURE 2. Projection of solutions of N_2

corresponds to a solution $\underline{x} \in \mathbf{Z}^3$ of the norm form equation

$$x_1^3 + x_2^3 + x_3^3 - x_1^2 x_2 + 5x_1^2 x_3 - 2x_1 x_2^2 + 6x_1 x_3^2 - x_2^2 x_3 - 2x_2 x_3^2 - x_1 x_2 x_3 = 1. \qquad (N_2)$$

After central projection, those solutions look like figure 2. In this example, there are four maximal real families of solutions, each contributing to three of the six cluster points. Theorem 2 describes precisely the distribution of points close to the cluster lines.

For the third example, take $K_3 = \mathbf{Q}(2^{\frac{1}{3}})$ which has ring of integers $\mathbf{Z}[2^{\frac{1}{3}}]$. Choosing the basis $1, 2^{\frac{1}{3}}, 2^{\frac{2}{3}}$ for the ring of integers gives the non-CM, norm form equation

$$x_1^3 + 2x_2^3 + 4x_3^3 - 6x_1 x_2 x_3 = 1. \qquad (N_3)$$

Figure 3 shows the distribution of the projected solutions; essentially a line and two isolated cluster points. Due to the limited resolution, many images are printed on top of each other - out of 800 points in the whole picture, 261 are closer than 0.01 in distance to the isolated cluster points! Note that this time, the points around the cluster line are uniformly distributed. Formula (1.4) holds with V consisting of the union of a line and two points.

Next come two quartic cases, one is CM and the other is not. For a totally real quartic, take $K_4 = \mathbf{Q}(\alpha)$, where α is a root of $x^4 - 2x^3 + 3x + 2$. Again, $\mathbf{Z}[\alpha]$ is the ring of integers of K_4, and we choose the powers of α as basis. Then we consider the norm form equation

$$N_{K_4|\mathbf{Q}}(x_1 + x_2\alpha + x_3\alpha^2 + x_4\alpha^3) = 1. \qquad (N_4)$$

Figure 4A shows one face (a 3-dimensional cube) of the corresponding 4-dimensional unit ball with the projection of the smallest 700 solutions with respect to Euclidean norm. Depending

