COMPOSITIO MATHEMATICA

# Waring's problem with restricted digits

Ben Green

Compositio Math. **161** (2025), 341–364.

doi: 10.1112/S0010437X24007723



FOUNDATION
COMPOSITIO
MATHEMATICA

LONDON
MATHEMATICAL
SOCIETY
EST. 1865

Compositio Math. 161 (2025) 341–364
doi:10.1112/S0010437X24007723

# Waring's problem with restricted digits

Ben Green

Abstract

Let $k \geqslant 2$ and $b \geqslant 3$ be integers, and suppose that $d_1, d_2 \in \{0, 1, \dots, b-1\}$ are distinct and coprime. Let $\mathcal{S}$ be the set of non-negative integers, all of whose digits in base $b$ are either $d_1$ or $d_2$. Then every sufficiently large integer is a sum of at most $b^{160k^2}$ numbers of the form $x^k$, $x \in \mathcal{S}$.

## Contents



## 1. Introduction

Let $k \geqslant 2$ be an integer. One of the most celebrated results in additive number theory is Hilbert's theorem that the $k$th powers are an asymptotic basis of finite order. That is, there is some $s$ such that every sufficiently large natural number can be written as a sum of at most $s$ $k$th powers of natural numbers.

One may ask whether a similar result holds if one passes to a subset $\{x^k : x \in \mathcal{S}\}$ of the full set of $k$th powers. This has been established in various cases, for instance, when $\mathcal{S}$ is the set of primes (the so-called Waring–Goldbach problem [KT05]), the set of smooth numbers with

Received 26 September 2023, accepted in final form 8 November 2024.
*2020 Mathematics Subject Classification* 11P05 (primary), 11A63 (secondary).
*Keywords:* Waring's problem; digits.

suitable parameters [DS16], the set of integers such that the sum of digits in base $b$ lies in some fixed residue class modulo $m$ [TT05], random sets with $\mathbf{P}(s \in \mathcal{S}) = s^{c-1}$ for some $c > 0$ [Vu00, Woo03a], or *all* sets with suitably large density [Sal21].

Our main result in this paper is that a statement of this type holds when $\mathcal{S}$ is the set of integers whose base $b$ expansion contains just two different (fixed) digits.

THEOREM 1.1. *Let $k \geqslant 2$ and $b \geqslant 3$ be integers, and suppose that $d_1, d_2 \in \{0, 1, \dots, b-1\}$ are distinct and coprime. Let $\mathcal{S}$ be the set of non-negative integers, all of whose digits in base $b$ are either $d_1$ or $d_2$. Then every sufficiently large integer is a sum of at most $b^{160k^2}$ numbers of the form $x^k$, $x \in \mathcal{S}$.*

*Remark.* While the basic form of the bound is the best the method gives, the constant 160 could certainly be reduced, especially for large values of $b$; I have not tried to optimise it. The restriction to $b \geqslant 3$ is helpful at certain points in the argument. Of course, the case $b = 2$ (in which case we must have $\{d_1, d_2\} = \{0, 1\}$) corresponds to the classical Waring problem, for which much better bounds are known.

Although Theorem 1.1 seems to be new, one should certainly mention in this context the interesting work of Biggs [Big21, Big23] and Biggs and Brandes [BB23], who showed that, for some $s$, every sufficiently large integer is a sum of at most $s$ numbers of the form $x^k$, $x \in \mathcal{S}$, and one further $k$th power. (In their work $b$ is taken to be prime and larger than $k$.)

This paper is completely independent of the work of Biggs and Brandes, but it seems plausible that by combining their methods with ours one could significantly reduce the quantity $b^{160k^2}$ in Theorem 1.1, at least for prime $b$.

Finally, we note that sets of integers whose digits in some base are restricted to some set are often called *ellipsephic*, a term coined by Mauduit, as explained in [Big21, Big23].

### 1.1 Notation

If $x \in \mathbf{R}$, we write $\|x\|$ for the distance from $x$ to the nearest integer. The only other time we use the double vertical line symbol is for certain box norms $\|\cdot\|_\square$, which occur in Appendix A. There seems little danger of confusion so we do not resort to more cumbersome notation such as $\|x\|_{\mathbf{R}/\mathbf{Z}}$. Write $e(x) = e^{2\pi i x}$.

If $X$ is a finite set and $f : X \to \mathbf{C}$ is a function then we write $\mathbf{E}_{x \in X} f(x) = 1/|X| \sum_{x \in X} f(x)$.

All intervals will be discrete. Thus, $[A, B]$ denotes the set of all *integers* $x$ with $A \leqslant x \leqslant B$ (and here $A, B$ need not be integers). We will frequently encounter the discrete interval $[0, m)$, for positive integer $m$, which is the same thing as the set $\{0, 1, \dots, m-1\}$. Note carefully that at some points in § 6, the notation $[m_1, m_2]$ will also refer to the lowest common multiple of two integers $m_1, m_2$.

Throughout the paper we will fix a base $b \geqslant 3$, an exponent $k \geqslant 2$ and distinct coprime digits $d_1, d_2 \in [0, b)$. Denote by $\mathcal{S}$ the set of all non-negative integers $x$, all of whose digits in base $b$ are $d_1$ or $d_2$. We include 0 in $\mathcal{S}$. Write $\mathcal{S}^k := \{x^k : x \in \mathcal{S}\}$. Note that $\mathcal{S}^k$ might more usually refer to the $k$-fold product set of $\mathcal{S}$ with itself, but we have no use for that concept here.

We will reserve the letter $n$ for a variable natural number, which we often assume is sufficiently large, and which it is usually convenient to take to be divisible by $k$. We always write $N = b^n$, so $[0, N)$ is precisely the set of non-negative integers with at most $n$ digits in base $b$.

If $n$ is a natural number, we define the map $L_b : \{0, 1\}^{[0,n)} \to \mathbf{Z}$ by

$$L_b(\mathbf{x}) := \sum_{i \in [0,n)} x_i b^i, \tag{1.1}$$



https://doi.org/10.1112/S0010437X24007723 Published online by Cambridge University Press

where $\mathbf{x} = (x_i)_{i\in[0,n)}$. Although this map depends on $n$, we will not indicate this explicitly, since the underlying $n$ will be clear from context. Then

$$\tfrac{d_1(b^n-1)}{b-1} + (d_2 - d_1)L_b(\mathbf{x}) \tag{1.2}$$

is the number whose base $b$ expansion has a $b^i$ digit equal to $d_1$ if $x_i = 0$, and $d_2$ if $x_i = 1$.

## 2. An outline of the argument

Unsurprisingly, given its pre-eminence in work on Waring's problem, the basic mode of attack is the Hardy–Littlewood circle method. Let $n \in \mathbf{N}$, set $N = b^n$ and consider the subset of $\mathcal{S}$ consisting of integers with precisely $n$ digits. This is a set of size $2^n$. Denote by $\mu_n$ the normalised probability measure on the set of $k$th powers of the elements of this set. That is, $\mu_n(m) = 2^{-n}$ if $m = (\sum_{i\in[0,n)} x_i b^i)^k$ with all $x_i \in \{d_1, d_2\}$ for all $i$, and $\mu_n(m) = 0$ otherwise. The Fourier transform $\widehat{\mu_n}(\theta) := \sum_{m\in\mathbf{Z}} \mu_n(m)e(m\theta)$ is then a normalised version of what is usually called the exponential sum or Weyl-type sum, and as expected for an application of the circle method, it plays a central role in our paper.

Our main technical result is the following, which might be called a log-free Weyl-type estimate for $k$th powers with restricted digits.

PROPOSITION 2.1. *Suppose that $k \geqslant 2$ and $b \geqslant 3$. Set $B := b^{6k^2}$. Suppose that $\delta \in (0,1)$ and that $k \mid n$. Suppose that $|\widehat{\mu_n}(\theta)| \geqslant \delta$ and that $N \geqslant (2/\delta)^B$, where $N := b^n$. Then there is a positive integer $q \leqslant (2/\delta)^B$ such that $\|\theta q\| \leqslant (2/\delta)^B N^{-k}$.*

*Remarks.* If $\mu_n$ is replaced by the normalised counting measure on $k$th powers less than $N$ without any digital restriction, a similar estimate is true and is very closely related to Weyl's inequality. The most standard proof of Weyl's inequality such as [Vau97, Lemma 2.4], however, results in some extra factors of $N^{o(1)}$ (from the divisor bound). 'Log-free' versions may be obtained by combining the standard result with major arc estimates as discussed, for example, in [Woo03b], or by modifying the standard proof of Weyl's inequality to focus on this goal rather than on the quality of the exponents, as done in [GT10, § 4]. Our treatment here is most closely related to this latter approach.

Although we will only give a detailed proof of Proposition 2.1 in the case that $\mu_n$ is the measure on $k$th powers of integers with just two fixed digits, similar arguments ought to give a more general result in which the digits are restricted to an arbitrary subset of $\{0, 1, \dots, b-1\}$ of size at least 2. This would be of interest if one wanted to obtain an asymptotic formula in Theorem 1.1, with more general digital restrictions of this type.

Experts will consider it a standard observation that Proposition 2.1 implies that $\mathcal{S}^k$ is an asymptotic basis of some finite order $s$. Roughly, this is because one can use it to obtain a moment estimate $\sum_x \mu_n^{(t)}(x)^2 = \int_0^1 |\widehat{\mu_n}(\theta)|^{2t} d\theta \ll N^{-k}$ for a suitably large $t$. Here, $\mu_n^{(t)}$ denotes the $t$-fold convolution power of $\mu_n$; see immediately after (3.1) for full details. The Cauchy–Schwarz inequality then implies that the $t$-fold sumset $t\mathcal{S}^k$ has positive density in an interval of length $\gg N^k$, whereupon methods of additive combinatorics can be used to conclude.

However, by itself this kind of argument leads to $s$ having a double-exponential dependence on $k$. The reason is that Proposition 2.1 is not very effective in the regime $\delta \approx 1$. It is possible that the proof could be adapted so as to be more efficient in this range, but this seems nontrivial. Instead we provide, in § 4, a separate argument that is at first sight crude, but turns out to be more efficient for this task. This gives the following result.



https://doi.org/10.1112/S0010437X24007723 Published online by Cambridge University Press

Proposition 2.2. *Let $n \in \mathbf{N}$ and let $N = b^n$. Suppose that $n \geqslant k$. Then the measure of all $\theta \in \mathbf{R}/\mathbf{Z}$ such that $|\widehat{\mu_n}(\theta)| \geqslant 1 - 1/4b^{-3k^2}$ is bounded above by $2b^{k^2}N^{-k}$.*

In fact, we obtain a characterisation of these values of $\theta$, much as in Proposition 2.1; see § 4 for the detailed statement and proof.

Details of how to estimate the moment $\int_0^1 |\widehat{\mu_n}(\theta)|^{2t}d\theta$ using Propositions 2.1 and 2.2, and of the subsequent additive combinatorics arguments leading to the proof of Theorem 1.1, may be found in § 3.

This leaves the task of proving Proposition 2.1, which forms the bulk of the paper, and is where the less standard ideas are required. For the purposes of this overview, we mostly consider the case $k = 2$, and for definiteness set $\{d_1, d_2\} = \{0, 1\}$.

*Decoupling.* The first step is a kind of decoupling. Recall the definitions of the maps $L_b$ (see (1.1)). The idea is to split the variables $\mathbf{x} = (x_i)_{i\in[0,n)}$ into the even variables $\mathbf{y} = (x_{2i})_{i\in[0,n/2)}$ and the odd variables $\mathbf{z} = (x_{2i+1})_{i\in[0,n/2)}$, assuming that $n$ is even for this discussion. We have $L_b(\mathbf{x}) = L_{b^2}(\mathbf{y}) + bL_{b^2}(\mathbf{z})$. Here, there is a slight abuse of notation in that $L_b$ is defined on vectors of length $n$, whilst $L_{b^2}$ is defined on vectors of length $n/2$. We then have

$$\begin{aligned}\widehat{\mu_n}(\theta) = \mathbf{E}_{\mathbf{x}\in\{0,1\}^{[0,n)}} e(\theta L_b(\mathbf{x})^2) &= \mathbf{E}_{\mathbf{y},\mathbf{z}\in\{0,1\}^{[0,n/2)}} e(\theta(L_{b^2}(\mathbf{y}) + bL_{b^2}(\mathbf{z})^2))\\ &= \mathbf{E}_{\mathbf{y},\mathbf{z}\in\{0,1\}^{[0,n/2)}} \Psi(\mathbf{y})\Psi'(\mathbf{z})e(2b\theta L_{b^2}(\mathbf{y})L_{b^2}(\mathbf{z})),\end{aligned}$$

where $\Psi(\mathbf{y}) = e(\theta L_{b^2}(\mathbf{y})^2)$ and $\Psi'(\mathbf{z}) = e(b^2\theta L_{b^2}(\mathbf{z})^2)$, but the precise form of these functions is not important in what follows. By two applications of the Cauchy–Schwarz inequality (see Appendix A for a general statement), we may eliminate the $\Psi$ and $\Psi'$ terms, each of which depends on just one of $\mathbf{y}, \mathbf{z}$. Assuming, as in the statement of Proposition 2.1, that $|\widehat{\mu_n}(\theta)| \geqslant \delta$, we obtain

$$\delta^4 \leqslant \mathbf{E}_{\mathbf{y},\mathbf{z},\mathbf{y}',\mathbf{z}'\in\{0,1\}^{[0,n/2)}} e(2b\theta(L_{b^2}(\mathbf{y})L_{b^2}(\mathbf{z}) - L_{b^2}(\mathbf{y}')L_{b^2}(\mathbf{z}) - L_{b^2}(\mathbf{y})L_{b^2}(\mathbf{z}') + L_{b^2}(\mathbf{y}')L_{b^2}(\mathbf{z}'))).$$

We remove the expectation over the dashed variables, that is to say, there is some choice of $\mathbf{y}', \mathbf{z}'$ for which the remaining average over $\mathbf{y}, \mathbf{z}$ is at least $\delta^4$. For simplicity of discussion, suppose that $\mathbf{y}' = \mathbf{z}' = 0$ is such a choice; then

$$\delta^4 \leqslant \mathbf{E}_{\mathbf{y},\mathbf{z}\in\{0,1\}^{[0,n/2)}} e(2b\theta L_{b^2}(\mathbf{y})L_{b^2}(\mathbf{z})). \tag{2.1}$$

At the expense of replacing $\delta$ by $\delta^4$, we have replaced the quadratic form $L_b(\mathbf{x})^2$ by a product of two linear forms in disjoint variables, which is a far more flexible object to work with. I remark that I obtained this idea from the proof of [CTV06, Theorem 4.3], which uses a very similar method.

