Exceptional sets in Kolmogorov-Arnol'd-Moser theory

1986 J. Phys. A: Math. Gen. 19 349

(http://iopscience.iop.org/0305-4470/19/3/018)

J. Phys. A: Math. Gen. 19 (1986) 349-374. Printed in Great Britain

Exceptional sets in Kolmogorov–Arnol'd–Moser theory

M M Dodson† and J A G Vickers‡

† Department of Mathematics, University of York, Heslington, York YO1 5DD, UK
‡ Department of Mathematics, University of Southampton, Highfield, Southampton SO9 5NH, UK

Received 11 January 1985, in final form 7 June 1985

Abstract. The exceptional sets in the autonomous and periodic KAM theorems consist of sets of near-resonant frequencies $(\omega_1, \ldots, \omega_n) = \boldsymbol{\omega}$ for which the method of proof of the theorems breaks down. They are residual and of measure 0. In this paper their Hausdorff dimension is deduced from that of a related set of well approximable linear forms. The Hausdorff dimension of these linear forms is a special case of a more general number theoretic result on systems of linear forms established recently but a self-contained proof is given. The proof relies on the resonant hyperplanes being reasonably well distributed. A fairly general statistical 'second moment' argument is used but some geometrical ideas are introduced to simplify and improve the original proof. When the number of degrees of freedom is large, the Hausdorff dimension is nearly maximal, so that, although of measure 0, the exceptional sets are, roughly speaking, close to sets of positive Lebesgue measure. The implications for the stability of Hamiltonian systems with many degrees of freedom and for the onset of certain kinds of instability are discussed briefly.

1. Introduction

One of the oldest problems in mechanics has been to understand the behaviour of solutions for the motions of N bodies subject to Newtonian attraction. A case of particular importance is the Solar System, in which one of the masses m_N (the Sun) is much larger than the masses m_j, $j = 1, \ldots, N-1$, of the other bodies (the planets), so that the mass ratios m_j/m_N are small. If, as a first approximation, the centre of mass of the system is supposed to coincide with the Sun and any interactions between the planets are ignored, then the system decouples into $N-1$ equations, whose solutions may be chosen to be elliptical orbits about the Sun with frequencies ω_j, $j = 1, \ldots, N-1$. Such a solution for the whole system is said to be quasi-periodic. It is then of interest to know whether such solutions still exist (for small enough m_j/m_N) when the interactions are taken into account. A formal proof of the existence of such quasi-periodic solutions was known to Weierstrass (see Moser 1973, ch 1, § 2) who was able to construct series solutions but, owing to the presence of small divisors, was unable to establish their convergence. Indeed it appears that, because of doubts about the convergence of his series expansions, Poincaré suspected that quasi-periodic solutions did not exist (see Moser 1973, ch 1, § 2). The problem was not answered satisfactorily until 1962, when Arnol'd not only established in the Kolmogorov-Arnol'd-Moser (KAM) theorem the existence of quasi-periodic solutions for the more general problem of a perturbed integrable Hamiltonian system but also showed that the set of such solutions forms a complicated Cantor-type set of positive Lebesgue measure (see Moser 1973, p 8). This implied that for planets of sufficiently small mass compared to the

Sun and for the majority of initial conditions under which the instantaneous orbits are close to coplanar circles, the perturbations of the planets on each other cause little change over all time. It follows that there is a set of initial conditions, which has positive Lebesgue measure, such that the distances of the bodies from each other will remain perpetually bounded if the initial conditions belong to this set (Arnol'd 1963, p 125).

The magnitude of a permissible perturbation of the system depends upon how closely integer linear forms in the frequencies approach 0 or how nearly the frequencies are rationally dependent or resonant. When the frequencies are very close to being resonant, problems associated with so-called 'small divisors' arise and the method used to deduce the existence of quasi-periodic solutions fails. To avoid these problems, a Diophantine approximation condition, which controls the closeness to resonance and which is purely number theoretic in character, is imposed on the frequencies. This condition, which is governed by an exponent τ, holds almost everywhere for suitable τ and frequencies which do not satisfy this condition, i.e. which are very close to being resonant, form an exceptional set of Lebesgue measure 0. In the case of these frequencies, the presence of small divisors prevents convergence in an extension of Newton's tangent method, which lies at the heart of the KAM theorem, being established. By considering the KAM theorem in the setting of the infinite dimensional group of canonical transformations acting on the infinite dimensional manifold of Hamiltonian vector fields, the Diophantine approximation condition emerges as the natural requirement for the application of an infinite dimensional implicit function theorem (see Vickers and Dodson (1985) for more details).

However, even when a set of frequencies satisfies the Diophantine approximation condition, so that when perturbed the corresponding system still has quasi-periodic solutions, the permissible perturbations for certain frequencies are so small that they are physically meaningless. Hence to prevent the perturbation of the Hamiltonian system having to be very small (or to ensure a reasonably 'robust' stability), the exponent τ is fixed at a convenient value for which the Diophantine approximation condition holds for almost all frequencies, so that the complementary set of frequencies not satisfying the condition is of measure 0 and so may be neglected.

In this paper, the complementary set and related exceptional sets, including the corresponding sets in the periodic KAM theorem, are studied and their Hausdorff dimensions determined. The Hausdorff dimension of the exceptional sets associated with the (autonomous) KAM theorem is obtained from that of the corresponding sets in the periodic KAM theorem. The dimension in the periodic case is determined using fairly general geometrical and statistical arguments which rely on the resonant hyperplanes being reasonably well distributed. It turns out that the cost of excluding frequencies for which the permissible perturbations are very small is that the complementary exceptional sets are relatively large and that their Hausdorff dimensions are almost maximal when there are many degrees of freedom. This means that, even though the exceptional sets are of Lebesgue measure zero, they are close to sets of positive Lebesgue measure.

The rest of this paper is organised as follows: in § 2 the proof of the autonomous KAM theorem is sketched to indicate how the number theoretic conditions emerge and the corresponding exceptional sets E and $E(\tau)$ are explained. In § 3 the corresponding conditions and exceptional sets $\hat{E}$ and $\hat{E}(\tau)$ which arise in the periodic KAM theorem are set out. In § 4, well approximable forms $\hat{W}(\tau)$ and $W(\tau)$ are introduced and their relationship with the exceptional sets established. The Hausdorff dimension dim X of a set X in $\mathbb{R}^n$ is explained in § 5 and upward inequalities for dim $W(\tau)$ and dim $\hat{W}(\tau)$ obtained in § 6 and § 7, respectively. The main and most difficult part of the paper is § 8 where dim $\hat{W}(\tau)$ is determined. The dimension of $W(\tau)$ is deduced in § 9 and dim $\hat{E}(\tau)$ and dim $E(\tau)$ obtained in § 10.

2. The Kolmogorov–Arnol'd–Moser theorem

A Hamiltonian system

$$\dot{x}_k = \partial H/\partial y_k \qquad \dot{y}_k = -\partial H/\partial x_k \qquad k = 1, \ldots, n \tag{2.1}$$

is called integrable if coordinates $q_1, \ldots, q_n, p_1, \ldots, p_n$ can be introduced via an (invertible) canonical transformation

$$(\boldsymbol{x}, \boldsymbol{y}) = W(\boldsymbol{q}, \boldsymbol{p})$$

where W has period 2π in the q_i, such that the Hamiltonian $H = H(\boldsymbol{p})$ is independent of $\boldsymbol{q}$. Such coordinates are called *action* and *angle* variables and imply the existence of a *global* solution to the Hamilton–Jacobi equations. In these coordinates, Hamilton's equations take the form

$$\dot{q}_k = \partial H/\partial q_k \qquad \dot{p}_k = 0, \; k = 1, \ldots, n$$

and have the general solution

$$q_k(t) = \frac{\partial H}{\partial p_k}(c)t + q_k(0) \qquad p_k(t) = p_k(0) = c_k \text{ (say)} \qquad k = 1, \ldots, n.$$

We now consider a perturbation of the integrable system H: $T^n \times \mathbb{R}^n \to \mathbb{R}$ (T^n is the n-dimensional torus $\{(z_1, \ldots, z_n) \in \mathbb{C}^n : |z_1| = \ldots = |z_n| = 1\}$) given by

$$H'(\boldsymbol{q}, \boldsymbol{p}; \mu) = H_0(\boldsymbol{p}) + \mu H_1(\boldsymbol{q}, \boldsymbol{p}) + O(\mu^2) \tag{2.2}$$

where μ is a small parameter and where H' is assumed to be analytic in $2n+1$ variables, with period 2π in the q_i. Thus for $\mu = 0$ the $2n$-dimensional phase space is foliated into an n-parameter family of tori

$$p_k(t) = c_k = \text{constant} \qquad k = 1, \ldots, n$$

or $\boldsymbol{p}(t) = \boldsymbol{c}$, on which the flow is given by

$$\dot{q}_k = \frac{\partial H_0}{\partial p_k}(c) = \omega_k \text{ (say)} \qquad k = 1, \ldots, n.$$

By (2.2), $H'(\boldsymbol{q}, \boldsymbol{p}; 0) = H_0(\boldsymbol{p})$. It turns out that in order to investigate orbits which persist under small perturbations (or to prove the KAM theorem), it is not necessary to consider a general Hamiltonian $H_0(\boldsymbol{p})$ but it suffices to consider Hamiltonians which are at most quadratic in the p_i and satisfy

$$\partial H_0/\partial p_j = \omega_j + \sum_{k=1}^{n} a_{jk} p_k$$

for some invertible symmetric matrix (a_{jk}) (see Moser 1973, p 115). If we are given such a Hamiltonian H_0 and we are able to introduce new canonical coordinates $\boldsymbol{q}'$, $\boldsymbol{p}'$ so that Hamilton's equations become

$$\dot{q}'_j = \partial H'/\partial p'_j \qquad \dot{p}'_j = -\partial H'/\partial q'_j$$

where

$$\partial H'/\partial p_j' = \omega_j + \sum_k b_{jk} p_k' + O(p'^2)$$

$(b_{jk} = b_{kj})$ and $\partial H'/\partial q_j' = O(p'^2)$, then a solution to Hamilton's equation's for H' with initial conditions $p_j'(0) = 0$ is provided by

$$q_j'(t) = \omega_j t + q_j'(0) \qquad p_j'(t) = 0,\ k = 1, \ldots, n.$$

In terms of the original coordinates $\boldsymbol{q}, \boldsymbol{p}$, this gives rise to a quasi-periodic solution

$$(\boldsymbol{q}(t), \boldsymbol{p}(t)) = W'(\boldsymbol{\omega} t + \boldsymbol{q}'(0), 0)$$

of the original equations.