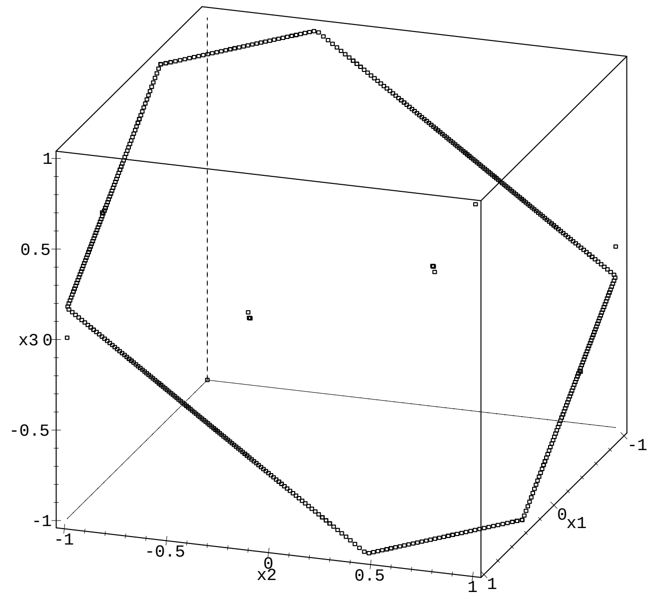

FIGURE 3. Projection of solutions of N_3

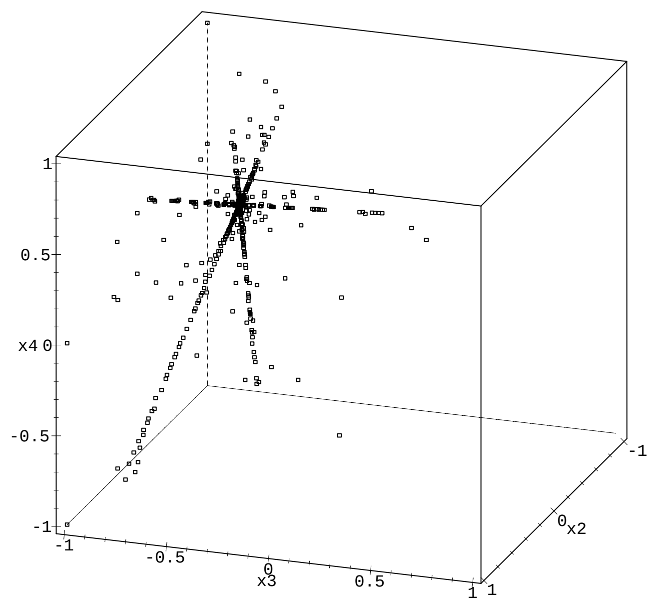

FIGURE 4A. Projection of solutions of N_4

on the choice of basis, some faces can have more than one cluster point, or none at all. Figure 4B shows another face, with the projection of 4217 solutions and two cluster points. It is clearly visible that the images of solutions cluster densely around the cluster lines, and yet more densely towards the cluster points. Once again, Theorem 2 goes beyond this in describing this phenomenon quantitatively.

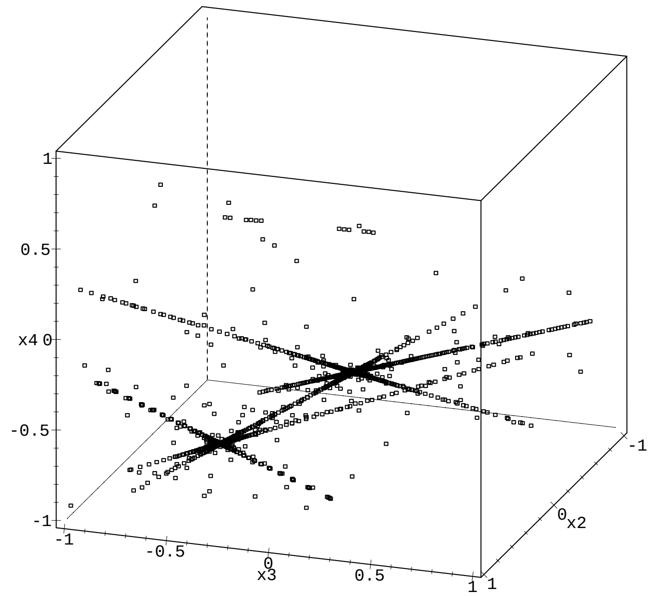

FIGURE 4B. Projection of solutions of N_4 - face 2

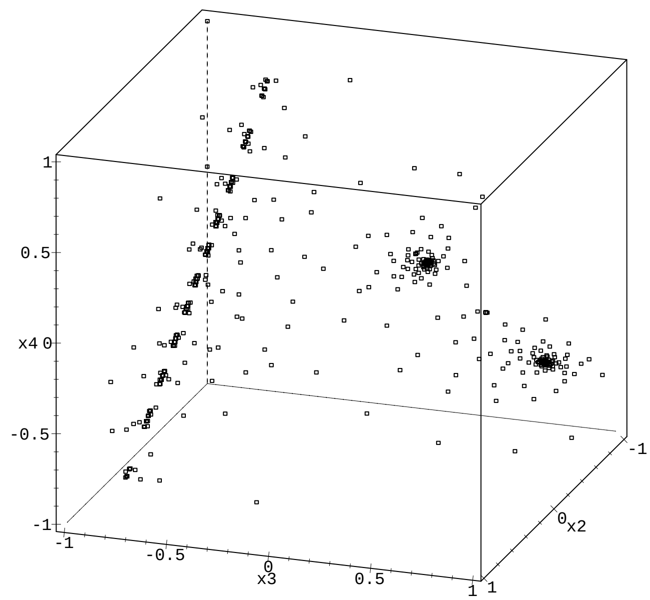

FIGURE 5A. Projection of solutions of N_5

For the next example, let K_5 denote the field $K_5 = \mathbf{Q}(2^{\frac{1}{4}})$. The ring of integers here is $\mathbf{Z}[2^{\frac{1}{4}}]$, and we have chosen the powers of $2^{\frac{1}{4}}$ as basis. The norm form equation is

$$N_{K_5|\mathbf{Q}}(x_1 + x_2 2^{\frac{1}{4}} + x_3 2^{\frac{2}{4}} + x_4 2^{\frac{3}{4}}) = 1. \tag{N_5}$$