Now, for fixed $\mathbf{z}$, the average over $\mathbf{y}$ in (2.1) can be estimated fairly explicitly. The conclusion is that for $\gg \delta^4 2^{n/2}$ values of $\mathbf{z}$, $2b\theta L_{b^2}(\mathbf{z})$ has $\ll \log(1/\delta)$ nonzero base $b$ digits, among the first $n$ digits after the radix point. Here, we use the *centred* base $b$ expansion in which digits lie in $(-b/2, b/2]$, discussed in more detail in § 5.

*Additive expansion.* The output of the decoupling step is an assertion to the effect that, for $m$ in a somewhat large set $\mathscr{M} \subset \{1, \dots, N\}$, $\theta m$ has very few nonzero digits in base $b$ among the first $n$ after the radix point. The set $\mathscr{M}$ is the set of $2bL_{b^2}(\mathbf{z})$ for $\gg \delta^4 2^{n/2}$ values of $\mathbf{z} \in \{0, 1\}^{[0,n/2)}$, and so has size $\sim N^{(\log 2)/2\log b}$ that, though 'somewhat large', is unfortunately appreciably smaller than $N$.



https://doi.org/10.1112/S0010437X24007723 Published online by Cambridge University Press

The next step of the argument is to show that the sum of a few copies of $\mathscr{M}$ is a considerably larger set, of size close to $N$. In fact, in the case $k=2$ under discussion, $b^2-1$ copies will do. This follows straightforwardly from the following result from the literature.

THEOREM 2.3. *Let $r, n \in \mathbf{N}$. Suppose that $A_1, \dots, A_r \subseteq \{0,1\}^n$ are sets with densities $\alpha_1, \dots, \alpha_r$. Then $A_1 + \cdots + A_r$ has density at least $(\alpha_1 \cdots \alpha_r)^\gamma$ in $\{0,1,\dots,r\}^n$, where $\gamma := r^{-1}\log_2(r+1)$.*

This theorem, which came from the study of Cantor-type sets in the 1970s and 1980s, seems not to be well known in modern-day additive combinatorics. The result has a somewhat complicated history, with contributions by no fewer than 10 authors, and I am unsure exactly how to attribute it. For comments and references pertinent to this, see Appendix B.

We remark that, for $k>2$, a considerably more elaborate argument is required at this point, and this occupies the bulk of § 6.

The conclusion is that $\theta m$ has $\ll \log(1/\delta)$ nonzero base $b$ digits among the first $n$ after the radix point, for all $m$ in a set $\mathscr{M} \subset \{1,\dots,N\}$ of size $\gg \delta^C N$.

*From digits to diophantine.* In the final step of the argument we extract the required diophantine conclusion (that is, the conclusion of Proposition 2.1) from the digital condition just obtained. The main ingredient is a result on the additive structure of sets with few nonzero digits, which may potentially have other uses. Recall that if $A$ is a set of integers then $E(A)$, the additive energy of $A$, is the number of quadruples $(a_1,a_2,a_3,a_4) \in A\times A\times A\times A$ with $a_1+a_2=a_3+a_4$.

PROPOSITION 2.4. *Let $r \in \mathbf{Z}_{\geqslant 0}$. Suppose that $A \subset \mathbf{Z}$ is a finite set, all of whose elements have at most $r$ nonzero digits in their centred base $b$ expansion. Then $E(A) \leqslant (2b)^{4r}|A|^2$.*

The proof of this involves passing to a quadripartite formulation (that is, with four potentially different sets $A_1, A_2, A_3, A_4$, and also allowing for the possibility of a 'carry' in the additive quadruples) and an inductive argument.

The final deduction of Proposition 2.1 uses this and some fibring arguments. This, and the proof of Proposition 2.4, may be found in § 7.

## 3. Reduction to a log-free Weyl-type estimate

In this section we show that our main result, Theorem 1.1, follows from the log-free Weyl-type estimate, Proposition 2.1. We begin by stating two results about growth under set addition. The first is a theorem of Nathanson and Sárközy [NS89, Theorem 1].

THEOREM 3.1. *Let $X \in \mathbf{N}$ and $r \in \mathbf{N}$. Suppose that $A \subset \{1,\dots,X\}$ is a set of size $\geqslant 1 + X/r$. Then there is an arithmetic progression of common difference $d$, $1 \leqslant d \leqslant r-1$ and length at least $\lfloor X/2r^2 \rfloor$ contained in $4rA$.*

*Proof.* In [NS89, Theorem 1], take $h=2r$, $z=\lfloor X/2r^2\rfloor$; the result is then easily verified. □

The second result we will need is a simple but slightly fiddly lemma on repeated addition of discrete intervals.

LEMMA 3.2. *Let $X \geqslant 1$ be real and suppose that $I \subset [0,X)$ is a discrete interval of length $L \geqslant 2$. Set $\eta := L/X$. Let $K \geqslant 4$ be a parameter. Then $\bigcup_{j \leqslant \lceil 2K/\eta^2\rceil} jI$ contains the discrete interval $[4/\eta X, K/\eta X]$.*



https://doi.org/10.1112/S0010437X24007723 Published online by Cambridge University Press

*Proof.* Write $I=[x_0,x_0+L-1]$, where $x_0\in\mathbf{Z}_{\geqslant 0}$. Then $jI=[jx_0,jx_0+j(L-1)]$. Note that if $j\geqslant x_0/(L-1)$, we have $jx_0+j(L-1)\geqslant (j+1)x_0$, and so the interval $(j+1)I$ overlaps the interval $jI$. Therefore, if we set $j_0:=\lceil x_0/(L-1)\rceil$, for any $j_1\geqslant j_0$, the union $I^*:=\bigcup_{j_0\leqslant j\leqslant j_1} jI$ is a discrete interval. Set $j_1:=\lceil 2K/\eta^2\rceil$. We have

$$\min I^* = j_0x_0 \leqslant \lceil \tfrac{X}{L-1}\rceil X \leqslant \lceil \tfrac{2X}{L}\rceil X \leqslant \tfrac{4X^2}{L} = \tfrac{4}{\eta}X,$$

and

$$\max I^* \geqslant j_1(L-1) \geqslant \tfrac{2K}{\eta^2}\tfrac{L}{2} = \tfrac{K}{\eta}X.$$

This concludes the proof. □

*Proof of Theorem 1.1, assuming Proposition 2.1.* Let $n$ be some large multiple of $k$ and consider the measure $\mu_n$ as described in § 2. Thus, $\mu_n$ is supported on $\mathcal{S}^k\cap[0,N^k)$, where $N=b^n$. Set

$$t:=8b^{9k^2}, \tag{3.1}$$

and write $\mu_n^{(t)}$ for the $t$-fold convolution power of $\mu_n$, that is to say,

$$\mu_n^{(t)}(x)=\sum_{x_1+\cdots+x_t=x}\mu_n(x_1)\cdots\mu_n(x_t).$$

Then $\widehat{\mu_n^{(t)}}=(\widehat{\mu_n})^t$ and so by Parseval's identity and the layer-cake representation,

$$\sum_x \mu_n^{(t)}(x)^2=\int_0^1 |\widehat{\mu_n}(\theta)|^{2t}d\theta = 2t\int_0^1 \delta^{2t-1}\,\mathrm{meas}\{\theta:|\widehat{\mu_n}(\theta)|\geqslant\delta\}d\delta = 2t(I_1+I_2+I_3), \tag{3.2}$$

where $I_1,I_2,I_3$ are the integrals over ranges $[0,2N^{-1/B}]$, $[2N^{-1/B},1-c]$ and $[1-c,1]$, respectively, with $c:=\frac{1}{4}b^{-3k^2}$, $B=b^{6k^2}$ (as in Proposition 2.1) and meas is the Lebesgue measure on the circle $\mathbf{R}/\mathbf{Z}$. We have, for $N$ large,

$$I_1\leqslant (2N^{-1/B})^{2t-1}<N^{-k}.$$

To bound $I_2$, we use Proposition 2.1, which tells us that the set $\{\theta\in\mathbf{R}/\mathbf{Z}:|\widehat{\mu_n}(\theta)|\geqslant\delta\}$ is contained in the set $\{\theta\in\mathbf{R}/\mathbf{Z}:\|\theta q\|\leqslant(2/\delta)^BN^{-k}$ for some positive $q\leqslant(2/\delta)^B\}$, and so $\mathrm{meas}\{\theta:|\widehat{\mu_n}(\theta)|\geqslant\delta\}\leqslant 2(2/\delta)^{2B}N^{-k}$. Since $2t-1-2B\geqslant t$, we therefore have

$$I_2\leqslant 2N^{-k}\int_0^{1-c}\delta^{2t-1}(2/\delta)^{2B}d\delta\leqslant 2N^{-k}(1-c)^t2^{2B}<N^{-k}.$$

For the last inequality, we used the fact that $t=2B/c$ and so $(1-c)^t\leqslant e^{-2B}$.

Finally, to bound $I_3$, we use Proposition 2.2, which immediately implies that

$$I_3\leqslant 2b^{k^2}N^{-k}.$$

Substituting these bounds for $I_1,I_2$ and $I_3$ into (3.2), we obtain that, for sufficiently large $N$, $\sum_x\mu_n^{(t)}(x)^2\leqslant 4tb^{k^2}N^{-k}=32b^{10k^2}N^{-k}$. On the other hand, it follows by the Cauchy–Schwarz inequality and the fact that $\sum_x\mu_n^{(t)}(x)=1$ that $1\leqslant|\operatorname{Supp}(\mu_n^{(t)})|\sum_x\mu_n^{(t)}(x)^2$, and so $|\operatorname{Supp}(\mu_n^{(t)})|\geqslant 2^{-5}b^{-10k^2}N^k$. Thus, since $\mu_n^{(t)}$ is supported on the $t$-fold sumset of $\mathcal{S}^k\cap[0,N^k)$, we see that $|t\mathcal{S}^k\cap[0,tN^k)|\geqslant 2^{-5}b^{-10k^2}N^k$. Applying Theorem 3.1 with $X=tN^k$ and $r=2^8b^{19k^2}$, we see that $4rt\mathcal{S}^k\cap[0,4rtN^k)$ contains an arithmetic progression $P$ of common difference $<r$ and length $|P|\geqslant L:=2^{-15}b^{-29k^2}N^k$.

Since $d_1^k$ and $d_2^k$ are coprime, every number greater than or equal to $(d_1^k-1)(d_2^k-1)<b^{2k}<r$ is a non-negative integer combination of these numbers. Therefore, it is certainly the case that $2r\mathcal{S}^k$ contains $[r,2r)$. Since the common difference of $P$ is less than $r$, $P+[r,2r)$ contains a



https://doi.org/10.1112/S0010437X24007723 Published online by Cambridge University Press

discrete interval $I$ of length $\geqslant L$. This interval is therefore contained in $(4rt+2r)\mathcal{S}^k \subset 8rt\mathcal{S}^k$. Note that by construction $I \subset [0, 8rtN^k)$.