Since H' is of period 2π in the q_i, H' can be regarded as defining a Hamiltonian vector field

$$m = \sum_{j=1}^{n} \frac{\partial H'}{\partial p_j} \frac{\partial}{\partial q_j} - \frac{\partial H'}{\partial q_j} \frac{\partial}{\partial p_j}$$

on $D = T^n \times P$, where P is an open domain in $\mathbb{R}^n$. Being able to find coordinates p' and q' is equivalent to showing that for an element m in the space M of real analytic Hamiltonian vector fields on D sufficiently near the vector field

$$m^* = \sum_j \left(\omega_j + \sum_k a_{jk} p_k \right) \partial/\partial q_j$$

there exists an element $g = g(m)$ in G, the (local) group of real analytic homogeneous canonical transformations on D, such that

$$(g^{*-1} \circ m \circ g)_1 = (m^*)_1 + d_0 \tag{2.3}$$

where

$$d_0 \in \mathrm{D} = \left\{ d = \sum_j \sum_k c_{jk} p_k\, \partial/\partial q_j \colon c_{jk} = c_{kj} \right\}$$

and $(\ )_1$ denotes linearisation with respect to the p variables, i.e.

$$(m(\boldsymbol{q}, \boldsymbol{p}))_1 = m(\boldsymbol{q}, 0) + \sum_k p_k \frac{\partial m}{\partial p_k}(\boldsymbol{q}, 0).$$

Let M_1 be the set of p-linearised Hamiltonian vector fields. Then we can define an action $\mathrm{G} \times M_1 \to M_1$ by

$$(g, m) \to g \cdot m = (g^{*-1} \circ m \circ g)_1.$$

The KAM theorem follows when (2.3) can be solved. The group action $g \cdot m$ induces a map $\phi : \mathscr{L} \to \tilde{M}_1$, given by

$$\phi : x \to (L_x m^*)_1$$

where $\mathscr{L}$ is the Lie algebra of G, $\tilde{M}_1 = T_{m^*} M_1$ and $L_x m^*$ is the Lie derivative. Note that since the vector field is linearised with respect to the p variables, it is enough to consider those elements of $\mathscr{L}$ which are linear in p. The linearised version of (2.3) is

$$\phi(x) + d_0 = (L_x m^*)_1 + d_0 = m_1 \tag{2.4}$$

where $x \in \mathscr{L}$ and $m_1 \in \tilde{M}_1$. By considering the coefficients of the components of the vector m_1 and splitting them up into terms independent of p or linear in p, (2.4) gives a system of independent scalar equations, each of the form

$$\partial U = \sum_k \omega_k \, \partial U/\partial q_k = f(\boldsymbol{q}) \tag{2.5}$$

where f is 2π-periodic in q and has constant term 0. It can be verified that the functions $e(\boldsymbol{j})$, $\boldsymbol{j} \in \mathbb{Z}^n$, where $e(\boldsymbol{j})(\boldsymbol{q}) = \exp(2\pi \mathrm{i} \boldsymbol{j} \cdot \boldsymbol{q})$, form a complete set of eigenfunctions for ∂ with eigenvalues $\boldsymbol{\omega} \cdot \boldsymbol{j}$. When expanded in terms of these eigenfunctions, the Fourier series for f is

$$f = \sum_{j \neq 0} f_j e(\boldsymbol{j})$$

and we can solve (2.5) with

$$U = \frac{1}{2\pi \mathrm{i}} \sum_{j \neq 0} f_j e(\boldsymbol{j})/(\boldsymbol{\omega} \cdot \boldsymbol{j}). \tag{2.6}$$

For the solution (2.6) to exist, $\boldsymbol{\omega} \cdot \boldsymbol{j}$ cannot vanish for non-zero j. But for any $\boldsymbol{\omega}$, the denominator $\boldsymbol{\omega} \cdot \boldsymbol{j}$ can be arbitrarily small for certain j and this problem of 'small divisors' makes convergence problematic, despite the exponential convergence to zero of the Fourier coefficients f_j.

One method of solving (2.3) is to use an iterative method based on Newton's tangent method and which involves repeatedly solving (2.4) (for more details see Arnol'd 1963, Moser 1973, Sternberg 1969, Vickers and Dodson 1985, Zehnder 1977). This method can be shown to converge and the KAM theorem established if the map ϕ satisfies a condition known as finite order (see Arnol'd 1968, p 27, Arnol'd 1981). It can be shown that if for some real τ and positive C, $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)$ satisfies the purely arithmetical condition

$$|\boldsymbol{\omega} \cdot \boldsymbol{j}| \geq C |\boldsymbol{j}|_1^{-\tau} \tag{2.7}$$

where $|\boldsymbol{j}|_1 = |j_1| + \ldots + |j_n|$, for all non-zero $\boldsymbol{j}$ in $\mathbb{Z}^n$, then ϕ is of finite order. Moreover this condition guarantees the convergence of the series (2.6) and so guarantees (under a natural non-degeneracy requirement) the KAM theorem. Indeed the full statement of the KAM theorem is as follows (see Moser 1973, p 44).

Let Y be an open set in $\mathbb{R}^n$ and let $H'(\boldsymbol{q}, \boldsymbol{p}, \mu)$ be a real analytic function of $\boldsymbol{q}$, $\boldsymbol{p}$ and μ for all $\boldsymbol{q}$, $\boldsymbol{p}$ in $\mathbb{R}^n$ and all μ near 0. Suppose H has period 2π in $q_1, \ldots, q_n$ and that $H_0 = H(\boldsymbol{q}, \boldsymbol{p}, 0)$ is independent of $\boldsymbol{q}$ (so that $H_0(\boldsymbol{q}, \boldsymbol{p}) = H_0(\boldsymbol{p})$). Let $\boldsymbol{c}$ in $\mathbb{R}^n$ be chosen so that the frequencies $\omega_1, \ldots, \omega_n$ given by

$$\omega_k = \partial H_0(c)/\partial p_k \qquad k = 1, \ldots, n \tag{2.8}$$

satisfy (2.7) for some τ and positive C and all non-zero j in $\mathbb{Z}^n$, and so that the Hessian

$$\det(\partial \omega_k / \partial p_j) = \det(\partial^2 H_0 / \partial p_j \, \partial p_k) \tag{2.9}$$

does not vanish at $\boldsymbol{p} = \boldsymbol{c}$.

Then there exists a positive μ_0 such that, for all μ with $|\mu| < \mu_0$, there exists an invariant torus

$$\boldsymbol{q} = \boldsymbol{\theta} + \boldsymbol{u}(\boldsymbol{\theta}, \mu) \qquad \boldsymbol{p} = \boldsymbol{c} + \boldsymbol{v}(\boldsymbol{\theta}, \mu) \tag{2.10}$$

where $u_i(\boldsymbol{\theta}, \mu)$ and $v_i(\boldsymbol{\theta}, \mu)$ are real analytic functions of μ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$, have period 2π in $\theta_1, \ldots, \theta_n$ and vanish for $\mu = 0$. Moreover the flow on this torus is given by

$$\dot{\theta}_k = \omega_k \qquad 1 \leqslant k \leqslant n.$$

This equation defines a subsystem of the system given by (2.2) and (2.10) defines the embedding of the torus.

Since the arithmetic condition (2.7) holds for almost all $\boldsymbol{\omega}$ in $\mathbb{R}^n$, the KAM theorem holds for almost all $\boldsymbol{\omega}$. Note that μ_0 depends on the exponent τ in (2.7); the smaller τ, the bigger the perturbation μ can be (Russman (1973) contains estimates connected with this question). Physically only perturbations which are not too small and correspond to 'robust' stability are of interest. If τ can be arbitrarily large, as is permitted in the mathematical definition, the allowable perturbations can become arbitrarily small and the stability can be considered as delicate.

We are going to study the set Ω of points $\boldsymbol{\omega}$ in $\mathbb{R}^n$ for which (2.7) holds, i.e. the set of frequencies for which the KAM theorems can be proved, and its complement, the set of frequencies for which the proof fails. To do this, it is convenient to introduce the auxiliary τ-stability set:

$$\Omega(\tau) = \{\boldsymbol{x} \in \mathbb{R}^n\text{: for some } C > 0,\ |\boldsymbol{q} \cdot \boldsymbol{x}| \geqslant C|\boldsymbol{q}|_1^{-\tau} \text{ for all non-zero } \boldsymbol{q} \in \mathbb{Z}^n\}$$

so that $\Omega(\tau') \supseteq \Omega(\tau)$ when $\tau' \geqslant \tau$ and $\Omega = \bigcup_{\tau \in \mathbb{R}} \Omega(\tau)$.

It follows from a Dirichlet box argument that $\Omega(\tau)$ is empty when $\tau < n - 1$ and it can be shown that $|\Omega(n-1)| = 0$ (and incidentally that the Hausdorff dimension of $\Omega(n-1)$ is n) and that $|\Omega(\tau)| = 1$ when $\tau > n - 1$ (Arnol'd 1963, Schmidt 1980, Rüssman 1973). Hence without loss of generality the union may be taken over all $\tau \geqslant n - 1$, and Ω has Lebesgue measure $|\Omega| = 1$. It follows that, provided H' is sufficiently near H_0, then almost all frequencies give rise to quasi-periodic solutions. It is worth repeating that the size of the allowable perturbation $H' - H_0$ depends upon the size of the exponent τ; the larger τ, the smaller the allowable perturbation. In order to study this aspect of the theory, we study the complements $E(\tau) = \mathbb{R}^n \backslash \Omega(\tau)$ of the constituent τ-stability sets $\Omega(\tau)$, i.e. we study the sets

$$E(\tau) = \{x \in \mathbb{R}^n\text{: for each } C > 0,\ |\boldsymbol{j} \cdot \boldsymbol{x}| < C|\boldsymbol{j}|_1^{-\tau} \text{ for some non-zero } \boldsymbol{j} \in \mathbb{Z}^n,\ 1 \leqslant k \leqslant n\}$$

for which the inequality (2.3) does not hold. Clearly $E(\tau') \subseteq E(\tau)$ when $\tau' \geqslant \tau$ and from the above when $\tau < n - 1$, $E(\tau) = \mathbb{R}^n$, $|E(n-1)| = 1$ and when $\tau > n - 1$, $|E(\tau)| = 0$. Clearly the set

$$E = \bigcap_{\tau} E(\tau) = \mathbb{R}^n \backslash \Omega$$

has Lebesgue measure 0 and so E and the sets $E(\tau)$ when $\tau > n - 1$ will be called exceptional. Note however that since the constant C in the definition of $E(\tau)$ can, without loss of generality, be taken to be a positive rational, it follows that

$$E(\tau) = \bigcap_{C \in \mathbb{Q}^+} \bigcup_{\substack{\boldsymbol{j} \in \mathbb{Z}^n \\ \boldsymbol{j} \neq 0}} \{\boldsymbol{x} \in \mathbb{R}^n\text{: } |\boldsymbol{j} \cdot \boldsymbol{x}| < C|\boldsymbol{j}|_1^{-\tau}\}$$

whence $E(\tau)$ is the countable intersection of open dense sets (containing $\mathbb{Q}^n$) and so is residual. Similarly since

$$E = \lim_{n \to \infty} \bigcap_{j=1}^{n} E(j)$$

E is a residual set, so that their respective complements $\Omega(\tau)$ and Ω are meagre, even though Ω and $\Omega(\tau)$ when $\tau > n-1$ have full measure. Note also that E and $E(\tau)$ are G_δ sets whence Ω and $\Omega(\tau)$ are F_σ sets.

Note that since the norm $|\ |_1$ and the supremum norm $|\ |$ on $\mathbb{R}^n$ given by

$$|v| = \max\{|v_1|, \ldots, |v_n|\}\ (=|v|_\infty)$$

are equivalent, both norms can be used in the definition of Ω, $\Omega(\tau)$ and $E(\tau)$; thus

$$E(\tau) = \{x \in \mathbb{R}^n\text{: for each } C>0,\ |j\cdot\omega| < C|j|^{-\tau} \text{ for some } j \in \mathbb{Z}^n\backslash\{0\}\}. \tag{2.11}$$

3. The periodic KAM theorem

A closely related version of the autonomous KAM theorem just discussed deals with the perturbation of a Hamiltonian of the form $H_0(p, t)$ which has period 2π in the time variable t. In the periodic KAM theorem, a real analytic Hamiltonian $H: T^{n+1} \times \mathbb{R}^n \to \mathbb{R}$ is perturbed to the Hamiltonian

$$H'(q, p, t; \mu) = H_0(p, t) + \mu H_1(q, p, t) + O(\mu^2)$$

which is real analytic in the $2n+2$ variables and has period 2π in q_i and t. The theorem asserts that, given some point c in P for which the point $\omega = (\omega_1, \ldots, \omega_n)$ (where the ω_k are given by (2.8) with the Hessian (2.9) not vanishing) satisfies the stronger condition that for some real τ and positive C,

$$|j\cdot\omega - m| \geq C|j|_1^{-\tau} \tag{3.1}$$

for all $m \in \mathbb{Z}$ and non-zero $j \in \mathbb{Z}^n$, then there exists a corresponding quasi-periodic solution to the original equations. Write

$$\begin{aligned}\hat\Omega(\tau) &= \{x \in \mathbb{R}^n\text{: for some } C>0,\ |j\cdot x - m| \geq C|j|_1^{-\tau}, j \in \mathbb{Z}^n\backslash\{0\}, m \in \mathbb{Z}\}\\ &= \{x \in \mathbb{R}^n\text{: for some } C>0,\ |j\cdot x - m| \geq C|j|^{-\tau}, j \in \mathbb{Z}^n\backslash\{0\}, m \in \mathbb{Z}\}\end{aligned}$$

so that if the vector ω of frequencies $\omega_1, \ldots, \omega_n$ lies in $\hat\Omega = \cup\, \hat\Omega(\tau)$, where the union is over all real τ, a quasi-periodic solution exists. Note that $\hat\Omega(\tau') \subseteq \hat\Omega(\tau)$ when $\tau' \geq \tau$.