There are 3000 solutions represented in figure 5A. This example is not CM and the clustering behaviour is different. Formula (1.4) holds with the actual V consisting two cluster points

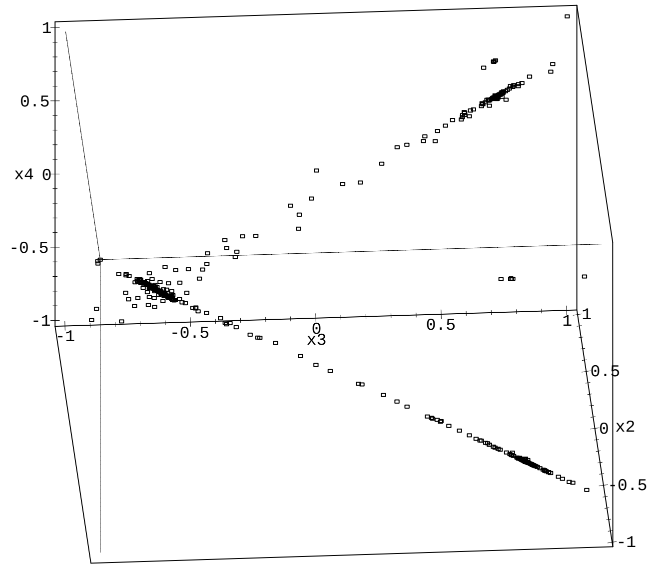

FIGURE 5B. The same face as Figure 5A, tilted upwards

(both of which are captured on the face shown) and one cluster line which appears very curiously 'dashed'. However, it is possible to show that the distribution around this line is in fact uniform. This line corresponds to units where the complex conjugates dominate in absolute value. The points correspond to units where one of the real conjugates dominates. Another new phenomenon appears when we tilt the picture upwards - see figure 5B. The solutions are nearly all confined to two 'cluster planes' determined by the cluster line and one of the cluster points. If we tilted the face a little bit more, we could shrink the cluster line to a point on the paper, and the cluster planes would appear as lines. Between the two cluster points, there lie second order cluster points which are discrete, having the first order cluster points as limit points. These are barely recognisable from the picture because of their proximity to the limit points. Formula (1.7) holds with the actual W consisting of finitely many planes and a discrete set of points which lie on finitely many lines.

The last two examples are decomposable form equations which are the product of two Pellian equations (see (1.6)). Firstly

$$(x_1^2 - 2x_2^2)(x_3^2 - 3x_4^2) = 1. \qquad (P_{2,3})$$

Figure 6 shows one 3-dimensional face of the 4-dimensional unit cube containing the projections of 4121 solutions which belong to two different families. Each family contributes two V-shapes. In many respects, the distribution is the same as in the 2nd and 4th examples. Formulae (1.2) and (1.5) do not hold but (1.7) holds with V a finite set of points and the actual W consisting of a finite union of lines. Also, formula (1.11) holds for this example.

Now consider the equation

$$(x_1^2 - 3x_2^2)(x_3^2 - 3x_4^2) = 1. \qquad (P_{3,3})$$

Figure 7 shows one 3-dimensional face of the 4-dimensional cube. The distribution is markedly different. The projections of some 4000 solutions are on view but the picture appears to