Apply Lemma 3.2, taking $X = X(n) = 8rtN^k$, $\eta = \frac{L}{X} = 2^{-29}b^{-57k^2}$ and $K = 4b^{k^2}$. Since $\mathcal{S}$ contains 0, we see that $\lfloor 2K/\eta^2 \rfloor 8rt\mathcal{S}^k = 2^{75}b^{142k^2}\mathcal{S}^k$ contains the interval $I_n := [\frac{4}{\eta}X(n), \frac{K}{\eta}X(n)]$. Remember that here $n$ is any sufficiently large multiple of $k$. By the choice of $K$, $\frac{K}{\eta}X(n) = \frac{4}{\eta}X(n+k)$, and so these intervals overlap. Thus, $\bigcup_n I_n$ consists of all sufficiently large integers, and hence, so does $2^{75}b^{142k^2}\mathcal{S}^k$. Finally, one may note that $2^{75} < b^{12k^2}$ for $b \geqslant 3$ and $k \geqslant 2$. □

## 4. Very large values of the Fourier transform

In this section we establish Proposition 2.2. We will in fact establish the following more precise result.

PROPOSITION 4.1. *Let $n \in \mathbf{N}$ and let $N = b^n$. Suppose that $n \geqslant k$. Let $\theta \in \mathbf{R}/\mathbf{Z}$. Suppose that $|\widehat{\mu_n}(\theta)| \geqslant 1 - 1/4b^{-3k^2}$. Then there is a positive integer $q \leqslant (2k!)b^{1/2k(k-1)+1}$ such that $\|\theta q\| \leqslant (2k!)^{-1}b^{1/2k(k+1)-1}N^{-k}$.*

Proposition 2.2 is a consequence of this and the observation that the measure of $\theta \in \mathbf{R}/\mathbf{Z}$ such that $\|\theta q\| \leqslant \varepsilon$ for some positive integer $q \leqslant q_0$ is bounded above by $2\varepsilon q_0$.

*Proof of Proposition 4.1.* Set $Q := 2k!b^{k(k-1)/2+1}$. Note that, since $2k! \leqslant 2^{k^2/2} \leqslant b^{k^2/2}$ for all $b, k \geqslant 2$, we have $Q \leqslant b^{k^2}$. By Dirichlet's theorem, there is some positive integer $q \leqslant Q$ and an $a$, coprime to $q$, such that $|\theta - a/q| \leqslant 1/qQ$. Set $\eta := \theta - a/q$, thus $|\eta| \leqslant 1/qQ$. There is a unique integer $j$ such that

$$\tfrac{1}{2bq} < |(d_2 - d_1)k!b^j\eta| \leqslant \tfrac{1}{2q}. \tag{4.1}$$

Now if we had $j < k(k-1)/2$ then

$$|(d_2 - d_1)k!b^j\eta| \leqslant (b-1)k!b^{1/2k(k-1)-1}|\eta| < k!b^{1/2k(k-1)}/qQ = 1/2bq,$$

contrary to (4.1). If $j > kn - k(k+1)/2$ then, using (4.1),

$$\|\theta q\| = |\eta q| \leqslant |2(d_2 - d_1)k!b^j|^{-1} \leqslant (2k!)^{-1}b^{1/2k(k+1)-1}N^{-k},$$

in which case the conclusion of the proposition is satisfied.

Suppose, then, that $k(k-1)/2 \leqslant j \leqslant kn - k(k+1)/2$. Then there is a set $I \subset [0, n)$, $|I| = k$, such that $j = \sum_{i\in I} i$. As usual, write $\mathbf{x} = (x_i)_{i\in[0,n)}$. It is convenient to write $\mathbf{x}_I$ for the variables $x_i$, $i \in I$ and $\mathbf{x}_{[0,n)\setminus I}$ for the other variables. For any fixed choice of $\mathbf{x}_{[0,n)\setminus I}$, we can write, setting $u := d_1(b^n - 1)/b - 1$,

$$\begin{aligned}(u + (d_2 - d_1)L_b(\mathbf{x}))^k &= \left(u + (d_2 - d_1)\sum_{i\in[0,n)} x_i b^i\right)^k \\ &= (d_2 - d_1)k!b^j \prod_{i\in I} x_i + \sum_{i\in I} \psi_i(\mathbf{x}_{[0,n)\setminus I}; \mathbf{x}_I)\end{aligned}$$

for some functions $\psi_i$, where $\psi_i$ does not depend on $x_i$. It follows that

$$\begin{aligned}|\widehat{\mu}_n(\theta)| &= |\mathbf{E}_{\mathbf{x}\in\{0,1\}^{[0,n)}} e((u + (d_2 - d_1)L_b(\mathbf{x}))^k)| \\ &\leqslant \mathbf{E}_{\mathbf{x}_{[0,n)\setminus I}\in\{0,1\}^{[0,n)\setminus I}} \left|\mathbf{E}_{\mathbf{x}_I\in\{0,1\}^I} \prod_{i\in I} \Psi_i(\mathbf{x}_{[0,n)\setminus I}; \mathbf{x}_I) e\left((d_2 - d_1)k!b^j\theta \prod_{i\in I} x_i\right)\right|,\end{aligned}$$



https://doi.org/10.1112/S0010437X24007723 Published online by Cambridge University Press

where $\Psi_i := e(\psi_i)$ is a 1-bounded function, not depending on $x_i$. By Proposition A.2 (and the accompanying definition of the Box norm, Definition A.1) it follows that

$$|\widehat{\mu_n}(\theta)|^{2^k} \leqslant \mathbf{E}_{\mathbf{x}_I,\mathbf{x}'_I\in\{0,1\}^I} e\Big((d_2-d_1)k!b^j\theta \prod_{i\in I}(x_i-x'_i)\Big).$$

(The right-hand side here is automatically a non-negative real number). On the right, we now bound all the terms trivially (by 1) except for two: the term with $x_i = x'_i = 0$ for all $i \in I$, and the term with $x_i = 1$ and $x'_i = 0$ for all $i \in I$. This gives, using the inequality $2 - |1 + e(t)| = 4\sin^2 \pi\|t\|/2 \geqslant 4\|t\|^2$,

$$\begin{aligned}|\widehat{\mu_n}(\theta)|^{2^k} &\leqslant 1 - \tfrac{2}{4^k} + \tfrac{1}{4^k}|1 + e((d_2-d_1)k!b^j\theta)| \\ &\leqslant 1 - 2^{2-2k}\|(d_2-d_1)k!b^j\theta\|^2. \end{aligned} \tag{4.2}$$

There are now two slightly different cases, according to whether or not $q \mid (d_2-d_1)k!b^j a$. If this is the case, then by (4.1),

$$\|(d_2-d_1)k!b^j\theta\| = |(d_2-d_1)k!b^j\eta| \geqslant 1/2bq.$$

If, on the other hand, $q \nmid (d_2-d_1)k!b^j a$ then by (4.1) we have

$$\|(d_2-d_1)k!b^j\theta\| \geqslant \tfrac{1}{q} - |(d_2-d_1)k!b^j\eta| \geqslant \tfrac{1}{2q}.$$

In both cases, $\|(d_2-d_1)k!b^j\theta\| \geqslant 1/2bQ = (4k!)^{-1}b^{-2-\frac{1}{2}k(k-1)}$. It follows from (4.2) that

$$|\widehat{\mu_n}(\theta)| \leqslant (1 - 2^{2-2k}(4k!)^{-2}b^{-4-k(k-1)})^{1/2^k} < 1 - \tfrac{1}{4}b^{-3k^2},$$

that is to say, the hypothesis of the proposition is not satisfied. Here, the second inequality follows from the Bernoulli inequality $(1-x)^{1/2^k} \leqslant 1 - x/2^k$ and the crude bounds $k! \leqslant b^{k^2/4}$, $2^{3k} \leqslant b^{2k}$, both valid for $b \geqslant 3$ and $k \geqslant 2$. □

## 5. Decoupling

We now turn to the somewhat lengthy task of proving Proposition 2.1. In this section we give the details of what we called the decoupling argument in the outline of § 2. The main result of the section is Proposition 5.2 below. We begin with a definition.

Definition 5.1. Let $\alpha \in \mathbf{R}/\mathbf{Z}$. Then we define

$$\tilde{\mathrm{w}}_n(\alpha) := \sum_{i\in[0,n)} \|\alpha b^i\|^2. \tag{5.1}$$

The reason for the notation is that $\tilde{\mathrm{w}}_n(\alpha)$ is closely related to the more natural quantity $\mathrm{w}_n(\alpha)$, which is the number of nonzero digits among the first $n$ digits after the radix point in the (centred) base $b$ expansion of $\alpha$. For a careful definition of this, see § 7. However, $\tilde{\mathrm{w}}_n$ has more convenient analytic properties.

Now we come to the main result of the section. As we said before, it is a little technical to state. However, it is rather less technical in the case $k = 2$, in which case the reader may wish to compare it with the outline in § 2.

Proposition 5.2. *Let $n \in \mathbf{N}$ be divisible by $k$ and set $N := b^n$. Suppose that $\delta \in (0, 1]$ and that $|\widehat{\mu_n}(\theta)| \geqslant \delta$. Then there are $t_1, \dots, t_{k-1} \in \mathbf{Z}$ with $|t_j| \leqslant N$ for all $j$ and a positive integer $q_0 \leqslant b^{k^2}$ such that, for at least $1/2\delta^{2^k}2^{(k-1)n/k}$ choices of $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k-1)} \in \{0,1\}^{[0,n/k)}$, we have*



https://doi.org/10.1112/S0010437X24007723 Published online by Cambridge University Press

$$\tilde{\mathrm{w}}_n\left(\theta q_0 \prod_{i=1}^{k-1}(L_{b^k}(\mathbf{x}^{(i)})+t_i)\right) \leqslant 2^k b^{2k}\log(2/\delta).$$

*Proof.* By (1.2) and the definition of the measure $\mu_n$, we have

$$\widehat{\mu_n}(\theta) = \mathbf{E}_{\mathbf{x}\in\{0,1\}^n} e(\theta(u+(d_2-d_1)L_b(\mathbf{x}))^k), \tag{5.2}$$

where $u := d_1(b^n-1)/(b-1)$. The first stage of the decoupling procedure is to split the variables $\mathbf{x}$ into $k$ disjoint subsets of size $n/k$. If $\mathbf{x} = (x_i)_{i\in[0,n)} \in \{0,1\}^{[0,n)}$, for each $j\in[0,k)$, we write $\mathbf{x}^{(j)} = (x_{ik+j})_{i\in[0,n/k)} \in \{0,1\}^{[0,n/k)}$. Then

$$L_b(\mathbf{x}) = \sum_{j\in[0,k)} b^j L_{b^k}(\mathbf{x}^{(j)}). \tag{5.3}$$

(Note here that $L_b$ is defined on $\{0,1\}^{[0,n)}$, whereas $L_{b^k}$ is defined on $\{0,1\}^{[0,n/k)}$.) By (5.2) we have

$$\widehat{\mu_n}(\theta) = \mathbf{E}_{\mathbf{x}^{(0)},\dots,\mathbf{x}^{(k-1)}\in\{0,1\}^{[0,n/k)}} e\left(\theta\left(u+(d_2-d_1)\sum_{i\in[0,k)} b^i L_{b^k}(\mathbf{x}^{(i)})\right)^k\right).$$

Expanding out the $k$th power and collecting terms, this can be written as

$$\mathbf{E}_{\mathbf{x}^{(0)},\dots,\mathbf{x}^{(k-1)}\in\{0,1\}^{[0,n/k)}}\left(\prod_{j\in[0,k)}\Psi_j(\mathbf{x})\right) e\left(\theta q_0 \prod_{i\in[0,k)} L_{b^k}(\mathbf{x}^{(i)})\right),$$

where

$$q_0 := k!(d_2-d_1)^k b^{k(k-1)/2}$$

and $\Psi_j$ is some 1-bounded function of the variables $\mathbf{x}^{(i)}$, $i\in[0,k)\setminus\{j\}$, the precise nature of which does not concern us. The inequality $q_0 \leqslant b^{k^2}$ follows using $|d_1-d_1|\leqslant b$ and the estimate $k!\leqslant 3^{k(k-1)/2}$, since $b\geqslant 3$.