Again it can be shown by using a Dirichlet box argument that when $\tau < n$, $\hat\Omega(\tau)$ is empty and it can also be shown that $|\hat\Omega(n)| = 0$ and that $|\hat\Omega(\tau)| = 1$ when $\tau > n$ (Arnol'd 1963, Schmidt 1980, Sprindžuk 1979). Thus (3.1) never holds when $\tau < n$, holds for a set of measure 0 when $\tau = n$ (incidentally the Hausdorff dimension $\dim \hat\Omega(n) = n$ (Schmidt 1980)) and for a set of full measure otherwise; and thus $|\hat\Omega| = 1$.

The exceptional set $\hat E(\tau) = \mathbb{R}^n\backslash\hat\Omega(\tau)$ of frequencies ω for which (3.1) does not hold for the exponent τ, so that the periodic KAM theorem cannot be deduced, is of the form

$$\begin{aligned}\hat E(\tau) &= \{x \in \mathbb{R}^n\text{: for each } C>0,\ |j\cdot x - m| < C|j|_1^{-\tau} \text{ for some } j \in \mathbb{Z}^n\backslash\{0\}, m \in \mathbb{Z}\}\\ &= \{x \in \mathbb{Z}^n\text{: for each } C>0,\ |\langle j\cdot x\rangle| < C|j|^{-\tau} \text{ for some } j \in \mathbb{Z}^n\backslash\{0\}\}\end{aligned} \tag{3.2}$$

where $\langle u\rangle$ is the unique number $u-m$, $m \in \mathbb{Z}$, in $(-\frac{1}{2}, \frac{1}{2}]$; thus

$$|\langle u\rangle| = \|u\| = \inf\{|u-m|: m \in \mathbb{Z}\}$$

the distance of u from the integer nearest to u.

Because $\hat{E}(\tau)=\mathbb{R}^n\backslash\hat{\Omega}(\tau)$, when $\tau<n$, $\hat{E}(\tau)=\mathbb{R}^n$, $|\hat{E}(n)|=1$ and when $\tau>n$, $|\hat{E}(\tau)|=0$. As with $E(\tau)$ and E,

$$\hat{E}(\tau)=\bigcap_{C\in\mathbb{Q}^+}\bigcup_{\substack{j\in\mathbb{Z}^n\\ j\neq 0}}\bigcup_{m\in\mathbb{Z}}\{x\in\mathbb{R}^n: |j\cdot x-m|<C|j|^{-\tau}\}\qquad \hat{E}=\lim_{n\to\infty}\bigcap_{j=1}^{n}\hat{E}(j)$$

are residual, so that their respective complements $\hat{\Omega}(\tau)$ and $\hat{\Omega}$ are meagre, even though $\hat{\Omega}$ and $\hat{\Omega}(\tau)$ when $\tau>n$ have full measure. Note also that $\hat{E}$ and $\hat{E}(\tau)$ are G_δ sets whence $\hat{\Omega}$ and $\hat{\Omega}(\tau)$ are F_σ sets. For large τ, vectors x in $\hat{E}(\tau)$ have coordinates $x_1,\ldots,x_n$ which are close to being integrally and hence rationally dependent with 1. The exceptional sets $\hat{E}(\tau)$ are closely related to sets of well approximable linear forms, as we shall see in the next section.

4. Exceptional sets and well approximable forms

The set of real linear forms $\boldsymbol{\xi}\cdot x=\Sigma_{j=1}^{n}\xi_j x_j$ satisfying

$$|\langle q\cdot x\rangle|=\|q\cdot x\|<|q|^{-\tau}$$

for infinitely many q in $\mathbb{Z}^n$ will be written

$$\hat{W}(\tau)=\{x\in\mathbb{R}^n: |\langle q\cdot x\rangle|<|q|^{-\tau} \text{ for infinitely many } q\in\mathbb{Z}^n\}\tag{4.1}$$

i.e. the form $\boldsymbol{\xi}\cdot x$ is identified with the vector x in $\mathbb{R}^n$. If a vector $x=(x_1,\ldots,x_n)$ is in $\hat{W}(\tau)$, $\tau>n$, the numbers $x_1,\ldots,x_n,1$ are close to being integrally (and hence rationally) dependent or resonant. The set $\hat{W}(\tau)$ depends on n and when it is desirable to make this dependence explicit, we shall write $\hat{W}(\tau)$ as $\hat{W}(\tau,n)$. When $\tau<n$, it follows from a box argument that $\hat{W}(\tau)=\mathbb{R}^n$. Since

$$\sum_q |q|^{-\tau}\begin{cases}=\infty & \text{when } \tau\leqslant n\\ <\infty & \text{when } \tau>n\end{cases}$$

where the sum is over all non-zero q in $\mathbb{Z}^n$, it follows from Groshev's generalisation of Khintchine's theorem (Sprindžuk 1979) that the Lebesgue measure $|\hat{W}(n)|$ of $\hat{W}(n)$ is 1 and that when $\tau>n$, $|\hat{W}(\tau)|=0$ (see also Schmidt 1980). The set $\hat{W}(\tau)$ and the exceptional set $\hat{E}(\tau)$ given by (3.2) and associated with the periodic KAM theorem are evidently closely related and both involve the linear form $\langle\boldsymbol{\xi}\cdot x\rangle$.

In the same way the exceptional set $E(\tau)$ given by (2.11) and associated with the autonomous KAM theorem is also evidently closely related to the set $W(\tau)$ of linear forms $\boldsymbol{\xi}\cdot x$ which satisfy

$$|q\cdot x|<|q|^{-\tau}\tag{4.2}$$

for infinitely many q in $\mathbb{Z}^n$, i.e. again identifying the form $\boldsymbol{\xi}\cdot x$ with x, $E(\tau)$ is closely related to the set

$$W(\tau)=\{x\in\mathbb{R}^n: |q\cdot x|<|q|^{-\tau} \text{ for infinitely many } q \text{ in } \mathbb{Z}^n\}.\tag{4.3}$$

As with $W(\tau)$, when the dependence on n needs to be made explicit, we will write $W(\tau)=W(\tau,n)$.

Although not equal to the sets of linear forms, the exceptional sets are included in the corresponding sets and contain slightly smaller versions of them.

Lemma 4.1. For each positive ε,

$$\hat{W}(\tau+\varepsilon)\subseteq \hat{E}(\tau)\subseteq \hat{W}(\tau) \qquad W(\tau+\varepsilon)\subseteq E(\tau)\subseteq W(\tau).$$

Proof. Let $x\in\mathbb{R}^n\setminus \hat{W}(\tau+\varepsilon)$, so that for all but finitely many q in $\mathbb{Z}^n$, $q_1,\dots,q_m$ say, $|\langle q\cdot x\rangle|\geq |q|^{-\tau}$. The constant $K=K(x)$ given by

$$K=\max\{|\langle q_r\cdot x\rangle|\cdot|q_r|^{-\tau}: r=1,\dots,m\}$$

is positive since if $\langle q_0, x\rangle=0$, then $\langle rq_0, x\rangle=0$ for each integer r, so that $\langle q\cdot x\rangle=0$ has infinitely many solutions in $\mathbb{Z}^n$. It follows that for all non-zero q in $\mathbb{Z}^n$

$$|\langle q\cdot x\rangle|>(K/2)|q|^{-\tau}$$

so that $x\in\mathbb{R}^n\setminus\hat{E}(\tau)=\hat{\Omega}(\tau)$, whence $\hat{E}(\tau)\subseteq\hat{W}(\tau)$.

Next let $x\in \hat{W}(\tau+\varepsilon)$, so that there exists an infinite sequence $\{q_r\in\mathbb{Z}^n: r=1,2,\dots\}$ satisfying $|\langle q_r\cdot x\rangle|<|q_r|^{-\tau-\varepsilon}$. Let C be any positive constant and choose r such that $|q_r|^{-\varepsilon}<C$. Then

$$|\langle q_r\cdot x\rangle<|q_r|^{-\tau-\varepsilon}<C|q_r|^{-\tau}$$

whence $x\in\hat{E}(\tau)$ and $\hat{W}(\tau+\varepsilon)\subset\hat{E}(\tau)$.

Since $q\cdot x=0$ implies that $rq\cdot x=0$ for each integer r, the result can also be shown to hold for $W(\tau)$ and $E(\tau)$.

Because of the close relationship between the linear forms $\langle \xi\cdot x\rangle$ and $\xi\cdot x$, the set $W(\tau,n)\cap\{x\in I^n: \frac14<x_1<\frac12\}$, where $I=(-\frac12,\frac12]$, can be mapped into $(\frac14,\frac12)\times(\hat{W}(\tau-\varepsilon,n-1)\cap I^{n-1})$ by a function T which satisfies a Lipschitz condition (so that the Hausdorff dimension of the domain of T is greater than that of the image, a fact which will be used in § 9). Moreover for each positive ε, $T(W(\tau,n)\cap\{x\in I^n: \frac14<x_1<\frac12\})$ contains $(\frac14,\frac12)\times(\hat{W}(\tau+\varepsilon,n-1)\cap I^{n-1})$.

Lemma 4.2. For each $\varepsilon>0$ and $n\geq 2$ the function T: $(\frac14,\frac12)\times I^{n-1}\to\mathbb{R}^n$ given by

$$T(x)=T(x_1,x_2,\dots,x_n)=(x_1,x_2/x_1,\dots,x_n/x_1)$$

sends $W(\tau,n)\cap\{x\in I^n:\frac14<x_1<\frac12\}$ into $(\frac14,\frac12)\times(\hat{W}(\tau-\varepsilon,n-1)\cap I^{n-1})$, the image $T(W(\tau,n)\cap\{x\in I^n:\frac14<x_1<\frac12\})$ contains $(\frac14,\frac12)\times(\hat{W}(\tau+\varepsilon,n-1)\cap I^{n-1})$ and T satisfies $|T(x)-T(x')|\leq 16|x-x'|$.

Proof. For each x in $W(\tau,n)\cap\{x\in I^n:\frac14<x_1<\frac12\}$, there are infinitely many q in $\mathbb{Z}^n$ such that $|q_1x_1+q_2x_2+\dots+q_nx_n|<|q|^{-\tau}$, and so infinitely many q such that $|q_1+q_2x_2/x_1+\dots+q_nx_n/x_1|<4|q|^{-\tau}<q^{-\tau+\varepsilon}$, where $q=\max\{|q_i|: 2\leq i\leq n\}>4^{1/\varepsilon}$. Hence, by definition,

$$T(x)=(x_1,x_2/x_1,\dots,x_n/x_1)\in(\tfrac14,\tfrac12)\times\hat{W}(\tau-\varepsilon,n-1)\cap I^n.$$

Also for each $y=(y_1,\dots,y_n)$ in the subset $(\frac14,\frac12)\times(\hat{W}(\tau-\varepsilon,n-1)\cap I^n)$, the point $(y_1,y_1y_2,\dots,y_1y_n)$ is mapped by T to y and is contained in the set $W(\tau,n)\cap\{x\in I^n:\frac14<x_1<\frac12\}$ since there are infinitely many $(q_2,\dots,q_n)$ in $\mathbb{Z}^{n-1}$ and q_1 in $\mathbb{Z}$ such that $|q_1+q_2y_2+\dots+q_ny_n|<q^{-\tau-\varepsilon}$ and so

$$|q_1y_1+q_2y_1y_2+\dots+q_1y_1y_n|<\tfrac12 q^{-\tau-\varepsilon}.$$

But $|q_1|<(n-1)q/2+\frac12$, whence $|q|=\max\{|q_1|,q\}<nq$ and so

$$|q_1+q_2y_2+\dots+q_ny_n|<n^{\tau+\varepsilon}|q|^{-\tau-\varepsilon}<|q|^{-\tau}$$

for infinitely many $\boldsymbol{q}$ ($|\boldsymbol{q}| > (n^{\tau+\varepsilon})^{1/\varepsilon}$) in $\mathbb{Z}^n$. Thus

$$T(W(\tau, n) \cap \{x \in I^n : \tfrac{1}{4} < x_1 < \tfrac{1}{2}\}) \supseteq (\tfrac{1}{4}, \tfrac{1}{2}) \times (\hat{W}(\tau+\varepsilon, n-1) \cap I^{n-1})$$

and by considering components, it is readily verified that for each $\boldsymbol{x}$, $\boldsymbol{x}'$ in $(\frac{1}{4}, \frac{1}{2}) \times I^{n-1}$, $|T(\boldsymbol{x}) - T(\boldsymbol{x}')| \leqslant 16|\boldsymbol{x} - \boldsymbol{x}'|$.