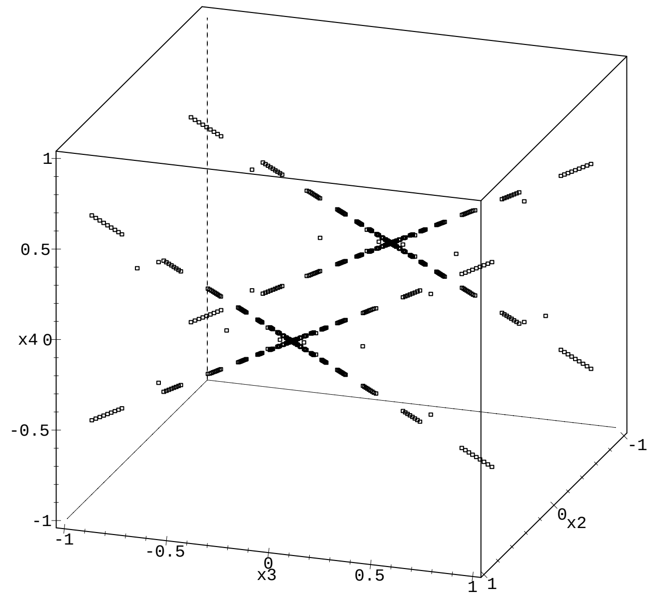

FIGURE 6. Projection of solutions of $P_{2,3}$

contain far fewer points due to the limited resolution. The solutions belong to two families, each one contributing two V-shapes. Formulae (1.2) and (1.5) do not hold but (1.7) holds with V a finite set of points. In contrast to the previous example, the actual W is an infinite set of points which lie on finitely many lines and have V as the only limit points. Also, formula (1.11) does not hold for this example.

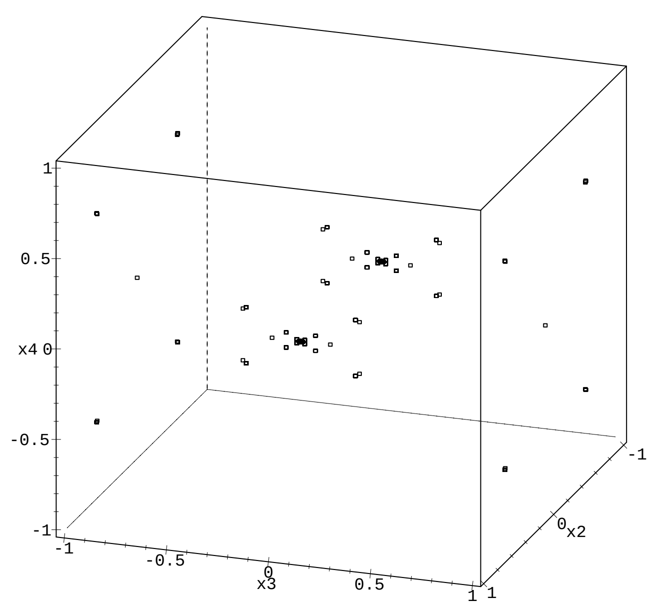

FIGURE 7. Projection of solutions of $P_{3,3}$

References

[E1] G. R. Everest, *Angular distribution of units in abelian group rings-an application to Galois-module theory*, J. reine angew. Math. (1987), 24-41.

[E2] G. R. Everest, *Units in abelian group rings and meromorphic functions*, Illinois J. Math. **33** (1989), 542-553.

[EG] G. R. Everest, K. Győry, *Counting solutions of decomposable form equations*, Acta Arith. **79** (1997), 173-191.

[EvG] J.-H. Evertse, K. Győry, *The number of families of solutions of decomposable form equations*, Acta Arith. **80** (1997), 367-394.

[G] K. Győry, *On the number of families of solutions of systems of decomposable form equations*, Publ. Math. Debrecen **22** (1993), 65-101.

[K] G. Karpilovsky, *Unit Groups of Classical Rings*, Oxford University Press, 1988.

[Ka] M. Daberkow, C. Fieker, J. Klüners, M. Pohst, K. Roegner and K. Wildanger, *KANT V4*, J. Symbolic Comp. **24** (1997), 267-283.

[KN] L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences*, Wiley, New York, 1974.

[Sc1] W. Schmidt, *Norm form equations*, Ann. of Math. **96** (1972), 526-551.

[Sc2] W. Schmidt, *Diophantine Approximation, Lecture Notes in Mathematics vol. 785*, Springer, 1980.

[Sc3] W. Schmidt, *Diophantine Approximation and Diophantine Equations, Lecture Notes in Mathematics vol. 1467*, Springer, 1991.

[Se] S. Sehgal, *Topics in Group Rings*, Dekker, New York 1978.

(Everest and Röttger) School of Mathematics, University of East Anglia, Norwich, Norfolk NR4 7TJ. UK g.everest@uea.ac.uk

(Gaal and Győry) Institute of Mathematics and Informatics, Lajos Kossuth University, H-4010 Debrecen, Pf 12, Hungary gyory@math.klte.hu