One may now apply the Cauchy–Schwarz inequality $k$ times to eliminate the functions $\Psi_j$ in turn. This procedure is well known from the theory of hypergraph regularity [Gow07] or from the proofs of so-called generalised von Neumann theorems in additive combinatorics [GT10]. For a detailed statement, see Proposition A.2. From this it follows that

$$\delta^{2^k} \leqslant \mathbf{E} e\left(\theta q_0 \sum_{\omega\in\{0,1\}^{[0,k)}} (-1)^{|\omega|} \prod_{i\in[0,k)} L_{b^k}(\mathbf{x}_{\omega_i}^{(i)})\right),$$

where the average is over $\mathbf{x}_0^{(0)},\dots,\mathbf{x}_0^{(k-1)},\mathbf{x}_1^{(0)},\dots,\mathbf{x}_1^{(k-1)}\in\{0,1\}^{[0,n/k)}$, and we write $\omega = (\omega_i)_{i\in[0,k)}$ and $|\omega| = \sum_{i=1}^k |\omega_i|$. By pigeonhole there is some choice of $\mathbf{x}_1^{(0)},\dots,\mathbf{x}_1^{(k-1)}$ such that the remaining average over $\mathbf{x}_0^{(0)},\dots,\mathbf{x}_0^{(k-1)}$ is at least $\delta^{2^k}$. This may be written as

$$\delta^{2^k} \leqslant |\mathbf{E}_{\mathbf{x}^{(0)},\dots,\mathbf{x}^{(k-1)}\in\{0,1\}^{[0,n/k)}} e\left(\theta q_0 \prod_{i\in[0,k)}(L_{b^k}(\mathbf{x}^{(i)})+t_i)\right)|,$$

where $t_i := -L_{b^k}(\mathbf{x}_1^{(i)})$. It follows that, for at least $1/2\delta^{2^k}2^{(k-1)n/k}$ choices of $\mathbf{x}^{(1)},\dots,\mathbf{x}^{(k-1)}\in\{0,1\}^{[0,n/k)}$, we have

$$|\mathbf{E}_{\mathbf{x}^{(0)}\in\{0,1\}^{[0,n/k)}} e\left(\theta q_0 L_{b^k}(\mathbf{x}^{(0)})\prod_{i=1}^{k-1}(L_{b^k}(\mathbf{x}^{(i)})+t_i)\right)| \geqslant \delta^{2^k}/2. \tag{5.4}$$



https://doi.org/10.1112/S0010437X24007723 Published online by Cambridge University Press

Let $\alpha \in \mathbf{R}/\mathbf{Z}$ be arbitrary. Note that

$$\tilde{\mathrm{w}}_n(\alpha) = \sum_{i\in[0,n-1)} \|\alpha b^i\|^2 = \sum_{j\in[0,k)} \sum_{i\in[0,n/k)} \|\alpha b^{j+ik}\|^2$$
$$\leqslant \Bigl(\sum_{j\in[0,k)} b^{2j}\Bigr) \sum_{i\in[0,n/k)} \|\alpha b^{ik}\|^2 \leqslant b^{2k} \sum_{i\in[0,n/k)} \|\alpha b^{ik}\|^2.$$

Therefore, using the inequality $|1+e(t)| = 2|\cos(\pi t)| \leqslant 2\exp(-\|t\|^2)$, we have

$$|\mathbf{E}_{\mathbf{y}\in\{0,1\}^{[0,n/k)}} e(\alpha L_{b^k}(\mathbf{y}))| = \prod_{i\in[0,n/k)} \left|\frac{1+e(\alpha b^{ik})}{2}\right|$$
$$\leqslant \exp\Bigl(-\sum_{i\in[0,n/k)} \|\alpha b^{ik}\|^2\Bigr) \leqslant \exp(-b^{-2k}\tilde{\mathrm{w}}_n(\alpha)).$$

Combining this with (5.4), Proposition 5.2 follows. □

## 6. Sums of products of linear forms

We now turn to the next step of the outline in § 2, which we called additive expansion. The main result of the previous section, Proposition 5.2, is roughly of the form 'for quite a few $m \sim N^{k-1}$, $\tilde{\mathrm{w}}_n(\theta m) \lesssim \log(2/\delta)$'. (The reader should not attach any precise meaning to the symbols $\sim, \lesssim$ here.) The shortcoming of the statement as it stands is that the set of $m$ is of size $\sim 2^{(k-1)n/k}$, which is substantially smaller than $N^{k-1}$ (recall that $N = b^n$). The aim of this section is to upgrade the conclusion of Proposition 5.2 to get a much larger set of $m$. Here is the statement we will prove.

Proposition 6.1. *Set $C := b^{7k^2/2}$. Suppose that $\delta \in (0,1]$ and that $k \mid n$. Suppose that $|\widehat{\mu_n}(\theta)| \geqslant \delta$ and that $N \geqslant (2/\delta)^C$, where $N := b^n$. Then for at least $(\delta/2)^C N^{k-1}$ values of $m$, $|m| \leqslant CN^{k-1}$, we have $\tilde{\mathrm{w}}_n(\theta m) \leqslant C\log(2/\delta)$.*

The basic idea of the proof is to take sums of a few copies of the set of $m$ produced in Proposition 5.2 that (it turns out) expands this set of $m$ dramatically, whilst retaining the property of $\tilde{\mathrm{w}}_n(\theta m)$ being small.

We assemble some ingredients. The key input is Theorem 2.3 (see, in addition to § 2, Appendix B). We will also require some other lemmas of a miscellaneous type, and we turn to these first.

Lemma 6.2. *Let $\varepsilon, U, V$ be real parameters with $0 < \varepsilon \leqslant 2^{-44}$ and $U, V \geqslant 64/\varepsilon$. Suppose that $\Omega \subset [-U,U]\times[-V,V]$ has size at least $\varepsilon UV$. Then at least $\varepsilon^7 UV$ integers $n \in [-2UV, 2UV]$ may be written as $u_1v_1 + u_2v_2$ with $(u_1,v_1),(u_2,v_2) \in \Omega$.*

*Proof.* The conclusion is invariant under applying any of the four involutions $(u,v) \mapsto (\pm u, \pm v)$ to $\Omega$, so without loss of generality we may suppose that $\Omega \cap ([0,U]\times[0,V])$ has size at least $\varepsilon UV/4$. It then follows that $\Omega \cap ([\varepsilon U/32, U]\times[\varepsilon V/32, V])$ has size at least $\varepsilon UV/8$. Covering this box by disjoint dyadic boxes $[2^i, 2^{i+1})\times[2^j, 2^{j+1})$ contained in $[\varepsilon U/64, 2U]\times[\varepsilon V/64, 2V]$, we see that there is some dyadic box $[U', 2U')\times[V', 2V')$, $\varepsilon U/64 \leqslant U' \leqslant U$, $\varepsilon V/64 \leqslant V' \leqslant V$, on which the density of $\Omega$ is at least $\varepsilon/32$. Without loss of generality, suppose that $U' \leqslant V'$, and set $X := U'V' \geqslant 1$. Set $\Omega' := \Omega \cap ([U', 2U')\times[V', 2V'))$.

For $n \in \mathbf{Z}$, denote by $r(n)$ the number of representations of $n$ as $u_1v_1 + u_2v_2$ with $(u_1,v_1)$ and $(u_2,v_2)$ in $\Omega'$, and by $\tilde{r}(n)$ the number of representations as $u_1v_1 + u_2v_2$ with $(u_1,v_1),(u_2,v_2) \in$



https://doi.org/10.1112/S0010437X24007723 Published online by Cambridge University Press

$[U', 2U') \times [V', 2V')$. Thus, $r(n) \leqslant \tilde{r}(n)$. By the Cauchy–Schwarz inequality,

$$(\varepsilon X/32)^4 \leqslant |\Omega'|^4 = \Big(\sum_n r(n)\Big)^2 \leqslant |\operatorname{Supp}(r)| \sum_n \tilde{r}(n)^2. \tag{6.1}$$

Now, denoting by $\nu(n)$ the number of divisors of $n$ in the range $[U', 2U')$,

$$\tilde{r}(n) \leqslant \sum_{m \leqslant 4X} \nu(m)\nu(n-m) = \sum_{\substack{d,e \in [U',2U') \\ (d,e)|n}} \sum_{\substack{m \leqslant 4X \\ d|m, e|n-m}} 1 \leqslant 8X \sum_{\substack{d,e \in [U',2U') \\ (d,e)|n}} \tfrac{1}{[d,e]}.$$

Here, in the last step we used the fact that the set of $m$ satisfying $d \mid m$ and $e \mid n-m$ is a single residue class modulo $[d,e]$ (the lowest common multiple of $d$ and $e$), whose intersection with the interval $[1, 4X)$ has size $\leqslant 1 + 4X/[d,e] \leqslant 8X/[d,e]$ since $[d,e] \leqslant (2U')^2 \leqslant 4X$.

Setting $\delta := (d,e)$ and $d = \delta d'$, $e = \delta e'$, so that $[d,e] = \delta d'e'$, it then follows that

$$\tilde{r}(n) \leqslant 8X \sum_{\delta|n} \tfrac{1}{\delta} \sum_{d',e' \in [U'/\delta, 2U'/\delta)} \tfrac{1}{d'e'} \leqslant 8X \sum_{\delta|n} \tfrac{1}{\delta}.$$

Since $\tilde{r}(n)$ is supported where $n \leqslant 8X$, we have

$$\begin{aligned}\sum_n \tilde{r}(n)^2 &\leqslant (8X)^2 \sum_{n \leqslant 8X} \Big(\sum_{\delta|n} \frac{1}{\delta}\Big)^2 = (8X)^2 \sum_{\delta_1,\delta_2 \leqslant 8X} \frac{1}{\delta_1\delta_2} \sum_{n \leqslant 8X} 1_{[\delta_1,\delta_2]|n} \\ &\leqslant (8X)^2 \sum_{\delta_1,\delta_2 \leqslant 8X} \tfrac{1}{\delta_1\delta_2} \Big(\frac{8X}{[\delta_1,\delta_2]} + 1\Big).\end{aligned}$$

The contribution from the $+1$ term is $\leqslant (8X)^2(1+\log 8X)^2 < 2^{10}X^3$, since $X \geqslant 1$. Since $[\delta_1,\delta_2] \geqslant \sqrt{\delta_1\delta_2}$, the contribution from the main term is $\leqslant 2^8\zeta(\frac{3}{2})^2X^3 < 2^{11}X^3$. It follows that $\sum_n \tilde{r}(n)^2 \leqslant 2^{12}X^3$. Comparing with (6.1), we obtain $|\operatorname{Supp}(r)| \geqslant 2^{-32}\varepsilon^4 X \geqslant 2^{-44}\varepsilon^6 UV$. Since we are assuming that $\varepsilon \leqslant 2^{-44}$, this is at least $\varepsilon^7 UV$, and the proof is complete. □

LEMMA 6.3. *Let $X \geqslant 1$ be real, and suppose that $S_1, \dots, S_t \subseteq [-X, X]$ are sets of integers with $|S_i| \geqslant \eta X$. Then $|\bigcap_{i=1}^t (S_i - S_i)| \geqslant (\eta/5)^t X$.*

*Proof.* We have

$$\sum_{h_2,\dots,h_t} \Big(\sum_x 1_{S_1}(x) 1_{S_2}(x+h_2) \cdots 1_{S_t}(x+h_t)\Big) = \prod_{i=1}^t |S_i| \geqslant \eta^t X^t.$$

Since the $h_i$ may be restricted to range over $[-2X, 2X]$, which contains at most $5X$ integers, there is some choice of $h_2, \dots, h_t$ so that $\sum_x 1_{S_1}(x)1_{S_2}(x+h_2)\cdots 1_{S_t}(x+h_t) \geqslant (\eta/5)^t X$. That is, there is a set $S$, $|S| \geqslant (\eta/5)^t X$, such that $S \subseteq S_1 \cap (S_2 - h_2) \cap \cdots \cap (S_t - h_t)$. But then $S - S \subseteq \bigcap_{i=1}^t (S_i - S_i)$, and the result is proved. □

We now turn to the heart of the proof of Proposition 6.1. The key technical ingredient is the following.

PROPOSITION 6.4. *Let $d, r$ be positive integers with $d \geqslant 2$. Let $\alpha \in (0,1]$. Let $m$ be an integer, set $N := d^m$ and suppose that $N \geqslant (2/\alpha)^{(32d)^r}$. Suppose that $t_1, \dots, t_r$ are integers with $|t_j| \leqslant N$. Define $L_d : \{0,1\}^{[0,m)} \to [0, N)$ as in (1.1). Suppose that $A \subset (\{0,1\}^{[0,m)})^r$ is a set of size at least $\alpha 2^{mr}$. Then at least $(\alpha/2)^{(32d)^r} N^r$ integers $x$ with $|x| \leqslant (8dN)^r$ may be written as a $\pm$ sum of at most $(4d)^r$ numbers $\prod_{j=1}^r (L_d(\mathbf{y}_j) + t_j)$ with $(\mathbf{y}_1, \dots, \mathbf{y}_r) \in A$.*



https://doi.org/10.1112/S0010437X24007723 Published online by Cambridge University Press

*Proof.* It is convenient to write $\phi_j(\mathbf{y}) := L_d(\mathbf{y}) + t_j$, $j = 1, \dots, r$. Note, for further use, the containment

$$\phi_j(\{0,1\}^{[0,m)}) \subset [-2N, 2N], \tag{6.2}$$

which follows from the fact that $|t_j| \leqslant N$.

Turning to the proof, we proceed by induction on $r$. In the case $r = 1$, we can apply Theorem 2.3. Noting that $L_d(\{0,1,\dots,d-1\}^m) = \{0,1,\dots,N-1\}$, we see that at least $\alpha^{\log_2 d} N$ elements of $\{0,1,\dots,N-1\}$ are the sum of $d-1$ elements $L_d(\mathbf{y}_1)$, $\mathbf{y}_1 \in A$. Since, for any $\mathbf{y}_1^{(1)}, \dots, \mathbf{y}_1^{(d-1)} \in A$, we have

$$\sum_{i=1}^{d-1} \phi_1(\mathbf{y}_1^{(i)}) = \sum_{i=1}^{d-1} L_d(\mathbf{y}_1^{(i)}) + (d-1)t_1,$$

we see that at least $\alpha^{\log_2 d} N$ elements of $[-dN, dN]$ are the sum of $d-1$ elements $\phi_1(\mathbf{y}_1)$, $\mathbf{y}_1 \in A$, which gives the required result in this case.