It can also be shown that the set $(0, \frac{1}{2}) \times (\hat{W}(\tau+\varepsilon, n-1) \cap I^{n-1})$ can be embedded analytically into $W(\tau, n)$ for any positive ε; such embeddings preserve the Hausdorff dimension.

5. The Hausdorff dimension of the exceptional sets

As has been said, when τ is large enough each of the sets $\hat{\Omega}(\tau)$ and $\Omega(\tau)$ discussed above is of Lebesgue measure 1, so that the measure of each of the corresponding complementary exceptional sets $\hat{E}(\tau)$ and $E(\tau)$ is 0. But although the exceptional sets can thus be regarded as thin or negligible, sets of measure 0 can be very different. The Hausdorff dimension (which, when there is no risk of confusion, will be referred to for convenience simply as the dimension and which will be explained below) provides more information about the size of a set of measure 0 and can distinguish between many such sets. For example, both the uncountable (perfect) Cantor middle third set and the countable (dense) set of rationals in $[0, 1]$ have Lebesgue measure 0 but their respective dimensions are $(\log 2)/(\log 3)$ and 0 (Falconer 1985). On the other hand, different sets can have the same dimension and, for example, the set $\hat{W}((\log 9/\log 2) - 1, 1)$ also has dimension $2/(\tau+1) = (\log 2)/(\log 3)$ (Jarnik 1929, Besicovitch 1934). Indeed, although $|\hat{\Omega}(n)| = 0$ and $|\hat{E}(n)| = 1$ (Arnol'd 1963, Sprindžuk 1979), it can be shown that $\hat{\Omega}(n)$ and $\hat{E}(n)$ both have Hausdorff dimension n (Schmidt 1969, Rüssman 1973). Hausdorff dimension (which is also referred to as Hausdorff-Besicovitch or fractional dimension) is a generalisation of the familiar notion of dimension but with the fundamental difference that any subset of $\mathbb{R}^n$ can be assigned a Hausdorff dimension. The price of this generality is a somewhat complicated definition (Rogers 1970, Falconer 1985) but the Hausdorff dimension does give an indication of the size of a set. A set with Hausdorff dimension close to n will, roughly speaking, be close to a set of positive Lebesgue measure.

The Hausdorff dimension of a set X in $\mathbb{R}^n$ will be denoted by $\dim X$ and can be defined as follows. Let ρ be any positive number and let Γ_ρ be any finite or countable cover of X by n-dimensional hypercubes C, where the length $L(C)$ of a side of each hypercube is at most ρ. For each real number s define the s volume to be

$$L^s(\Gamma_\rho) = \sum_{C \in \Gamma_\rho} L(C)^s.$$

Clearly $\inf L^s(\Gamma_\rho)$, where the infimum is taken over all covers Γ_ρ of X, cannot increase as ρ decreases, and if $s' \geqslant s$, then

$$\inf L^{s'}(\Gamma_\rho) \leqslant \rho^{s'-s} \inf L^s(\Gamma_\rho).$$

Thus if $s' > s$ and $\sup \inf L^s(\Gamma_\rho)$ is finite, where the supremum is over all positive ρ, then $\sup \inf L^{s'}(\Gamma_\rho) = 0$. The Hausdorff dimension $\dim X$ of X is the supremum over all real s for which $\sup \inf L^s(\Gamma_\rho)$ is positive, i.e.

$$\dim X = \sup\left\{ s \in \mathbb{R} : \sup_{\rho > 0} \inf_{\Gamma_\rho} L^s(\Gamma_\rho) > 0 \right\}.$$

It follows that if X can be covered by a collection Γ_ρ with arbitrarily small s volume $L^s(\Gamma_\rho)$, then $\dim X \leq s$. On the other hand, if for each positive ε, there exists a positive number $\rho = \rho(\varepsilon)$ such that every cover Γ_ρ of X with $L(C) \leq \rho$ satisfies $L^s(\Gamma_\rho) > \varepsilon$, then $\dim X \geq s$. Roughly speaking, if the s volume of covers consisting of small hypercubes of X is large, then $\dim X \geq s$. An equivalent condition is that if there exists a positive ε such that for any positive ρ, collections Γ_ρ satisfying $L^s(\Gamma_\rho) < \varepsilon$ cannot cover X, then $\dim X > s$. In other words, if collections of small hypercubes and small s volume cannot cover X, then $\dim X > s$.

Clearly a cover Γ of X will be a cover for any subset X' of X and it follows from the definition that if $X' \subset X \subset \mathbb{R}^n$, then

$$\dim X' \leq \dim X \leq n. \tag{5.1}$$

The determination of the dimension can often be simplified by the observation that when

$$X = \bigcup_{j=1}^{\infty} X_j$$

then

$$\dim X = \sup\{\dim X_j : j = 1, 2, \dots\}. \tag{5.2}$$

6. dim $W(\tau) \leq n - (\tau - n + 1)/(\tau + 1)$

To illustrate these ideas we shall show that $n - (\tau - n + 1)/(\tau + 1)$ is an upper bound for the Hausdorff dimension of $W(\tau)$ when $\tau > n - 1$. It is readily verified that $W(\tau, 1) = \{0\}$ when $\tau > 0$, so that from now on $n \geq 2$ unless otherwise stated. Since $\mathbb{R}^n$ is the union over all $\boldsymbol{q}$ in $\mathbb{Z}^n$ of hypercubes $I^n + \boldsymbol{q}$, it suffices to determine the dimension of $W(\tau) \cap I^n$.

To obtain the upper bound for dim $W(\tau)$, we construct a cover Γ for $W(\tau) \cap I^n$ of hypercubes C such that for each $s > n - (\tau - n + 1)/(\tau + 1)$, the s volume $L^s(\Gamma)$ can be made arbitrarily small, whence $\inf L^s(\Gamma) = 0$, so that by definition and (4.2), $\dim W(\tau) \leq n - (\tau - n + 1)/(\tau + 1)$.

For each $\boldsymbol{q}$ in $\mathbb{Z}^n$, write

$$H(\boldsymbol{q}) = \{\boldsymbol{x} \in I^n : |\boldsymbol{q} \cdot \boldsymbol{x}| = 0\}$$

so that $H(\boldsymbol{q})$ is a resonant hyperplane (Arnol'd 1979) and let

$$U(\boldsymbol{q}) = \{\boldsymbol{x} \in I^n : |\boldsymbol{q} \cdot \boldsymbol{x}| < |\boldsymbol{q}|^{-\tau}\}$$

be a neighbourhood of $H(\boldsymbol{q})$. Then for each positive integer N, $\{U(\boldsymbol{q}) : |\boldsymbol{q}| \geq N\}$ is a cover for $W(\tau) \cap I^n$. But each $U(\boldsymbol{q})$ has a cover $\Gamma(\boldsymbol{q})$ consisting of at most $K|\boldsymbol{q}|^{(n-1)(\tau+1)}$ n-dimensional hypercubes C with $L(C) = 8|\boldsymbol{q}|^{-(\tau+1)}$, where K is a constant (see figure 1). Hence for each N, $W(\tau) \cap I^n$ is covered by the collection $\Gamma_N = \{\Gamma(\boldsymbol{q}) : \boldsymbol{q} \in \mathbb{Z}^n, |\boldsymbol{q}| \geq N\}$, in which $L(C) \leq 8N^{-(\tau+1)}$, so that given any $\rho > 0$, Γ_N is a cover for $W(\tau) \cap I^n$ with $L(C) \leq \rho$ when N is sufficiently large. Now

$$L^s(\Gamma_N) = \sum_{C \in \Gamma_N} L(C)^s \ll \sum_{|\boldsymbol{q}| \geq N} |\boldsymbol{q}|^{(n-1)(\tau+1)} |\boldsymbol{q}|^{-s(\tau+1)}$$

where $\ll$ indicates an inequality with a positive constant factor. Rearranging the sum, we get

$$L^s(\Gamma_N) \ll \sum_{m=N}^{\infty} m^{(\tau+1)(n-1-s)} \sum_{|\boldsymbol{q}|=m} 1 \ll \sum_{m=N}^{\infty} m^{(\tau+1)(n-1-s)+n-1} \ll N^{n-(\tau+1)(s-n+1)}.$$

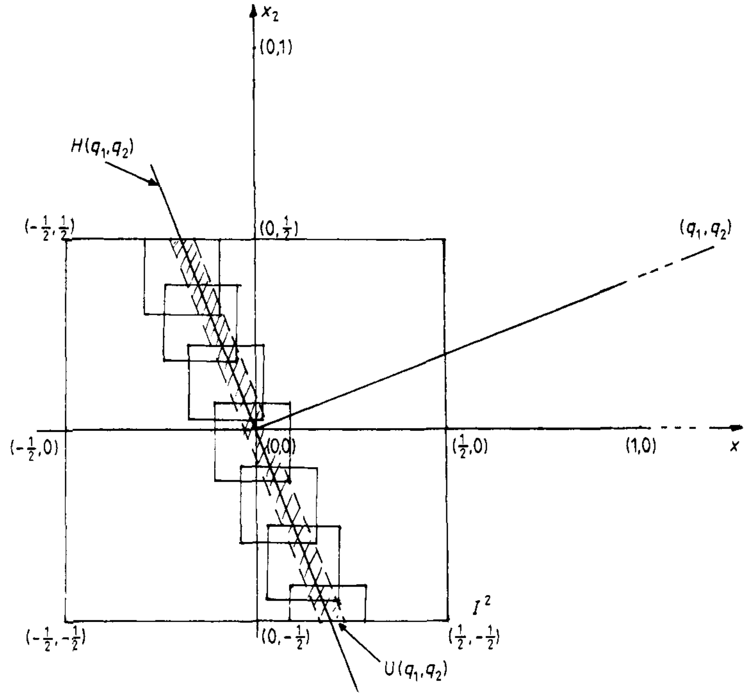

Figure 1.

Hence given $\varepsilon > 0$, $L^s(\Gamma_N) < \varepsilon$ for N sufficiently large, and it follows that $\inf L^s(\Gamma) = 0$, whence $s > \dim W(\tau) \cap I^n$. Since s is an arbitrary number greater than $n - (\tau - n + 1)/(\tau + 1)$, it follows from the definition that

$$\dim W(\tau)_n \cap I^n = \dim W(\tau) \leqslant n - (\tau - n + 1)/(\tau + 1).$$

The complementary inequality is much harder and will be dealt with later.

7. $\dim \hat{W}(\tau) \leqslant n - (\tau - n)/(\tau + 1)$

When $\tau > 2$, the set $\hat{W}(\tau, 1)$ is essentially the set of well approximable numbers and its dimension was shown to be $2/(\tau + 1)$ by Jarnik (1929, 1931) and Besicovitch (1934) (see also Eggleston 1952). Recently the Hausdorff dimension of a set more general than $\hat{W}(\tau)$ has been determined and using this result it can be shown that

$$\dim \hat{W}(\tau) = n - (\tau - n)/(\tau + 1)$$

when $\tau > n$ (Bovey and Dodson 1985). However, to keep this paper reasonably self-contained and because the arguments are of interest, the dimension of $\hat{W}(\tau)$ will be established for $n \geqslant 2$.

As the first (and easy) step in the determination of $\dim \hat{W}(\tau)$, we obtain the following upper bound which holds for $n \geqslant 1$.

Lemma 7.1. When $\tau > n$,

$$\dim \hat{W}(\tau) \leqslant n - (\tau - n)/(\tau + 1) = n - 1 + (n + 1)/(\tau + 1).$$

Proof. Let $\varepsilon > 0$ and $\delta > 0$ be given and let $t = n - 1 + (n+1)/(\tau+1)$. For each $\boldsymbol{q} \in \mathbb{Z}^n$ and $m \in \mathbb{Z}$, cover the $(m-1)$-dimensional resonant hyperplane

$$H(\boldsymbol{q}, m) = \{\boldsymbol{x} \in \mathbb{R}^n : |\boldsymbol{q} \cdot \boldsymbol{x} - m| = 0\}$$

in $\mathbb{R}^n$ with n-dimensional hypercubes of side $8|\boldsymbol{q}|^{-(\tau+1)}$ in such a way that the vertices lie on a lattice of width $|q|^{-(\tau+1)}$. Then $I^n \cap H(\boldsymbol{q}, m)$ can be covered by $\ll |\boldsymbol{q}|^{(\tau+1)(n-1)}$ such cubes (the notation $\ll$ indicates an inequality with an unspecified constant factor), as shown in figure 2. The collection of such cubes $\Gamma(\boldsymbol{q}, m)$, say, also forms a covering for the neighbourhood or thickened hyperplane

$$B(\boldsymbol{q}, m) = \{\boldsymbol{x} \in \mathbb{R}^n : |\boldsymbol{q} \cdot \boldsymbol{x} - m| < |\boldsymbol{q}|^{-\tau}\}.$$

Now for each $N = 1, 2, \ldots,$ the collection

$$\Gamma_N = \{\Gamma(\boldsymbol{q}, m) : |m| < \tfrac{1}{2}|\boldsymbol{q}|, |\boldsymbol{q}| > N\}$$

covers $\hat{W}(\tau)$ and the t volume of Γ_N is given by

$$\begin{aligned} L^t(\Gamma_N) &= \sum_{\boldsymbol{q}} \sum_{m} \sum_{C \in \Gamma(\boldsymbol{q}, m)} 8^t |\boldsymbol{q}|^{-(\tau+1)t} \\ &\ll \sum_{q > N} \sum_{|\boldsymbol{q}| = q} \sum_{m} |\boldsymbol{q}|^{-(\tau+1)t} |\boldsymbol{q}|^{(\tau+1)(n-1)} \\ &\ll \sum_{q > N} q^{-(\tau+1)t} q^{(\tau+1)(n-1)} q^n \ll \sum_{q > N} q^{-1-(\tau+1)\delta} < \varepsilon \end{aligned}$$

for N sufficiently large. Thus for all $t > n - (\tau - n)/(\tau+1)$, there exists a cover of $\hat{W}(\tau)$ with arbitrarily small t volume and hence

$$\dim \hat{W}(\tau) \leqslant n - (\tau - n)/(\tau + 1).$$

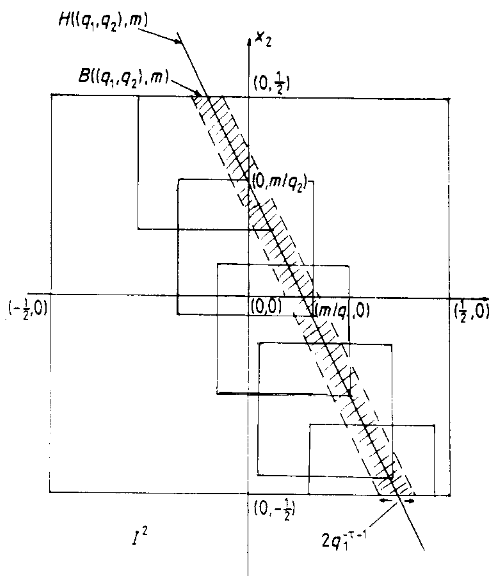

Figure 2.

8. dim $\hat{W}(\tau) = n - (\tau - n)/(\tau + n)$

As is common in the determination of Hausdorff dimension, the complementary inequality

$$\dim \hat{W}(\tau) = \dim \hat{W}(\tau, n) \geq n - (\tau - n)/(\tau + 1) \tag{8.1}$$

$(\tau > n)$, is much harder to establish. The methods used in Bovey and Dodson (1985) have some features in common with those of Jarnik (1929, 1931) and Besicovitch (1934), who first proved the result for $n = 1$, but also include a 'variance' or 'second moment' argument. These and subsequent arguments are somewhat elaborate and rely upon the distribution of certain resonant hyperplanes being roughly regular with an associated variance not being too large, suggesting a parallel with the notion of independence in probability. There is also a similarity with self-similar sets which reproduce themselves at certain scales and particularly with 'statistically' self-similar sets (Mandelbrot 1983, Falconer 1985). Recently Sullivan (1982) has used the probabilistic independence of certain sets in I^n, where $n = 1$ or 2, to give a new proof of Khintchine's approximation theorem (corresponding to $n = 1$) and a complex version when $n = 2$. There are a number of similarities between Sullivan (1982) and Bovey and Dodson (1985), such as some general statistical arguments, including the use of independence, which arises in Sullivan from a collection of disjoint spheres and in Bovey and Dodson from a special set of hyperplanes; also the flows in Sullivan are ergodic and even-mixing as are flows implicit in Bovey and Dodson. When $n \geq 2$, some of the arguments in Bovey and Dodson can be simplified and made more geometrical, giving the fundamental 'invariance of measure' and 'independence' results (8.4) and (8.5) below. These results are sharper than the corresponding ones in Bovey and Dodson and lead to sharper mean and second moment results (lemmas 8.1 and 8.2, respectively). In view of this and of the results of Jarnik and Besicovitch, we shall take $n \geq 2$ for this section.

We start by introducing some additional definitions and notation (recall that $\tau > n$). Let δ be any positive number and put

$$s = n - (\tau - n)/(\tau + 1) - \delta.$$

Suppose for some positive ε, the countable collection Γ of hypercubes C with $L(C) \leq \xi$, where ξ is an arbitrary positive number, satisfies

$$L^s(\Gamma) = \sum L(C)^s < \varepsilon. \tag{8.2}$$

We will show that no such collection Γ can cover $\hat{W}(\tau, n)$ and hence that the inequality (8.1) holds.

Let N be a sufficiently large positive integer and let η satisfy

$$0 < \eta < \min\{\tau - n, (\tau + 1)\delta\}. \tag{8.3}$$

Let $\boldsymbol{p}$ denote any vector $(p_1, \ldots, p_n)$ in $\mathbb{Z}^n$ satisfying $|\boldsymbol{p}| = p_1$ with

$$N < p_1 < 2N \qquad |p_i| < N \qquad \text{for } i = 2, \ldots, n.$$

Then

$$\sum_{\boldsymbol{p}} 1 \sim 2^{n-1} N^{(n-1)+1}$$

where $a(N) \sim b(N)$ means $a(N) = b(N)(1 + o(1))$. Let

$$S_N = \{H(\boldsymbol{p}, m) : |m| < \tfrac{1}{2}|\boldsymbol{p}| = \tfrac{1}{2}p_1\} \subset \hat{W}(\tau, n)$$

so that S_N is a set of resonant hyperplanes (see Bovey and Dodson 1985) and $|S_N|$ is the number of hyperplanes in S_N and satisfies

$$|S_N| \sim \sum_p \sum_m 1 \sim \sum_p p_1 \sim 3\times 2^{n-2}\times N^{n+1}$$

where, as always, m satisfies $|m|<\frac{1}{2}p_1$ unless otherwise stated.

We will show that the hyperplanes $H(\boldsymbol{p}, m)$ in S_N are asymptotically regularly distributed and can be used to construct a 'sampling' set $T(N)\subset \hat{W}(\tau)$ which can be used to show that Γ cannot cover $\hat{W}(\tau)$ (in fact, Γ fails to cover at least one point in a particular subset of $\hat{W}(\tau)$).

Choose η so that $0<\eta<\min\{\tau-n, (\tau+1)\delta\}$ and let

$$\rho=\rho(N)=N^{-n+\eta}$$

so that $\rho\to 0$ as $N\to\infty$. Let $\chi=\chi_{(-\rho,\rho)}$ be the characteristic function of the interval $(-\rho,\rho)$. Then for each $\boldsymbol{p}$

$$\int_{I^n}\chi(\boldsymbol{p}\cdot\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}=2\rho p_1^{-1}.$$

Suppose that $|\boldsymbol{p}\cdot\boldsymbol{x}-m|<\rho$, i.e. $\chi(\boldsymbol{p}\cdot\boldsymbol{x}-m)=1$. Then since N is sufficiently large, for any integer $m\neq m'$,

$$|\boldsymbol{p}\cdot\boldsymbol{x}-m'|\geqslant|m-m'|-|\boldsymbol{p}\cdot\boldsymbol{x}-m|>1-N^{-n+\eta}>\rho$$

i.e. $\chi(\boldsymbol{p}\cdot\boldsymbol{x}-m')=0$. It follows that for each $\boldsymbol{x}$ in I^n

$$\sum_m \chi(\boldsymbol{p}\cdot\boldsymbol{x}-m)=\begin{cases}1 & \text{if } |\boldsymbol{p}\cdot\boldsymbol{x}-m|<\rho \text{ for some } m,\ (|m|<\frac{1}{2}p_1)\\ 0 & \text{otherwise.}\end{cases}$$

For each vector $\boldsymbol{v}$ in $\mathbb{R}^n$ define the function $\Phi_{\boldsymbol{v}}:\mathbb{R}^n\to I$ by

$$\Phi_{\boldsymbol{v}}(\boldsymbol{x})=\langle\boldsymbol{v}\cdot\boldsymbol{x}\rangle.$$

Note also that

$$\sum_m \chi(\boldsymbol{p}\cdot\boldsymbol{x}-m)=\chi_A(\boldsymbol{x})$$

where the set $A=\Phi_{\boldsymbol{p}}^{-1}(-\rho,\rho)$. When $\boldsymbol{q}\in\mathbb{Z}^n$, the function $\Phi_{\boldsymbol{q}}$ is periodic and as a consequence, when $n\geqslant 2$, the measure of the inverse image under $\Phi_{\boldsymbol{q}}$ is preserved, i.e. for any interval J in I,

$$|\Phi_{\boldsymbol{q}}^{-1}(J)|=|J| \tag{8.4}$$

and the map $\Phi_{\boldsymbol{q}}$ is 'stochastic' or independent in the sense that for any independent $\boldsymbol{q}$, $\boldsymbol{q}'$,

$$|\Phi_{\boldsymbol{q}}^{-1}(J)\cap\Phi_{\boldsymbol{q}'}^{-1}(J')|=|\Phi_{\boldsymbol{q}}^{-1}(J)|\cdot|\Phi_{\boldsymbol{q}'}^{-1}(J')|=|J|\cdot|J'|. \tag{8.5}$$

These results are proved in a more general setting by Sprindžuk (1979, ch I, lemmas 8 and 9), but the arguments given here are of a more geometric character and require only translation invariance rather than linearity.