Now suppose that $r \geqslant 2$, and that we have proven the result for smaller values of $r$. For each $\mathbf{y}_r \in \{0,1\}^{[0,m)}$, denote by $A(\mathbf{y}_r) \subseteq (\{0,1\}^{[0,m)})^{r-1}$ the maximal set such that $A(\mathbf{y}_r) \times \{\mathbf{y}_r\} \subseteq A$. By a simple averaging argument there is a set $Y$ of at least $(\alpha/2)2^m$ values of $\mathbf{y}_r$ such that $|A(\mathbf{y}_r)| \geqslant (\alpha/2)2^{m(r-1)}$. By the inductive hypothesis, for each $\mathbf{y}_r \in Y$, there is a set

$$B(\mathbf{y}_r) \subseteq [-(8dN)^{r-1}, (8dN)^{r-1}], \tag{6.3}$$

with

$$|B(\mathbf{y}_r)| \geqslant (\alpha/4)^{(32d)^{r-1}} N^{r-1}, \tag{6.4}$$

such that everything in $B(\mathbf{y}_r)$ is a $\pm$ sum of at most $(4d)^{r-1}$ elements $\phi_1(\mathbf{y}_1) \cdots \phi_{r-1}(\mathbf{y}_{r-1})$ with $(\mathbf{y}_1, \dots, \mathbf{y}_{r-1}) \in A(\mathbf{y}_r)$. Observe that everything in $(B(\mathbf{y}_r) - B(\mathbf{y}_r))\phi_r(\mathbf{y}_r)$ is then a $\pm$ combination of at most $2(4d)^{r-1}$ elements $\phi_1(\mathbf{y}_1) \cdots \phi_r(\mathbf{y}_r)$ with $(\mathbf{y}_1, \dots, \mathbf{y}_r) \in A$.

Suppose now that $z \in (d-1)\phi_r(Y) = \phi_r(Y) + \cdots + \phi_r(Y)$. Note that, by (6.2),

$$|z| < 2dN. \tag{6.5}$$

For each such $z$, pick a representation $z = \phi_r(\mathbf{y}_r^{(1)}) + \cdots + \phi_r(\mathbf{y}_r^{(d-1)})$ with $\mathbf{y}_r^{(i)} \in Y$ for $i = 1, \dots, d-1$, and define $S(z) := \bigcap_{i=1}^{d-1} (B(\mathbf{y}_r^{(i)}) - B(\mathbf{y}_r^{(i)}))$. By (6.3), (6.4) and Lemma 6.3 (taking $X := (8dN)^{r-1}$, $\eta := (8d)^{-(r-1)}(\alpha/4)^{(32d)^{r-1}}$ and $t := d-1$ in that lemma), we have

$$\begin{aligned} |S(z)| &\geqslant 5^{-(d-1)}(8d)^{-(r-1)(d-2)}(\alpha/4)^{(32d)^{r-1}(d-1)} N^{r-1} \\ &\geqslant (\alpha/2)^{4d(32d)^{r-1}} N^{r-1}. \end{aligned} \tag{6.6}$$

Here, the second bound is crude and uses the inequality

$$(2d+2)(32d)^{r-1} \geqslant (d-1)\log_2 5 + (r-1)(d-2)\log_2(8d),$$

valid for $d \geqslant 2$ and $r \geqslant 1$ (by a large margin if $r > 1$).

Note that everything in $S(z)z$ is a $\pm$ combination of at most $2(d-1)(4d)^{r-1}$ elements $\phi_1(\mathbf{y}_1) \cdots \phi_r(\mathbf{y}_r)$ with $(\mathbf{y}_1, \dots, \mathbf{y}_r) \in A$. Set $\Omega := \bigcup_{z \in (d-1)\phi_r(Y)} (S(z) \times \{z\})$. Then $\Omega \subset [-U, U] \times [-V, V]$ where by (6.3) and (6.5) we can take $U := 2(8dN)^{r-1}$ and $V := 2dN$. Now by Theorem 2.3, and recalling that $|Y| \geqslant (\alpha/2)2^m$, we have $|(d-1)\phi_r(Y)| = |(d-1)L_d(Y)| \geqslant (\alpha/2)^{\log_2 d} N$. From this and (6.6), we have $|\Omega| \geqslant (\alpha/2)^{4d(32d)^{r-1} + \log_2 d} N^r$. Thus, noting that $UV = 2^{3r-1+r\log_2 d} N^r$, it follows that $|\Omega| \geqslant \varepsilon UV$ with

$$\varepsilon := (\alpha/2)^{4d(32d)^{r-1} + 3r + (r+1)\log_2 d}. \tag{6.7}$$



https://doi.org/10.1112/S0010437X24007723 Published online by Cambridge University Press

Now we aim to apply Lemma 6.2. For such an application to be valid, we require $\varepsilon < 2^{-44}$, which is comfortably a consequence of (6.7). We also need that $U, V \geqslant 64/\varepsilon$, which follows from (6.7) and the lower bound on $N$ in the hypotheses of the proposition. Note that if $(u_1, v_1) = (S(z), z)$, $(u_2, v_2) = (S(z'), z') \in \Omega$ then $u_1v_1 + u_2v_2 = S(z)z' + S(z')z'$ is a $\pm$ combination of at most $(4d)^r$ elements $\phi_1(\mathbf{y}_1) \cdots \phi_r(\mathbf{y}_r)$ with $(\mathbf{y}_1, \dots, \mathbf{y}_r) \in A$, and by Lemma 6.2 there are $\geqslant \varepsilon^7 UV > \varepsilon^7 N^r$ such elements. To conclude the argument, we need only check that $\varepsilon^7 \geqslant (\alpha/2)^{(32d)^r}$, which, using (6.7), comes down to checking that $4d(32d)^{r-1} \geqslant 7(3r + (r+1)\log_2 d)$, which is comfortably true for all $d, r \geqslant 2$. □

Finally, we are ready for the proof of the main result of the section, Proposition 6.1, which results from combining Propositions 5.2 and 6.4.

*Proof of Proposition 6.1.* In the following proof we suppress a number of short calculations, showing that various constants are bounded by $C = b^{7k^2/2}$. These calculations are all simple finger exercises using the assumption that $b \geqslant 3$ and $k \geqslant 2$.

First apply Proposition 5.2. As in the statement of that Proposition, we obtain $t_1, \dots, t_{k-1} \in \mathbf{Z}$, $|t_j| \leqslant N$ such that, for at least $1/2\delta^{2^k} 2^{(k-1)n/k}$ choices of $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k-1)} \in \{0,1\}^{[0,n/k)}$, we have

$$\tilde{\mathrm{w}}_n\left(\theta q_0 \prod_{i=1}^{k-1} (L_{b^k}(\mathbf{x}^{(i)}) + t_i)\right) \leqslant 2^k b^{2k} \log(2/\delta) \tag{6.8}$$

for some positive integer $q_0 \leqslant b^{k^2}$. (For the definition of $\tilde{\mathrm{w}}_n$, see Definition 5.1.) To this conclusion, we apply Proposition 6.4, taking $m := n/k$, $r := k-1$ and $d := b^k$ in that proposition, and taking $A$ to be the set of all $(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k-1)})$ as just described; thus, we may take $\alpha := \delta^{2^k}/2$. Note that $N = d^m = b^n$ is the same quantity. The reader may check that the lower bound on $N$ required for this application of Proposition 6.4 is a consequence of the assumption on $N$ in Proposition 6.1.

We conclude that at least $(\delta^{2^k}/4)^{(32b^k)^{k-1}} N^{k-1} > (\delta/2)^C N^{k-1}$ integers $x$ with $|x| \leqslant (8b^k N)^{k-1}$ may be written as a $\pm$ sum of at most $(4b^k)^{k-1}$ numbers of the form $\prod_{i=1}^{k-1}(L_{b^k}(\mathbf{x}^{(i)}) + t_i)$, with $(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k-1)}) \in A$. By (6.8), the fact that $\tilde{\mathrm{w}}_n(-\alpha) = \tilde{\mathrm{w}}_n(\alpha)$, as well as the (easily verified) subadditivity property

$$\tilde{\mathrm{w}}_n(\alpha_1 + \cdots + \alpha_s) \leqslant s(\tilde{\mathrm{w}}_n(\alpha_1) + \cdots + \tilde{\mathrm{w}}_n(\alpha_s)),$$

we see that, for all such $x$, we have

$$\tilde{\mathrm{w}}_n(\theta q_0 x) \leqslant (4b^k)^{2(k-1)} 2^k b^{2k} \log(2/\delta) < C \log(2/\delta).$$

Finally, note that for all these $x$, we have $|q_0 x| \leqslant b^{k^2}(8b^k)^{k-1} N^{k-1}$, which is less than $CN^{k-1}$. This concludes the proof. □

## 7. From digital to diophantine

In this section we turn to the final step in the outline of § 2, the aim of which is to convert the 'digital' conclusion of Proposition 6.1 to the 'diophantine' conclusion of Proposition 2.1. Before turning to detailed statements, we comment on the notion of a centred base $b$ expansion.

*Centred base $b$ expansions.* Consider $\alpha \in \mathbf{R}/\mathbf{Z}$. Then there are essentially unique choices of integers $\alpha_j \in (-b/2, b/2]$ such that

$$\alpha = \alpha_0 + \alpha_1 b^{-1} + \alpha_2 b^{-2} + \cdots \pmod 1. \tag{7.1}$$

We call this the *centred* base $b$ expansion of $\alpha \pmod 1$.



https://doi.org/10.1112/S0010437X24007723 Published online by Cambridge University Press

Let us pause to explain the existence of such expansions. When $b$ is odd, so that $(-b/2, b/2] = \{-1/2(b-1), \dots, 1/2(b-1)\}$, the centred expansion may be obtained from the more usual base $b$ expansion of $\alpha + b/2$, noting that $b/2 = 1/2(b-1)(1 + b^{-1} + b^{-2} + \cdots)$. As usual, there is some ambiguity when all the digits from some point on are $1/2(b-1)$; any such number can also be written with all digits from some point on being $-1/2(b-1)$. For consistency with the usual base $b$ expansions, we always prefer the latter representation. When $b$ is even, so that $(-b/2, b/2] = \{-1/2(b-2), \dots, 1/2b\}$, one instead considers the usual base $b$ expansion of $\alpha + b(b-2)/2(b-1)$, noting now that $b(b-2)/2(b-1) = 1/2(b-2)(1 + b^{-1} + b^{-2} + \cdots)$.

Definition 7.1. *Given $\alpha \in \mathbf{R}/\mathbf{Z}$, denote by $\mathrm{w}_n(\alpha)$ the number of nonzero digits among the first $n$ digits $\alpha_0, \alpha_1, \dots, \alpha_{n-1}$ in the centred expansion (7.1).*

We record the connection between $\mathrm{w}_n$ and the 'analytic' proxy $\tilde{\mathrm{w}}_n$, introduced in Definition 5.1.

Lemma 7.2. *Suppose that $b \geqslant 3$. Then $\tilde{\mathrm{w}}_n(\alpha) \leqslant \mathrm{w}_n(\alpha) \leqslant 16b^2 \tilde{\mathrm{w}}_n(\alpha)$.*

*Proof.* Let the centred expansion of $\alpha (\mathrm{mod}\ 1)$ be (7.1), and suppose that $\alpha_i$ is a nonzero digit. We have $\alpha b^{i-1} \equiv \sum_{j \geqslant 0} \alpha_{i+j} b^{-j-1} (\mathrm{mod}\ 1)$. However,

$$\left| \sum_{j \geqslant 0} \alpha_{i+j} b^{-j-1} \right| \leqslant \tfrac{b}{2} \sum_{j \geqslant 0} b^{-j-1} = \tfrac{b}{2(b-1)} \leqslant \tfrac{3}{4},$$

and, since $\alpha_i \neq 0$,

$$\left| \sum_{j \geqslant 0} \alpha_{i+j} b^{-j-1} \right| \geqslant \tfrac{1}{b} - \tfrac{b}{2} \sum_{j \geqslant 1} b^{-j-1} = \tfrac{b-2}{2b(b-1)} \geqslant \tfrac{1}{4b}.$$

Thus, $\|\alpha b^{i-1}\| \geqslant 1/4b$ and the upper bound follows.