We start by proving (8.4) in the case where $n=2$ and then consider the more general case. Let $Z\subset I^2$ be the set of points on which $\Phi_{\boldsymbol{q}}(\boldsymbol{x})$ vanishes, i.e. $Z=\Phi_{\boldsymbol{q}}^{-1}(0)$, the zero set of $\Phi_{\boldsymbol{q}}$. If we regard $I^2=(-\frac{1}{2},\frac{1}{2}]^2$ as the two-dimensional torus T^2, then the line $x_2=0$ and the set Z divide T^2 into strips S_i, $i=1,\ldots,q_1$, each of width $1/q_1$. Let $P=[0,\rho)$ and consider the set $B=\Phi_{\boldsymbol{q}}^{-1}(P)$. This set consists of the lines making up

Z thickened to form strips $\tilde{S}_i$ of width ρ/q_1 (see figure 3). Thus each strip S_i contains a substrip $\tilde{S}_i$ with the respective areas in the ratio

$$\text{area } S_1 : \text{area } \tilde{S}_i = 1/q_1 : \rho/q_1 = 1 : \rho.$$

But T^2 is covered by the strips S_i, so that

$$1 : \text{area } B = \text{area } T^2 : \text{area } B = \text{area } S_i : \text{area } \tilde{S}_i = 1 : \rho$$

and thus the area of B is ρ or

$$|\Phi_q^{-1}(P)| = |P|.$$

The translation invariance of Φ_q ensures that the result remains true, i.e. (8.4) holds, for a general interval and not just one of the form $[0, \rho)$. The above argument is readily generalised to the n-dimensional case by considering the way in which T^n is divided up into prisms by the zero set Z of Φ_q and the plane $x_3 = \ldots = x_n = 0$.

We now turn to (8.5) in the case when $n = 2$. Let Z be the zero set of Φ_q and Z' be the zero set of $\Phi_{q'}$. Then $Z \cup Z'$ divides T^2 into $|q \times q'|$ parallelograms Π_i, each of area $|q \times q'|^{-1}$ (see figure 4). Let $P = [0, \rho)$ and $P' = [0, \rho')$ and consider the sets $B = \Phi_q^{-1}(P)$ and $B' = \Phi_{q'}^{-1}(P')$. These sets are obtained by thickening the lines making up Z and Z' and taking the parallelograms $\tilde{\Pi}_i$ arising from their intersections. Each parallelogram Π_i contains just one parallelogram $\tilde{\Pi}_i$ and elementary geometry shows that the areas are in the ratio

$$\text{area } \Pi_i : \text{area } \tilde{\Pi}_i = 1 : \rho\rho'.$$

Now the torus T^2 is tessellated by the Π_i, so that

$$1 : \text{area } B \cap B' = \text{area } T^2 : \text{area } B \cap B' = \text{area } \Pi_i : \text{area } \tilde{\Pi}_i = 1 : \rho\rho'$$

whence the area of $B \cap B'$ is $\rho\rho'$ or

$$|\Phi_q^{-1}(P) \cap \Phi_{q'}^{-1}(P')| = |P| \cdot |P'|.$$

Again the translation invariance of Φ_q ensures that this result remains true for general intervals.

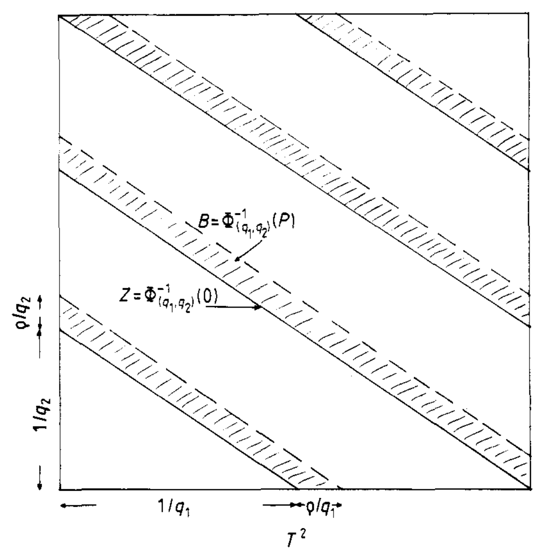

Figure 3.

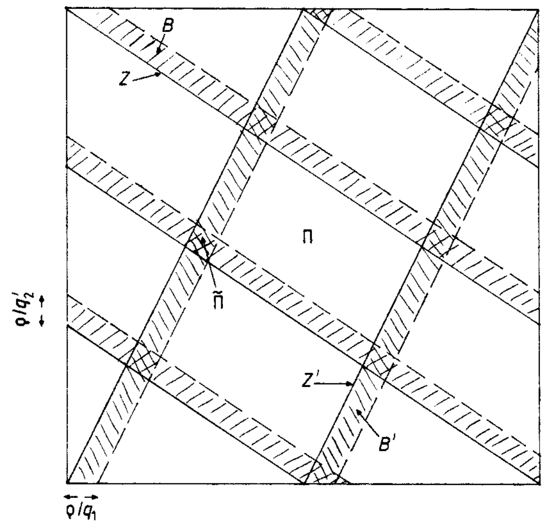

Figure 4.

To establish (8.5) for general n, we first observe that $Z=\Phi_q^{-1}(0)$ consists of a family of $(n-1)$-dimensional hyperplanes, each with normal $\boldsymbol{q}$, and that Z' is a family of hyperplanes with normal $\boldsymbol{q}'$. Let

$$H=\{\boldsymbol{x}\in I^n\colon \boldsymbol{x}=\alpha\boldsymbol{q}+\beta\boldsymbol{q}',\ \alpha,\beta\in\mathbb{R}\}$$

so that H is the 2-plane through the origin 0 of $\mathbb{R}^n$, spanned by $\boldsymbol{q}$ and $\boldsymbol{q}'$. Then $\Phi_q^{-1}(0)$ and $\Phi_{q'}^{-1}(0)$ are orthogonal to H and

$$\operatorname{vol}\{\Phi_q^{-1}(P)\cap\Phi_{q'}^{-1}(P')\}=\operatorname{area}\{\Phi_q^{-1}(P)\cap\Phi_{q'}^{-1}(P')\cap H\}$$

since $\Phi_q^{-1}(P)\cap\Phi_{q'}^{-1}(P')$ consists of a prism with base

$$\Phi_q^{-1}(P)\cap\Phi_{q'}^{-1}(P')\cap H$$

and with height 1 in each complementary dimension. But by the result for $n=2$, it follows that the area of the base is $\rho\rho'$, so that

$$|\Phi_q^{-1}(P)\cap\Phi_{q'}^{-1}(P')|=|P|\cdot|P'|.$$

The translational invariance of Φ_q ensures that this result holds for general intervals, i.e. that (8.5) holds.

Now the function $\nu_N\colon I^n\to\mathbb{Z}$ given by

$$\nu_N(\boldsymbol{x})=\sum_{\boldsymbol{p}}\sum_{m}\chi(\boldsymbol{p}\cdot\boldsymbol{x}-m)=\sum_{\boldsymbol{p}}\chi_A(\boldsymbol{x})$$

is the number of hyperplanes $H(\boldsymbol{p},k)$ in S_N within ρ of $\boldsymbol{x}$. Define

$$\mu_N=\int_{I^n}\nu_N(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

and

$$\sigma_N^2=\int_{I^n}(\nu_N(\boldsymbol{x})-\mu_N)^2\,\mathrm{d}\boldsymbol{x}=\int_{I^n}\nu_N(\boldsymbol{x})^2\,\mathrm{d}\boldsymbol{x}-\mu_N^2.$$

We show that the resonant hyperplanes $H(\boldsymbol{p}, m)$ are asymptotically regularly distributed in the sense that the variance σ_N^2 of ν_N is small, satisfying

$$\sigma_N^2 \leqslant \mu_N.$$

First we calculate the mean μ_N.

Lemma 8.1. The mean μ_N of ν_N is given by

$$\mu_N = 2\rho \sum_{\boldsymbol{p}} 1 \sim 2^n N^{\eta}.$$

Proof. By definition

$$\mu_N = \int_{I^n} \sum_{\boldsymbol{p}} \chi_A(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \sum_{\boldsymbol{p}} |\Phi_{\boldsymbol{p}}^{-1}(-\rho, \rho)| = 2\rho \sum_{\boldsymbol{p}} 1$$

by (8.4). Hence

$$\mu_N = 2N^{-n+\eta} 2^{n-1} N^{1+(n-1)}(1+\mathrm{o}(1)) \sim 2^n N^{\eta}.$$

Next we estimate the second moment of ν_N.

Lemma 8.2.

$$\int_{I^n} \nu_N^2(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \leqslant \mu_N + \mu_N^2.$$

Proof. Write $A = \Phi_{\boldsymbol{p}}^{-1}(-\rho, \rho)$, $A' = \Phi_{\boldsymbol{p}'}^{-1}(-\rho, \rho)$. By definition,

$$\begin{aligned}
\int_{I^n} \nu_N^2(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} &= \int_{I^n} \sum_{\boldsymbol{p}} \sum_{\boldsymbol{p}'} \chi_A(\boldsymbol{x})\chi_{A'}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \\
&= \sum_{\boldsymbol{p}} \int_{I^n} \chi_A(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} + \sum_{\boldsymbol{p} \neq \boldsymbol{p}'}\sum \int_{I^n} \chi_A(\boldsymbol{x})\chi_{A'}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \\
&= \mu_N + \sum_{\boldsymbol{p} \neq \boldsymbol{p}'}\sum |\Phi_{\boldsymbol{p}}^{-1}(-\rho, \rho)| \cap |\Phi_{\boldsymbol{p}'}^{-1}(-\rho, \rho)| \\
&= \mu_N + \sum_{\boldsymbol{p} \neq \boldsymbol{p}'}\sum |\Phi_{\boldsymbol{p}}^{-1}(-\rho, \rho)| \cdot |\Phi_{\boldsymbol{p}'}^{-1}(-\rho, \rho)|
\end{aligned}$$

by (8.5). Hence

$$\int_{I^n} \nu_N^2(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \mu N + \sum_{\boldsymbol{p} \neq \boldsymbol{p}'} 4\rho^2$$

by (8.4) and so

$$\begin{aligned}
\int_{I^n} \nu_N^2(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} &\leqslant \mu_N + \left(2\rho \sum_{\boldsymbol{p}} 1\right)^2 \\
&\leqslant \mu_N + \mu_N^2
\end{aligned}$$

by lemma 8.1.

The following result is immediate.

Corollary 8.2.

$$\sigma_N^2 \leqslant \mu_N.$$

Thus ν_N is not more spread than a Poisson distribution on the positive reals, since for such a process the mean and the variance are equal.

The set

$$\nu_N^{-1}(0) = \{x \in I^n : \nu_N(x) = 0\} = Z_N \qquad \text{say}$$

consists of points x which are not near any hyperplane $H(p, m)$ in S_N. The corollary implies that the volume $|Z_N|$ of Z_N is small.

Lemma 8.3.

$$|Z_N| = o(1).$$

Proof. The variance

$$\sigma_N^2 = \int_{I^n} (\nu_N(x) - \mu_N)^2 \, dx \geq \int_{Z_N} (\nu_N(x) - \mu_N)^2 \, dx = \mu_N^2 |Z_N|$$

whence

$$|Z_N| \leq \sigma_N^2/(\mu_N)^2 \leq 1/\mu_N = o(1).$$

We are now in a position to construct the sampling set $T(N)$ by selecting well distributed resonant hyperplanes in S_N and then thickening them slightly so that the resulting set is still a subset of $W(\tau) \cap I^n$.

Dissect I^n into $[N/(16\rho)]^n = [2^{-4}N^{n-\eta+1}]^n$ congruent hypercubes H with side $L(H) = [2^{-4}N^{n-\eta+1}]^{-1} \sim 16N^{-n-\eta-1}$. Now shrink each H by $\frac{1}{2}$ about its centre to obtain a similar hypercube H' with $L(H') = \frac{1}{2}L(H)$. Suppose there are M hypercubes such that for each $H(p, m)$ in S_N the $(n-1)$-dimensional volume of the intersection of H' with $H(p, m)$ is less than $(L(H)/\sqrt{2})^{n-1}$. Let H'' be the result of shrinking such a hypercube H' by $\frac{1}{2}$ about its centre (see figure 5) so that $L(H) = 4L(H'')$. Then every

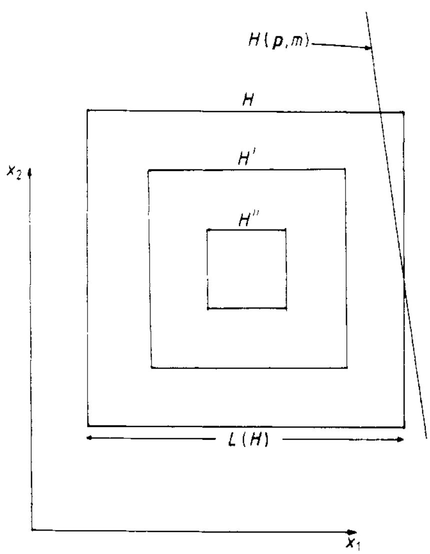

Figure 5.

point in H'' is at least $\frac{1}{8}L(H)(1+o(1))$ from each $H(p, m)$. But $\frac{1}{8}L(H)(1+o(1)) \sim \frac{16}{8}N^{\eta-n-1} = 2\rho/N$, so that when $x \in H''$ and $H(p, m) \in S_N$, $|p \cdot x - m| > \rho$ whence $\chi(p \cdot x - m) = 0$. Thus $x \in H''$ implies $\nu_N(x) = 0$, i.e. $x \in Z_N$. It follows that $M(\frac{1}{4}L(H))^n \leqslant |Z_N| = o(1)$, so that $M = o(L(H)^{-n}) = o(N^{(n+1-\eta)n}) = o(\rho/N)^n$. Thus there are $(N/16\rho)^n(1+o(1))$ hypercubes H' whose intersection with $H(p, m) \in S_N$ has volume at least $(L(H)/\sqrt{2})^{n-1}$. Pick one such intersection or 'slice' S, say, for each such hypercube H', so that asymptotically there are $(N/16\rho)^n \sim L(H)^{-n}$ such slices. Let

$$V = V(S) = \{x \in \mathrm{cl}\, H' \colon |x - y| < n^{-1}(2N)^{-\tau-1} \text{ for some } y \text{ in } S\}$$

i.e. $V = V(S)$ is a thickening of S, and let $T(N)$ be the collection of such V. Then the n-dimensional volume $|T(N)|$ of $T(N)$ is given by

$$|T(N)| = \sum_{V \in T(N)} |V|$$

and satisfies

$$L(H)^{-n}N^{-\tau-1}L(H)^{n-1} \ll |T(N)| \ll L(H)^{-1}N^{-\tau-1} \ll N^{n-\eta-\tau}.$$

The importance of the set $T(N)$ is that it is sufficiently regular and numerous to 'measure' the volume of a set.

Lemma 8.4. Let X be a set in I^n whose boundary is of measure 0. Then

$$|X \cap T(N)| \sim |X| \cdot |T(N)|.$$

Proof. Dissect I^n into $[N/(16\rho)]^n$ hypercubes as before and let P be the number which lie completely within X and P' the number which meet both X and $I^n \backslash X$ (see figure 6). Then $P \cdot L(H)^n \leqslant |X| \leqslant (P + P')L(H)$, and since $P' \cdot L(H)^n = o(1)$, it follows that $|X| \sim P \cdot L(H)^n$. The number of cubes in X which contain a set V from $T(N)$ is $P + o(L(H))^{-n}$ and thus

$$|X \cap T(N)| \sim (P + o(L(H))^{-n} \cdot |\bar{V}| \sim |X| \cdot L(H)^{-n}L(H)^n|T(N)| \sim |X| \cdot |T(N)|$$

where $|\bar{V}|$ is the mean of $|V|$ for V in $T(N)$.

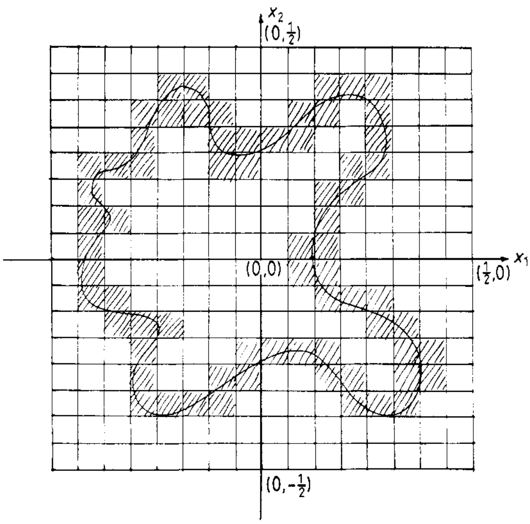

Figure 6.

When the set X depends on N, i.e. when $X = X(N)$, the above proof breaks down but the volume can be estimated in the case we are going to need.

Lemma 8.5. Let C be an n-dimensional hypercube with $L(C) \geq N^{-(\tau+1)}$. Then

$$|C \cap T(N)| \ll |C| \cdot |T(N)| + L(C)^{n-1} \cdot N^{-(\tau+1)}.$$

Proof. Since $\eta < \tau - n$ and N is sufficiently large,

$$N^{-(\tau+1)} < 16N^{-n-1-\eta} \sim [2^{-4}N^{1+n-\eta}]^{-1} = L(H).$$

As before, I^n is dissected into $[N/(16\rho)]^n$ cubes with $L(H) = [N/(16\rho)]^{-1}$. By using arguments similar to those in the preceding lemma it can be shown that the number of sets V which meet C is $\ll (L(C)/L(H))^n$ when $L(C) > L(H)$; and is $\ll 1$ when $L(C) \leq L(H)$. Hence by the above

$$\begin{aligned}|C \cap T(N)| &\ll L(C)^n \cdot L(H)^{-n} \cdot |V| + L(C)^{n-1} \cdot N^{-(\tau+1)} \\ &\ll |C| \cdot |T(N)| + L(C)^{n-1} \cdot N^{-(\tau+1)}.\end{aligned}$$

We now apply these estimates to hypercubes C from Γ where $L(C)$ lies in a suitable range. Let N_{s-1}, N_s be sufficiently large integers with $N_{s-1} < N_s$ and define

$$R(s) = \{C \in \Gamma \colon N_s^{-(\tau+1)} < L(C) \leq N_{s-1}^{-(\tau+1)}\}.$$

Then by using lemma 8.5, summing over all cubes in $R(s)$ and using (8.1), it can be shown that

$$|R(s) \cap T(N_s)| \ll |T(N_s)|\{|T(N_{s-1})| \cdot N_{s-1}^{\eta-(\tau+1)\delta} + N_s^{\eta-(\tau+1)\delta}\} \tag{8.6}$$

where the implied constants do not depend on N_{s-1} or N_s (a more general version of (8.6) is proved in Bovey and Dodson (1985)). We now use $T(N)$ to construct a set which is a subset of $\hat{W}(\tau, n)$ but which is not covered by Γ.

Let $G_0 = [-\frac{1}{2}, \frac{1}{2}]^n$ and define the compact sets G_s inductively by

$$G_s = [G_{s-1} \cap T(N_s)] \backslash R(s).$$

Lemma 8.6. For each $s = 1, 2, \ldots,$ the integers $N_0, N_1, \ldots, N_s$ can be chosen to increase with sufficient rapidity so that

$$|G_s| > 2^{-2s} \prod_{j=1}^{s} |T(N_j)| > 0.$$

Proof. Clearly $|G_0| = 1$ and from the definition of G_1,

$$|G_1| = |T(N_1)| - |R(1) \cap T(N_1)| \geq |T(N_1)|(1 - K\{|T(N_0)|N_0^{-\beta} + N_1^{-\beta}\})$$

where K is the implied constant in lemma 8.6 and where $\beta = (\tau+1)\delta - \eta > 0$ by the choice of η (8.3). Hence by choosing N_0 and N_1 sufficiently large,

$$|G_1| \geq 2^{-2}|T(N_1)|.$$

Now

$$|G_2| = |G_1 \cap T(N_2)| - |G_1 \cap T(N_2) \cap R(2)|$$

and by lemma 8.5 we can take N_2 sufficiently large so that

$$|G_1 \cap T(N_2)| \geq \tfrac{1}{2}|G_1||T(N_2)| \geq 2^{-3}|T(N_1)||T(N_2)|.$$

Moreover by lemma 8.5,

$$|G_1 \cap T(N_2) \cap R(2)| \leqslant K \cdot |T(N_2)|(|T(N_1)|N_1^\beta + N_2^\beta)$$

and we can take N_1 and N_2 sufficiently large so that

$$KN_1^{-\beta} > 2^{-5}KN_2^{-\beta} > 2^{-5}|T(N_1)|.$$

Hence

$$|G_2| > 2^{-4}|T(N_1)||T(N_2)|$$

and repeated application gives the required result (see Bovey and Dodson (1985) for a more general and detailed proof).

Lemma 8.6 implies that for each $s = 1, 2, \ldots,$ the compact set G_s is not empty; but by construction

$$G_s = \bigcap_{j=1}^{s} G_j \neq \varnothing.$$

Hence by the finite intersection property, G_∞ is not empty. Now each C in Γ does not meet G_∞, since every C in Γ is in $R(s)$ for some s and hence cannot be in $G_s \supset G_\infty$. Hence the collection Γ cannot cover G_∞ and we will show that $G_\infty \subset \hat{W}(\tau, n)$.

Suppose $\boldsymbol{x} \in G_\infty$. Then $\boldsymbol{x} \in G_s$ for $s = 1, 2, \ldots,$ and hence $\boldsymbol{x} \in T(N_s)$ for $s = 1, 2, \ldots.$ Thus for each $s = 1, 2, \ldots,$ there is a point $\boldsymbol{y}$ in some $H(\boldsymbol{p}, m) \in S_{N_s}$ with

$$|\boldsymbol{x} - \boldsymbol{y}| \leqslant n^{-1}(2N_s)^{-(\tau+1)}$$

and therefore, since $\boldsymbol{y} \in H(\boldsymbol{p}, m)$,

$$|\boldsymbol{p} \cdot \boldsymbol{x} - m| = |\boldsymbol{p} \cdot (\boldsymbol{x} - \boldsymbol{y})| \leqslant n \cdot |\boldsymbol{p}| \cdot |\boldsymbol{x} - \boldsymbol{y}| \leqslant |\boldsymbol{p}| \cdot (2N_s)^{-(\tau+1)} < |\boldsymbol{p}|^{-\tau}$$

since $|\boldsymbol{p}| < 2N_s$. Hence for each $\boldsymbol{x}$ in G_∞, there are infinitely many $(\boldsymbol{p}, m)$ in $\mathbb{Z}^{n+1}$ such that

$$|\boldsymbol{p} \cdot \boldsymbol{x} - m| < |\boldsymbol{p}|^{-\tau}$$

and so if $\boldsymbol{x} \in G_\infty$, then $\boldsymbol{x} \in \hat{W}(\tau, n)$, i.e. $G_\infty \subset \hat{W}(\tau, n)$. Because it is not a cover of G_∞, Γ cannot cover $\hat{W}(\tau, n)$ and it follows from the definition of dimension that

$$\dim \hat{W}(\tau, n) \geqslant n - (\tau - n)/(\tau + 1)$$

which, together with lemma 7.1, proves the following.

Theorem 8.7. When $\tau > n$, the set

$$\hat{W}(\tau) = \{\boldsymbol{x} \in \mathbb{R}^n : |\langle \boldsymbol{q} \cdot \boldsymbol{x} \rangle| < |\boldsymbol{q}|^{-\tau} \text{ for infinitely many } \boldsymbol{q} \text{ in } \mathbb{Z}^n\}$$

has Hausdorff dimension

$$\dim \hat{W}(\tau) = n - (\tau - n)/(\tau + 1)$$

and when $\tau \leqslant n$, $\hat{W}(\tau)$ has Lebesgue measure 1.

Proof. Only the result $|\hat{W}(\tau)| = 1$ when $\tau \leqslant n$ has to be proved and since, when $\tau \leqslant n$,

$$\sum_{\boldsymbol{q}} |\boldsymbol{q}|^{-\tau} = \infty$$

where the summation is over all non-zero $\boldsymbol{q}$ in $\mathbb{Z}^n$, this follows from a general theorem of Khintchine type established by Groshev (see Sprindžuk 1979).

9. The dimension of $W(\tau)$

Having obtained the dimension of the set of well approximable forms $\hat{W}(\tau)$, it is now possible to determine the dimension of $W(\tau)$.