The lower bound is not needed elsewhere in the paper, but we sketch the proof for completeness. Let $I := \{i : \alpha_i \neq 0\}$. Given $j$, denote by $i(j)$ the distance from $j$ to the smallest element of $I$ that is greater than $j$. Then

$$\|\alpha b^j\| = \left\| \sum_{i \in I, i > j} \alpha_i b^{-i+j} \right\| \leqslant \tfrac{b}{2} \sum_{m \geqslant i(j)} b^{-m} = \tfrac{b^2}{2(b-1)} b^{-i(j)}.$$

Now square this and sum over $j$, and use the fact that $\#\{j : i(j) = i\} \leqslant |I| = \mathrm{w}_n(\alpha)$ for all $i$. □

*Remarks.* This upper bound breaks down when $b = 2$, as may be seen by considering $\alpha$ of the form $1 - 2^{-m}$. This is the main reason for the restriction to $b \geqslant 3$ in the paper.

Here is the main result of the section.

Proposition 7.3. *Let $b \geqslant 3$ be an integer. Let $r, M, n$ be positive integers, and set $N := b^n$. Let $\eta \in (0, 1]$ be real. Suppose that $M, N \geqslant b^{20r} \eta^{-2}$. Suppose that $\theta \in \mathbf{R}$, and that $\mathrm{w}_n(\theta m) \leqslant r$ for at least $\eta M$ values of $m \in [-M, M]$. Then there is some positive integer $q \leqslant b^{20r} \eta^{-2}$ such that $\|\theta q\| \leqslant b^{20r} \eta^{-2} M^{-1} N^{-1}$.*

Before giving the proof, we assemble some lemmas. In the first of these, we will again be concerned with centred expansions in base $b$, but this time of integers. Every integer $x$ has a unique finite-length centred base $b$ expansion

$$x = x_0 + x_1 b + x_2 b^2 + \cdots, \tag{7.2}$$



https://doi.org/10.1112/S0010437X24007723 Published online by Cambridge University Press

with $x_i \in (-b/2, b/2]$. To see uniqueness, note that $x_0$ is uniquely determined by $x (\mathrm{mod}\ b)$, then $x_1$ is uniquely determined by $x - x_0/b (\mathrm{mod}\ b)$, and so on. Strictly speaking, we do not need the existence in this paper but one way to see it is to take the usual base $b$ expansion and modify from the right. For instance, in base 10 we have, denoting the 'digit' $-d$ by $\overline{d}$, $6277 = 628\overline{3} = 63\overline{23} = 1\overline{4}3\overline{23}$.

Denote by $\mathrm{d}_b(x)$ the number of nonzero digits in this expansion of $x$. The set of $x$ for which $\mathrm{d}_b(x) \leqslant r$ is a kind of 'digital Hamming ball'. As for true Hamming balls [Bon70, KS20], subsets of this set have little additive structure. Such a result was stated as Proposition 2.4. We recall the statement now. Recall that, if $A \subset \mathbf{Z}$ is a finite set, the additive energy $E(A)$ is the number of quadruples $(a_1, a_2, a_3, a_4) \in A \times A \times A \times A$ with $a_1 + a_2 = a_3 + a_4$.

PROPOSITION 2.4 *Let $r \in \mathbf{Z}_{\geqslant 0}$. Suppose that $A \subset \mathbf{Z}$ is a finite set, all of whose elements have at most $r$ nonzero digits in their centred base $b$ expansion. Then $E(A) \leqslant (2b)^{4r}|A|^2$.*

The proof of Proposition 2.4 will proceed by induction. However, to make this work, we need to prove a more general statement, involving four potentially different sets $A_1, A_2, A_3, A_4$ instead of just one, as well as the provision for a 'carry' in base $b$ arithmetic. Here is the more general statement, from which Proposition 2.4 follows immediately.

LEMMA 7.4. *Let $r_1, r_2, r_3, r_4 \in \mathbf{Z}_{\geqslant 0}$. For each $i \in \{1, 2, 3, 4\}$, suppose that $A_i \subset \mathbf{Z}$ is a finite set, all of whose elements have at most $r_i$ nonzero digits in their centred base $b$ expansion. Let $e \in \mathbf{Z}$, $|e| < b$. Then the number of quadruples $(a_1, a_2, a_3, a_4) \in A_1 \times A_2 \times A_3 \times A_4$ with $a_1 + a_2 = a_3 + a_4 + e$ is at most $(2b)^{r_1+r_2+r_3+r_4}|A_1|^{1/2}|A_2|^{1/2}|A_3|^{1/2}|A_4|^{1/2}$.*

*Proof.* We proceed by induction on $\sum_{j=1}^4 |A_j| + \sum_{j=1}^4 r_j$, the result being obvious when this quantity is zero. Suppose now that $\sum_{j=1}^4 |A_j| + \sum_{j=1}^4 r_j = n > 0$ and that the result has been proven for all smaller values of $n$. If any of the $A_j$ are empty, or if $A_1 = A_2 = A_3 = A_4 = \{0\}$, the result is obvious.

Suppose this is not the case, but that $b$ divides every element of $\bigcup_{j=1}^4 A_j$. Let $b^m$ be the largest power of $b$ that divides every element of $\bigcup_{j=1}^4 A_j$, this being well defined since this set contains at least one nonzero element. Then, if the number of quadruples in $A_1 \times A_2 \times A_3 \times A_4$ with $a_1 + a_2 = a_3 + a_4 + e$ is nonzero, we must have $e = 0$, and the number of such quadruples is the same as the number in $1/b^m A_1 \times 1/b^m A_2 \times 1/b^m A_3 \times 1/b^m A_4$. Thus, replacing $A_j$ by $\frac{1}{b^m} A_j$, we may assume that not all the elements of $\bigcup_{j=1}^4 A_j$ are divisible by $b$.

For each $j \in \{1, 2, 3, 4\}$ and for each $i \in (-b/2, b/2]$, write $A_j^{(i)}$ for the set of $x \in A_j$ whose first digit $x_0$ (in the centred base $b$ expansion (7.2)) is $i$. Write $\alpha_j(i)$ for the relative density of $A_j^{(i)}$ in $A_j$, that is to say, $|A_j^{(i)}| = \alpha_j(i)|A_j|$. Any quadruple $(a_1, a_2, a_3, a_4)$ with $a_1 + a_2 = a_3 + a_4 + e$ must have $a_j \in A_j^{(i_j)}$, where $i_1 + i_2 \equiv i_3 + i_4 + e (\mathrm{mod}\ b)$. Let us estimate the number of such quadruples $(a_1, a_2, a_3, a_4)$ for each quadruple $(i_1, i_2, i_3, i_4) \in (-b/2, b/2]^4$ satisfying this condition.

First note that $i_1 + i_2 = i_3 + i_4 + e + e'b$ for some integer $e'$, where

$$|e'| \leqslant \tfrac{1}{b}(|i_1 + i_2 - i_3 - i_4| + |e|) \leqslant \tfrac{3(b-1)}{b} < b,$$

where here we noted that $|i_1 - i_3|, |i_2 - i_4|, |e| \leqslant b - 1$. We then have $1/b(a_1 - i_1) + 1/b(a_2 - i_2) - 1/b(a_3 - i_3) - 1/b(a_4 - i_4) = -e'$. Now the set $A_j' := \frac{1}{b}(A_j^{(i_j)} - i_j)$ is a finite set of integers, all of whose elements $x$ have $\mathrm{d}_b(x) \leqslant r_j' := r_j - 1_{i_j \neq 0}$. Note that $\sum_{j=1}^4 |A_j'| + \sum_{j=1}^4 r_j' < \sum_{j=1}^4 |A_j| + \sum_{j=1}^4 r_j$; if any $i_j$ is not zero, this follows from the fact that $r_j' = r_j - 1$, whereas if $i_1 = i_2 =$



https://doi.org/10.1112/S0010437X24007723 Published online by Cambridge University Press

$i_3 = i_4 = 0$ we have $\sum_{j=1}^4 |A'_j| = \sum_{j=1}^4 |A_j^{(0)}| < \sum_{j=1}^4 |A_j|$, since not every element of $\bigcup_{j=1}^4 A_j$ is a multiple of $b$.

It follows from the inductive hypothesis that the numbers of quadruples $(a_1, a_2, a_3, a_4)$ with $a_1 + a_2 = a_3 + a_4 + e$, and with $a_j \in A_j^{(i_j)}$, $j = 1, \dots, 4$, is bounded above by

$$(2b)^{r_1+r_2+r_3+r_4-\#\{j:i_j\neq 0\}} \prod_{j=1}^4 |A_j^{(i_j)}|^{1/2}.$$

To complete the inductive step, it is therefore enough to show that

$$\sum_{i_1+i_2\equiv i_3+i_4+e (\mathrm{mod}\ b)} (2b)^{-\#\{j:i_j\neq 0\}} \prod_{j=1}^4 \alpha_j(i_j)^{1/2} \leqslant 1. \tag{7.3}$$

If $e \not\equiv 0 (\mathrm{mod}\ b)$ then we have $\#\{j : i_j \neq 0\} \geqslant 1$ for all $(i_1, i_2, i_3, i_4)$ in this sum, and moreover (where all congruences are (mod $b$))

$$\begin{aligned}
&\sum_{i_1+i_2\equiv i_3+i_4+e} \prod_{j=1}^4 \alpha_j(i_j)^{1/2} \\
&= \sum_{x\in \mathbf{Z}/b\mathbf{Z}} \left( \sum_{i_1+i_2\equiv x+e} \alpha_1(i_1)^{1/2}\alpha_2(i_2)^{1/2} \right) \left( \sum_{i_3+i_4\equiv x} \alpha_3(i_3)^{1/2}\alpha_4(i_4)^{1/2} \right) \\
&\leqslant \sum_{x\in \mathbf{Z}/b\mathbf{Z}} \left( \sum_{i_1+i_2\equiv x+e} \frac{\alpha_1(i_1)+\alpha_2(i_2)}{2} \right) \left( \sum_{i_3+i_4\equiv x} \frac{\alpha_3(i_3)+\alpha_4(i_4)}{2} \right) = b,
\end{aligned}$$

since $\sum_i \alpha_j(i) = 1$ for each $j$. Therefore, (7.3) holds in this case.

Suppose, then, that $e \equiv 0 (\mathrm{mod}\ b)$, which means that $e = 0$. Then, if $i_1 + i_2 \equiv i_3 + i_4 (\mathrm{mod}\ b)$ we either have $(i_1, i_2, i_3, i_4) = (0, 0, 0, 0)$, or else $\#\{j : i_j \neq 0\} \geqslant 2$, and so to establish (7.3) it suffices to show that

$$\prod_{j=1}^4 \alpha_j(0)^{1/2} + (2b)^{-2} \sum_{\substack{i_1+i_2\equiv i_3+i_4 (\mathrm{mod}\ b)\\ (i_1,i_2,i_3,i_4)\neq(0,0,0,0)}} \prod_{j=1}^4 \alpha_j(i_j)^{1/2} \leqslant 1. \tag{7.4}$$

Write $\varepsilon_j := 1 - \alpha_j(0)$. We first estimate the contribution to the sum where none of $i_1, i_2, i_3, i_4$ are zero. We have, similarly to the above (and again with congruences being (mod $b$)),

$$\begin{aligned}
&\sum_{\substack{i_1+i_2\equiv i_3+i_4\\ i_1i_2i_3i_4\neq 0}} \prod_{j=1}^4 \alpha_j(i_j)^{1/2} \\
&= \sum_{x\in \mathbf{Z}/b\mathbf{Z}} \left( \sum_{\substack{i_1+i_2\equiv x\\ i_1i_2\neq 0}} \alpha_1(i_1)^{1/2}\alpha_2(i_2)^{1/2} \right) \left( \sum_{\substack{i_3+i_4\equiv x\\ i_3i_4\neq 0}} \alpha_3(i_3)^{1/2}\alpha_4(i_4)^{1/2} \right) \\
&\leqslant \sum_{x\in \mathbf{Z}/b\mathbf{Z}} \sum_{\substack{i_1+i_2\equiv x\\ i_1i_2\neq 0}} \left( \frac{\alpha_1(i_1)+\alpha_2(i_2)}{2} \right) \left( \sum_{\substack{i_3+i_4\equiv x\\ i_3i_4\neq 0}} \frac{\alpha_3(i_3)+\alpha_4(i_4)}{2} \right) \\
&\leqslant b \left( \frac{\varepsilon_1+\varepsilon_2}{2} \right) \left( \frac{\varepsilon_3+\varepsilon_4}{2} \right) < b \sum_{j=1}^4 \varepsilon_j.
\end{aligned}$$



https://doi.org/10.1112/S0010437X24007723 Published online by Cambridge University Press

Next we estimate the contribution to the sum in (7.4) from the terms where at least one, but not all, of $i_1, i_2, i_3, i_4$ are zero. In each such term, at least two $i_j, i_{j'}$ are not zero, say with $j < j'$. Fix a choice of $j, j'$. Then, for each $i_j, i_{j'}$, there are at most two choices of the other $i_t$, $t \in \{1,2,3,4\} \setminus \{j, j'\}$, one of which must be zero and the other then being determined by the relation $i_1 + i_2 \equiv i_3 + i_4 (\mathrm{mod}\ b)$. It follows that the contribution to the sum in (7.4) from this choice of $j, j'$ is

$$\leqslant 2 \sum_{i_j, i_{j'} \neq 0} \alpha_j(i_j)^{1/2} \alpha_{j'}(i_{j'})^{1/2} = 2\Bigg( \sum_{i \neq 0} \alpha_j(i)^{1/2} \Bigg) \Bigg( \sum_{i \neq 0} \alpha_{j'}(i)^{1/2} \Bigg) \leqslant 2b\varepsilon_j^{1/2} \varepsilon_{j'}^{1/2} \leqslant b(\varepsilon_j + \varepsilon_{j'}),$$

where in the middle step we used the Cauchy–Schwarz inequality and the fact that $\sum_{i \neq 0} \alpha_j(i) = \varepsilon_j$. Summing over the six choices of $j, j'$ gives an upper bound of $3b \sum_{j=1}^4 \varepsilon_j$. Putting all this together, we see that the left-hand side of (7.4) is bounded above by $\prod_{j=1}^4 (1 - \varepsilon_j)^{1/2} + 1/b \sum_{j=1}^4 \varepsilon_j$. Using $\prod_{j=1}^4 (1 - \varepsilon_j)^{1/2} \leqslant 1 - 1/2 \sum_{j=1}^4 \varepsilon_j$, it follows that this is at most 1. This completes the proof of (7.4), and, hence, of Lemma 7.5. □

Now we turn to the proof of Proposition 7.3.

*Proof of Proposition 7.3.* Consider the map $\psi : \mathbf{R} \to \mathbf{Z}$ defined as follows. If $\alpha (\mathrm{mod}\ 1)$ has centred base $b$ expansion as in (7.1), set $\psi(\alpha) := \alpha_0 b^{n-1} + \cdots + \alpha_{n-2} b + \alpha_{n-1}$. Observe that

$$\mathrm{d}_b(\psi(\alpha)) = \mathrm{w}_n(\alpha). \tag{7.5}$$

Note that

$$\|\alpha - b^{1-n} \psi(\alpha)\| \leqslant \sum_{i \geqslant n} \tfrac{b}{2} b^{-i} \leqslant \tfrac{3}{4} b^{1-n}. \tag{7.6}$$

Thus, if $\alpha_1 + \alpha_2 = \alpha_3 + \alpha_4$ then

$$\|b^{1-n}(\psi(\alpha_1) + \psi(\alpha_2) - \psi(\alpha_3) - \psi(\alpha_4))\| \leqslant 3b^{1-n}.$$

Note also that, since $\psi$ takes values in $\mathbf{Z} \cap [-\frac{3}{4}b^n, \frac{3}{4}b^n]$, we have

$$|\psi(\alpha_1) + \psi(\alpha_2) - \psi(\alpha_3) - \psi(\alpha_4)| \leqslant 3b^n.$$

Now if $x \in \mathbf{Z}$ is an integer with $\|b^{1-n} x\| \leqslant 3b^{1-n}$ and $|x| \leqslant 3b^n$ then $x$ takes (at most) one of the $7(6b+1)$ values $\lambda b^{n-1} + \lambda'$, $\lambda \in \{-3b, \ldots, 3b\}$, $\lambda' \in \{0, \pm 1, \pm 2, \pm 3\}$. Denoting by $\Sigma$ the set consisting of these $7(6b+1)$ values, we see that $\psi$ has the following almost-homomorphism property: if $\alpha_1 + \alpha_2 = \alpha_3 + \alpha_4$ then

$$\psi(\alpha_1) + \psi(\alpha_2) - \psi(\alpha_3) - \psi(\alpha_4) \in \Sigma.$$

With parameters as in the statement of Proposition 7.3, consider the map $\pi : [-M, M] \to \mathbf{Z}$ given by

$$\pi(m) := \psi(\theta m). \tag{7.7}$$

Since the map $m \mapsto \theta m$ is a homomorphism from $\mathbf{Z}$ to $\mathbf{R}$, we see that $\pi$ also has an almost-homomorphism property, namely that if $m_1 + m_2 = m_3 + m_4$ then

$$\pi(m_1) + \pi(m_2) - \pi(m_3) - \pi(m_4) \in \Sigma. \tag{7.8}$$

Denote by $\mathscr{M}$ the set of all $m \in [-M, M]$ such that $\mathrm{w}_n(\theta m) \leqslant r$. Thus, by the assumptions of Proposition 7.3, $|\mathscr{M}| \geqslant \eta M$. Denote $A := \pi(\mathscr{M})$. By the definition (7.7) of $\pi$, (7.5) and the definition of $\mathscr{M}$, we see that $\mathrm{d}_b(a) \leqslant r$ for all $a \in A$. For $a \in A$, denote by $X_a := \pi^{-1}(a) \cap \mathscr{M}$ the $\pi$-fibre above $a$. Decompose $A$ according to the dyadic size of these fibres, thus, for $j \in \mathbf{Z}_{\geqslant 0}$ set,



https://doi.org/10.1112/S0010437X24007723 Published online by Cambridge University Press

$$A_j := \{a \in A : 2^{-j-1}M < |X_a| \leqslant 2^{-j}M\}. \tag{7.9}$$

Denote by $\mathscr{M}_j \subset \mathscr{M}$ the points of $\mathscr{M}$ lying above $A_j$, that is to say, $\mathscr{M}_j := \bigcup_{a\in A_j} X_a$. Define $\eta_j$ by $|\mathscr{M}_j| = \eta_j M$. Since $\mathscr{M}$ is the disjoint union of the $\mathscr{M}_j$, we have

$$\sum_j \eta_j \geqslant \eta. \tag{7.10}$$

By (7.9) we have $2^{-j-1}M|A_j| \leqslant |\mathscr{M}_j| \leqslant 2^{-j}M|A_j|$, and so

$$2^j\eta_j \leqslant |A_j| \leqslant 2^{j+1}\eta_j. \tag{7.11}$$

Now by a simple application of the Cauchy–Schwarz inequality any subset of $[-M, M]$ of size at least $\varepsilon M$ has at least $\varepsilon^4 M^3/4$ additive quadruples. In particular, for any $j \in \mathbf{Z}_{\geqslant 0}$, there are $\geqslant \eta_j^4 M^3/4$ additive quadruples in $\mathscr{M}_j$. By (7.8), there is some $\sigma_j \in \Sigma$ such that, for $\geqslant 2^{-10}b^{-1}\eta_j^4 M^3$ additive quadruples in $\mathscr{M}_j$, we have

$$\pi(m_1) + \pi(m_2) = \pi(m_3) + \pi(m_4) + \sigma_j. \tag{7.12}$$

For each $j$, fix such a choice of $\sigma_j$. Now the number of such quadruples with $\pi(m_i) = a_i$ for $i = 1, 2, 3, 4$ is, for a fixed choice of $a_1, \dots, a_4$, satisfying

$$a_1 + a_2 = a_3 + a_4 + \sigma_j, \tag{7.13}$$

the number of additive quadruples in $X_{a_1} \times X_{a_2} \times X_{a_3} \times X_{a_4}$, which is bounded above by $|X_{a_1}||X_{a_2}||X_{a_3}| \leqslant 2^{-3j}M^3$ since three elements of an additive quadruple determine the fourth. It follows that the number of $(a_1, a_2, a_3, a_4) \in A_j^4$ satisfying (7.13) is $\geqslant 2^{-10}b^{-1}2^{3j}\eta_j^4$. By (7.11), this is $\geqslant 2^{-13}b^{-1}\eta_j|A_j|^3$.

Now if $S_1, S_2, S_3, S_4$ are additive sets then $E(S_1, S_2, S_3, S_4)$, the number of solutions to $s_1 + s_2 = s_3 + s_4$ with $s_i \in S_i$, is bounded by $\prod_{i=1}^4 E(S_i)^{1/4}$, where $E(S_i)$ is the number of additive quadruples in $S_i$. This is essentially the Gowers–Cauchy–Schwarz inequality for the $U^2$-norm; it may be proven by two applications of the Cauchy–Schwarz inequality or alternatively from Hölder's inequality on the Fourier side. Applying this with $S_1 = S_2 = S_3 = A_j$ and $S_4 = A_j + \sigma_j$, and noting that $E(A_j + \sigma_j) = E(A_j)$, we see that $E(A_j) \geqslant 2^{-13}b^{-1}\eta_j|A_j|^3$.

By Proposition 2.4, we have $|A_j| \leqslant 2^{4r+13}b^{4r+1}\eta_j^{-1}$. Comparing with (7.11) gives $\eta_j \leqslant 2^{2r+7-j/2}b^{2r+1/2}$. Take $J$ to be the least integer such that $2^{J/2} \geqslant 2^{2r+9}b^{2r+1/2}\eta^{-1}$; then $\sum_{j\geqslant J}\eta_j < \eta$, and so by (7.10), some $\mathscr{M}_j$, $j \leqslant J-1$, is nonempty. In particular, by (7.9) there is some value of $a$ such that $|X_a| \geqslant 2^{-J}M \geqslant 2^{-4r-20}b^{-4r-1}\eta^2 M$. Fix this value of $a$ and set $\mathscr{M}' := X_a$. Thus, to summarise,

$$|\mathscr{M}'| \geqslant 2^{-4r-20}b^{-4r-1}\eta^2 M \tag{7.14}$$

and if $m \in \mathscr{M}'$ then $\pi(m) = a$. Note that the condition on $M$ in the statement of Proposition 7.3 implies (comfortably) that $|\mathscr{M}'| \geqslant 2$.

Note that, by (7.6) and the definition (7.7) of $\pi$, we have that if $m \in \mathscr{M}'$ then

$$\|\theta m - b^{1-n}a\| \leqslant \tfrac{3}{4}b^{1-n}. \tag{7.15}$$

Pick some $m_0 \in \mathscr{M}'$, and set $\mathscr{M}'' := \mathscr{M}' - m_0 \subset [-2M, 2M]$. By the triangle inequality and (7.15), we have

$$\|\theta m\| \leqslant \tfrac{3}{2}b^{1-n} < 2bN^{-1} \tag{7.16}$$

for all $m \in \mathscr{M}''$. (Recall that, by definition, $N = b^n$.) Replacing $\mathscr{M}''$ by $-\mathscr{M}''$ if necessary (and since $|\mathscr{M}''| \geqslant 2$), it follows that there are at least $2^{-4r-22}b^{-4r-1}\eta^2 M$ integers $m \in \{1, \dots, 2M\}$ satisfying (7.16).



https://doi.org/10.1112/S0010437X24007723 Published online by Cambridge University Press

Now we apply Lemma C.1, taking $L = 2M$, $\delta_1 = 2bN^{-1}$ and $\delta_2 = 2^{-4r-22}b^{-4r-1}\eta^2$ in that result. The conditions of the lemma hold under the assumptions that $M, N \geqslant b^{20r}\eta^{-2}$ (using here the fact that $b \geqslant 3$). The conclusion implies that there is some positive integer $q \leqslant b^{20r}\eta^{-2}$ such that $\|\theta q\| \leqslant b^{20r}\eta^{-2}N^{-1}M^{-1}$, which is what we wanted to prove. □

Finally, we are in a position to prove Proposition 2.1, whose statement we recall now.

PROPOSITION 2.1 *Suppose that $k \geqslant 2$ and $b \geqslant 3$. Set $B := b^{6k^2}$. Suppose that $\delta \in (0,1)$ and that $k \mid n$. Suppose that $|\widehat{\mu_n}(\theta)| \geqslant \delta$ and that $N \geqslant (2/\delta)^B$, where $N := b^n$. Then there is a positive integer $q \leqslant (2/\delta)^B$ such that $\|\theta q\| \leqslant (2/\delta)^B N^{-k}$.*

*Proof.* First apply Proposition 6.1. The conclusion is that, for at least $(\delta/2)^C N^{k-1}$ values of $m$, $|m| \leqslant CN^{k-1}$, we have $\tilde{\mathrm{w}}_n(\theta m) \leqslant C\log(2/\delta)$, where $C := b^{7k^2/2}$. By Lemma 7.2, for these values of $m$, we have $\mathrm{w}_n(\theta m) \leqslant 16b^2C\log(2/\delta)$. (For the definitions of $\tilde{\mathrm{w}}_n$ and $\mathrm{w}_n$, see Definitions 5.1 and 7.1 respectively.) Now apply Proposition 7.3 with $\eta := (\delta/2)^C C^{-1}$, $r = \lceil 16b^2C\log(2/\delta)\rceil$, $N = b^n$ (as usual) and $M := CN^{k-1}$.

To process the resulting conclusion, note that $b^{20r}\eta^{-2} \leqslant (2/\delta)^{C'}$, with

$$C' := 2C + 320b^2C\log b + \log_2(C^2b^{20}) < 321b^2C\log b < b^8C < B.$$

Proposition 2.1 then follows. □

## Appendix A. Box norm inequalities

In this appendix we prove an inequality, Proposition A.2, which is in a sense well known: indeed, it underpins the theory of hypergraph regularity [Gow07] and is also very closely related to generalised von Neumann theorems and the notion of the Cauchy–Schwarz complexity in additive combinatorics. We begin by recalling the basic definition of Gowers box norms as given in [GT10, Appendix B].