First the Hausdorff dimension of the interval $(\frac{1}{4}, \frac{1}{2})$ is 1 and that of the cartesian product $(\frac{1}{4}, \frac{1}{2}) \times X$ is $\dim(\frac{1}{4}, \frac{1}{2}) + \dim X = 1 + \dim X$ (Bovey and Dodson 1978, Falconer 1985). Next when a map $f\colon X \to Y$ satisfies a Lipschitz condition and is onto, then $\dim X \leqslant \dim Y$ (Falconer 1985, p 10, lemma 1.8). It follows from lemma 4.2 that for each positive ε,

$$\dim W(\tau, n) \geqslant \dim T\{(\tfrac{1}{4}, \tfrac{1}{2}) \times \hat{W}(\tau+\varepsilon, n-1) \cap I^{n-1}\} = 1 + \dim \hat{W}(\tau+\varepsilon, n-1)$$

whence for $\tau > n-1$,

$$\dim W(\tau, n) \geqslant 1 + n - 1 - \frac{\tau+\varepsilon-n+1}{\tau+1+\varepsilon} = n - \frac{\tau+\varepsilon-n+1}{\tau+1+\varepsilon}.$$

Since ε is an arbitrary positive number, $\dim W(\tau, n) \geqslant n - (\tau - n + 1)/(\tau+1)$ when $\tau > n-1$, and so by § 6,

$$\dim W(\tau, n) = \dim W(\tau) = n - \frac{\tau - n + 1}{\tau + 1} \tag{9.1}$$

when $\tau > n-1$. When $\tau \leqslant n-1$, $W(\tau, n)$ has full Lebesgue measure; this follows from lemma 4.1 when $\tau < n-1$ and $|W(n-1)| = 1$ follows from (10.1).

10. The dimension of the exceptional sets

The dimensions of the exceptional sets $\hat{E}(\tau)$ and $E(\tau)$ are obtained by a continuity argument from the dimensions of $\hat{W}(\tau, n)$ and $W(\tau, n)$. By lemma 4.1, given any $\varepsilon > 0$, $\hat{E}(\tau)$ is trapped between $\hat{W}(\tau+\varepsilon)$ arnd $\hat{W}(\tau)$ and $E(\tau)$ between $W(\tau+\varepsilon)$ and $W(\tau)$. Hence, by (5.1),

$$\begin{aligned} &\dim \hat{W}(\tau+\varepsilon) \leqslant \dim \hat{E}(\tau) \leqslant \dim \hat{W}(\tau) \\ &\dim W(\tau+\varepsilon) \leqslant \dim E(\tau) \leqslant \dim W(\tau). \end{aligned} \tag{10.1}$$

Thus by theorem 8.7, $n - (\tau+\varepsilon-n)/(\tau+\varepsilon+1) \leqslant \dim \hat{E}(\tau) \leqslant n - (\tau - n)/(\tau+1)$ for any positive ε when $\tau > n$. When $\tau < n$, $\hat{E}(\tau) = \mathbb{R}^n$ and $|\hat{E}(n)| = 1$ (since $|\hat{\Omega}(n)| = 0$). Hence, since $\varepsilon > 0$ is arbitrary,

$$\dim \hat{E}(\tau) = \dim \hat{W}(\tau) = \begin{cases} n - (\tau - n)/(\tau+1) & \text{when } \tau > n \\ n & \text{when } \tau \leqslant n. \end{cases}$$

The dimension of $E(\tau)$ is determined similarly. By (10.1) and (9.1), for any $\varepsilon > 0$, $n - (\tau+\varepsilon-n+1)/(\tau+\varepsilon+1) \leqslant \dim E(\tau) \leqslant n - (\tau-n+1)/(\tau+1)$ when $\tau > n-1$. Recall that $|E(n-1)| = 1$ and $E(\tau) = \mathbb{R}^n$ when $\tau < n-1$ (§ 2), so that

$$\dim E(\tau) = \dim W(\tau) = \begin{cases} n - (\tau - n + 1)/(\tau+1) & \text{when } \tau > n-1 \\ n & \text{when } \tau \leqslant n-1. \end{cases}$$

Since the exceptional sets $\hat{E}$ and E of frequencies $\omega_1, \ldots, \omega_n$ in $\mathbb{R}^n$ are given by

$$\hat{E} = \bigcap_\tau \hat{E}(\tau) \qquad E = \bigcap_\tau E(\tau)$$

respectively, it follows that

$$\dim \hat{E} = \lim_{\tau\to\infty} n - \frac{\tau - n}{\tau + 1} = n - 1$$

and

$$\dim E = \lim_{\tau\to\infty} n - \frac{\tau - n + 1}{\tau + 1} = n - 1.$$

Thus the exceptional sets $\hat{E}$ and E have Hausdorff codimension 1 in $\mathbb{R}^n$.

11. Conclusion

Returning to the KAM theory, we recall that the stability implied by very large values of τ is not significant physically. To prevent arbitrarily small perturbations, the value of the exponent τ in the exceptional set $E(\tau)$ should be a conveniently small number which is still large enough to ensure that the Lebesgue measure of $E(\tau)$ is zero. In the case of the autonomous KAM theorem, the exceptional set is chosen to be $E(n+1)$ (Arnol'd 1978, p 405). The Hausdorff dimension of $E(n+1)$ is $n-2/(n+2)$, so that for a system with a large number of degrees of freedom, $\dim E(n+1)$ is almost its maximal value n. More generally, fixing $\tau = n + c$, for some positive constant c, $\dim E(n+c) = n-(c+1)/(n+c+1)$ which approaches n when n is large compared to c. Similar considerations apply to the periodic KAM theorem: $\dim \hat{E}(n+1) = n-1/(n+2)$, which also approaches n when n is large. These observations are important for systems with an infinite number of degrees of freedom and reflect the difficulty of avoiding near resonance when the number of frequencies is large.

It should perhaps be stated again that the fact that a set of frequencies is in $E(\tau)$ or even E does not imply that the corresponding system does not have a quasi-periodic solution. However, it seems probable that the presence of near resonances is likely to lead to instabilities and, in fact, this has been shown to be the case in some examples (Arnol'd 1964, Brjuno 1970, Siegel and Moser 1971). Indeed the way it emerges in the general setting suggests that the Diophantine approximation condition is really required and is not just a technical restriction which could be circumvented using more powerful methods. The above results, which show that for a system with a large number of degrees of freedom the Hausdorff dimension of $E(n+1)$ is very nearly maximal, suggest that for such systems the instabilities associated with reasonable perturbations (corresponding to small values of τ) might be more prevalent than suspected.

The KAM theorem shows that under a sufficiently small perturbation, the 'good' non-resonant tori will persist and remain topologically tori and invariant. For the unperturbed Hamiltonian, the volume in phase space of the complementary set consisting of the everywhere dense set of 'bad' almost resonant tori, whose conservation cannot be guaranteed by the theorem, is small. When the Hamiltonian is perturbed slightly away from an integrable system, the 'bad' tori will typically break up into finite sets of alternating elliptic and hyperbolic periodic orbits which are surrounded by regions known as stochastic layers in which the motion is irregular. For a system with only two degrees of freedom, the level set of the energy is three-dimensional and is partitioned by the two-dimensional invariant tori, so that the motion is stable despite the disintegration of the bad tori. For a system with more than two degrees of freedom, the n-dimensional invariant tori no longer partition the $(2n-1)$-dimensional energy surface. The stochastic layers corresponding to different frequencies will intersect to form a dense tangled resonance network or 'web' covering the whole of the energy surface. The stochastic motion within each layer means that for initial conditions on this web the system will slowly move along the interconnected stochastic layers over the entire energy surface. This important process is called 'Arnol'd diffusion' (see Abraham and Marsden 1978) and was first shown to occur in a specific example due to Arnol'd (1964) who conjectured that the mechanism of this sort of instability is generic. Some recent work of Chirikov (1979), Holmes and Marsden (1982) and Vivaldi (1984) lends support to this conjecture and they give a criterion involving Melnikov integrals for the occurrence of this process. Indeed, Arnol'd diffusion has been observed in an experiment involving electrons confined in a magnetic bottle (Chirikov 1979) and in various numerical experiments (Lichtenberg and Lieberman 1983).

At first sight it is somewhat surprising that this effect is observed in real systems since it occurs only for very special initial conditions which lie in the exponentially narrow stochastic layers. However, the full set of such layers is everywhere dense in the phase space and as we have seen above for a system with a large number of degrees of freedom the Hausdorff dimension of the set of frequencies corresponding to bad tori is very nearly maximal. Thus, even though the set of almost resonant frequencies has Lebesgue measure 0, the set is very close to having positive Lebesgue measure. The break-up of the bad tori is likely to have qualitatively important effects on a sufficiently long timescale. The near-maximal Hausdorff dimension of the frequencies of the bad tori suggests that the volume of the corresponding trajectories in phase space might be somewhat larger than expected. Although not sufficient to establish the existence of Arnol'd diffusion, the relatively large volume of these trajectories might explain to some extent why Arnol'd diffusion is observed in real physical systems with several degrees of freedom and negligible energy dissipation. The Solar System is of this kind and Arnol'd diffusion can account for the Kirkwood gaps in the asteroid belt (Brjuno 1970).

Acknowledgments

We would like to thank the referees for useful comments which have improved the clarity of the paper.

References

Abraham R and Marsden J E 1978 *Foundations of Mechanics* (Reading, MA: Benjamin/Cummings)
Arnol'd V 1963 *Russ. Math. Surveys* **18** 85-192
—— 1964 *Sov. Math. Dokl.* **5** 581-4
—— 1968 *Russ. Math. Surveys* **23** 1-433
—— 1978 *Mathematical Methods of Classical Mechanics* (Engl. transl. Vogtmann K and Weinstein A (Berlin: Springer))
—— 1981 *Singularity Theory. LMS Lecture Notes Ser. 53* 3-45
—— 1983 *Geometric Methods in the Theory of Ordinary Differential Equations* (Engl. transl. Szucs J (Berlin: Springer))
Besicovitch A S 1934 *J. London Math. Soc.* **9** 126-31
Bovey J D and Dodson M M 1978 *Bull. London Math. Soc.* **10** 213-8
—— 1985 *Acta Arith.* **45** to appear
Brjuno A D 1970 *Math. USSR. Sbornik* **12** 272-312
Chirikov B V 1979 *Phys. Rep.* **52** 265-379

Eggleston H G 1952 *Proc. London Math. Soc.* **54** 42-93
Falconer K J 1985 *The Geometry of Fractal Sets* (Cambridge: Cambridge University Press)
Holmes P J and Marsden J E 1982 *J. Math. Phys.* **23** 135-66
Jarnik V 1929 *Math. Sbornik* **36** 371-82
—— 1931 *Math. Z.* **33** 505-43
Lichtenberg A J and Lieberman M A 1983 *Regular and Stochastic Motion* (Berlin: Springer)
Mandelbrot B B 1983 *The Fractal Geometry of Nature* (San Francisco: W H Freeman)
Moser J K 1973 *Stable and Random Motions in Dynamical Systems. Ann. Math. Studies No 77* (Princeton: Princeton University Press)
Rogers C A 1970 *Hausdorff Measures* (Cambridge: Cambridge University Press)
Rüssman H 1973 *Lecture Notes in Physics 38* ed J K Moser (Berlin: Springer) pp 598-624
Schmidt W M 1969 *J. Number Theory* **1** 139-54
—— 1980 *Diophantine Approximation. Lecture Notes in Mathematics* **785** (Berlin: Springer)
Siegel C L and Moser J K 1971 *Lectures on Celestial Mechanics* (Berlin: Springer)
Sprindžuk V G 1979 *Metric Theory of Diophantine Approximations* (Engl. transl. Silverman R A (New York: Wiley))
Sternberg S 1969 *Celestial Mechanics* Part II (New York: Benjamin)
Sullivan D 1982 *Acta Math.* **149** 215-37
Vickers J A G and Dodson M M 1985 in preparation
Vivaldi F 1984 *Rev. Mod. Phys.* **56** 737-76
Zehnder E 1977 *Geometry and Topology* ed J Palis and M do Carmo (*Lecture Notes in Mathematics 597*) (Berlin: Springer) pp 855-66