DEFINITION A.1. Let $(X_i)_{i\in I}$ be a finite collection of finite nonempty sets, and denote by $X_I := \prod_{i\in I} X_i$ the Cartesian product of these sets. Let $f : X_I \to \mathbf{C}$ be a function. Then we define the (Gowers) box norm $\|f\|_{\square(X_I)}$ to be the unique non-negative real number such that

$$\|f\|_{\square(X_I)}^{2^{|I|}} = \mathbf{E}_{x_I^{(0)}, x_I^{(1)} \in X_I} \prod_{\omega_I \in \{0,1\}^I} \mathcal{C}^{|\omega_I|} f(x_I^{(\omega_I)}).$$

Here, $\mathcal{C}$ denotes the complex conjugation operator and, for any $x_I^{(0)} = (x_i^{(0)})_{i\in I}$ and $x_I^{(1)} = (x_i^{(1)})_{i\in I}$ in $X_I$ and $\omega_I = (\omega_i)_{i\in I} \in \{0,1\}^I$, we write $x_I^{(\omega_I)} = (x_i^{(\omega_i)})_{i\in I}$ and $|\omega_I| := \sum_{i\in I} |\omega_i|$.

It is not obvious that $\|f\|_{\square(X_I)}$ is well defined, but this is so; see [GT10, Appendix B] for a proof. Another non-obvious fact, whose proof may also be found in [GT10, Appendix B], is that $\|f\|_{\square(X_I)}$ is a norm for $|I| \geqslant 2$. When $|I| = 1$, say $I = \{1\}$, we have $\|f\|_{\square(X_I)} = |\sum_{x_1\in X_1} f(x_1)|$, which is only a seminorm.

To clarify notation, in the case $I = \{1,2\}$ we have

$$\|f\|_{\square(X_{\{1,2\}})}^4 = \mathbf{E}_{\substack{x_1^{(0)}, x_1^{(1)} \in X_1 \\ x_2^{(0)}, x_2^{(1)} \in X_2}} f(x_1^{(0)}, x_2^{(0)})\overline{f(x_1^{(0)}, x_2^{(1)})f(x_1^{(1)}, x_2^{(0)})}f(x_1^{(1)}, x_2^{(1)}).$$

Here is the inequality we will need. The proof is simply several applications of the Cauchy–Schwarz inequality, the main difficulty being one of notation.



https://doi.org/10.1112/S0010437X24007723 Published online by Cambridge University Press

Proposition A.2. *Suppose that the notation is as in Definition A.1. Suppose additionally that, for each $i \in I$, we have a 1-bounded function $\Psi_i : X_I \to \mathbf{C}$ that does not depend on the value of $x_i$, that is to say, $\Psi_i(x_I) = \Psi_i(x'_I)$ if $x_j = x'_j$ for all $j \neq i$. Let $f : X_I \to \mathbf{C}$ be a function. Then we have*

$$|\mathbf{E}_{x_I \in X_I}\Big(\prod_{i\in I}\Psi_i(x_I)\Big)f(x_I)| \leqslant \|f\|_{\square(X_I)}.$$

*Proof.* We proceed by induction on $|I|$, the result being a tautology when $|I| = 1$. Suppose now that $|I| \geqslant 2$, and that we have already established the result for smaller values of $|I|$. Let $\alpha$ be some element of $I$, and write $I' := I \setminus \{\alpha\}$. By the Cauchy–Schwarz inequality, the 1-boundedness of $\Psi_\alpha$, and the fact that $\Psi_\alpha$ does not depend on $x_\alpha$, we have

$$\begin{aligned}
&\Big|\mathbf{E}_{x_I \in X_I}\Big(\prod_{i\in I}\Psi_i(x_I)\Big)f(x_I)\Big|^2\\
&= \Big|\mathbf{E}_{x_{I'}\in X_{I'}}\Psi_\alpha(x_I)\mathbf{E}_{x_\alpha \in X_\alpha}\Big(\prod_{i\in I'}\Psi_i(x_I)\Big)f(x_I)\Big|^2\\
&\leqslant \mathbf{E}_{x_{I'}\in X_{I'}}\Big|\mathbf{E}_{x_\alpha \in X_\alpha}\Big(\prod_{i\in I'}\Psi_i(x_I)\Big)f(x_I)\Big|^2\\
&= \mathbf{E}_{x_\alpha^{(0)},x_\alpha^{(1)}\in X_\alpha}\mathbf{E}_{x_{I'}\in X_{I'}}\Big(\prod_{i\in I'}\Psi_i(x_{I'},x_\alpha^{(0)})\overline{\Psi_i(x_{I'},x_\alpha^{(1)})}\Big)f(x_{I'},x_\alpha^{(0)})\overline{f(x_{I'},x_\alpha^{(0)})}.
\end{aligned}$$

For fixed $x_\alpha^{(0)}, x_\alpha^{(1)}$, we may apply the induction hypothesis (with indexing set $I'$) with 1-bounded functions, i.e.

$$\tilde{\Psi}_i(x_{I'}) := \Psi_i(x_{I'},x_\alpha^{(0)})\overline{\Psi_i(x_{I'},x_\alpha^{(1)})}$$

and with

$$\tilde{f}(x_{I'}) = f(x_{I'},x_\alpha^{(0)})\overline{f(x_{I'},x_\alpha^{(0)})},$$

noting that $\tilde{\Psi}_i$ does not depend on $x_i$.

This gives

$$\Big|\mathbf{E}_{x_I\in X_I}\Big(\prod_{i\in I}\Psi_i(x_I)\Big)f(x_I)\Big|^2 \leqslant \mathbf{E}_{x_\alpha^{(0)},x_\alpha^{(1)}\in X_\alpha}\Big\|f(\cdot,x_\alpha^{(0)})\overline{f(\cdot,x_\alpha^{(1)})}\Big\|_{\square(X_{I'})}.$$

By Hölder's inequality, it follows that

$$\Big|\mathbf{E}_{x_I\in X_I}\Big(\prod_{i\in I}\Psi_i(x_I)\Big)f(x_I)\Big|^{2^{|I|}} \leqslant \mathbf{E}_{x_\alpha^{(0)},x_\alpha^{(1)}\in X_\alpha}\Big\|f(\cdot,x_\alpha^{(0)})\overline{f(\cdot,x_\alpha^{(1)})}\Big\|_{\square(X_{I'})}^{2^{|I|-1}}.$$

However, the right-hand side is precisely $\|f\|_{\square(X_I)}^{2^{|I|}}$, and the inductive step is complete. □

## Appendix B. Sumsets of subsets of $\{0,1\}^n$

In this appendix we provide some comments on Theorem 2.3, which seems to have a very complicated history. In the case $r = 2$ it is due to Woodall [Woo77], and independently to Hajela and Seymour [HS85].

In the general case, Theorem 2.3 is a consequence of the following real-variable inequality, which was conjectured in [HS85].



https://doi.org/10.1112/S0010437X24007723 Published online by Cambridge University Press

Proposition B.1. *Let $r \geqslant 2$ be an integer. Suppose that $1 \geqslant x_1 \geqslant x_2 \geqslant \cdots \geqslant x_r \geqslant 0$. Then*

$$(x_1 \cdots x_r)^\gamma + (x_1 \cdots x_{r-1}(1-x_r))^\gamma + \cdots + ((1-x_1)\cdots(1-x_r))^\gamma \geqslant 1,$$

*where $\gamma := r^{-1}\log_2(r+1)$.*

The deduction of Theorem 2.3 from Proposition B.1 is a straightforward 'tensorisation' argument, but no details are given in either [BKMP88] or [HS85]. For the convenience of the reader, we give the deduction below, claiming no originality whatsoever.

Proposition B.1 (and, hence, Theorem 2.3) was established by Landau, Logan and Shepp [LLS85], and 3 years later but seemingly independently (and in a more elementary fashion) by Brown, Keane, Moran and Pearce [BKMP88]. A discussion of the history of these and related problems is given by Brown [Bro88] but this appears to overlook [LLS85].

Finally, we note that a result that is weaker in the exponent than Theorem 2.3, but quite sufficient for the purpose of proving the qualitative form of Theorem 1.1, follows by an iterated application of a result of Gowers and Karam [GK22, Proposition 3.1]. This avoids the need for the delicate analytic inequality in Proposition B.1. Let us also note that the context in which Gowers and Karam use this result is in some ways analogous to ours, albeit in a very different setting.

*Proof of Theorem 2.3, assuming Proposition B.1.* As stated in [BKMP88], one may proceed in a manner 'parallel' to arguments in [BM83], specifically the proof of Lemma 2.6 there. We proceed by induction on $n$. First we check the base case $n = 1$. Here, one may assume without loss of generality that $A_1 = \cdots = A_s = \{0,1\}$ and $A_{s+1} = \cdots = A_r = \{1\}$ for some $s$, $0 \leqslant s \leqslant r$. The density of $A_1 + \cdots + A_r$ in $\{0,1,\ldots,r\}$ is then $(s+1)/(r+1)$, whilst $\alpha_1 = \cdots = \alpha_s = 1$ and $\alpha_{s+1} = \cdots = \alpha_r = 1/2$. The inequality to be checked is thus $(s+1)/(r+1) \geqslant 2^{-(r-s)\gamma}$. However, taking $x_1 = \cdots = x_s = 1/2$ and $x_{s+1} = \cdots = x_r = 0$ in Proposition B.1 yields $(s+1)2^{-s\gamma} \geqslant 1$. Since $2^{r\gamma} = r+1$, the desired inequality follows.

Now assume the result is true for $n-1$. Let $A_i^0$ be the elements of $A_i$ with first coordinate 0, and $A_i^1$ the elements of $A_i$ with first coordinate 1. Suppose that $|A_i^0| = x_i|A_i|$, and without loss of generality suppose that $x_1 \geqslant x_2 \geqslant \cdots \geqslant x_r$. Then the sets $A_1^0 + \cdots + A_j^0 + A_{j+1}^1 + \cdots + A_r^1$, $j = 0,\ldots,r$ are disjoint, since the first coordinate of every element of this set is $j$.

It follows that

$$|A_1 + \cdots + A_r| \geqslant \sum_{j=0}^{r} |A_1^0 + \cdots + A_j^0 + A_{j+1}^1 + \cdots + A_r^1|.$$

Note that $A_i^0$ is a subset of a copy $\{0,1\}^{n-1}$ of density $2\alpha_i x_i$, and that $A_i^1$ is a subset of (a translate of) $\{0,1\}^{n-1}$ of density $2\alpha_i(1-x_i)$.

By the inductive hypothesis,

$$\begin{aligned}|A_1^0 + \cdots + A_j^0 + A_{j+1}^1 + \cdots + A_r^1| &\geqslant (2^r\alpha_1\cdots\alpha_r x_1\cdots x_j(1-x_{j+1})\cdots(1-x_r))^\gamma (r+1)^{n-1}\\ &= (r+1)^n(\alpha_1\cdots\alpha_r)^\gamma(x_1\cdots x_j(1-x_{j+1})\cdots(1-x_r))^\gamma.\end{aligned}$$

Performing the sum over $j$ and applying Proposition B.1, the result follows. □

## Appendix C. A diophantine lemma

The following is a fairly standard type of lemma arising in applications of the circle method and is normally attributed to Vinogradov. We make no attempt to optimise the constants, contenting ourselves with a version sufficient for our purposes in the main paper.



https://doi.org/10.1112/S0010437X24007723 Published online by Cambridge University Press

Lemma C.1. *Suppose that $\alpha \in \mathbf{R}$ and that $L \geqslant 1$ is an integer. Suppose that $\delta_1, \delta_2$ are positive real numbers satisfying $\delta_2 \geqslant 32\delta_1$, and suppose that there are at least $\delta_2 L$ elements $n \in \{1, \dots, L\}$ for which $\|\alpha n\| \leqslant \delta_1$. Suppose that $L \geqslant 16/\delta_2$. Then there is some positive integer $q \leqslant 16/\delta_2$ such that $\|\alpha q\| \leqslant \delta_1 \delta_2^{-1} L^{-1}$.*

*Proof.* Write $S \subseteq \{1, \dots, L\}$ for the set of all $n$ such that $\|\alpha n\| \leqslant \delta_1$; thus, $|S| \geqslant \delta_2 L$. By Dirichlet's lemma, there is a positive integer $q \leqslant 4L$ and an $a$ coprime to $q$ such that $|\alpha - a/q| \leqslant 1/4Lq$. Write $\theta := \alpha - a/q$; thus,

$$|\theta| \leqslant \frac{1}{4Lq}. \tag{C.1}$$

The remainder of the proof consists of 'bootstrapping' this simple conclusion. First, we tighten the bound for $q$, and then the bound for $|\theta|$.

Suppose that $n \in S$. Then, by (C.1), we see that

$$\left\| \frac{an}{q} \right\| \leqslant \delta_1 + \frac{1}{4q}. \tag{C.2}$$

Now we bound the number of $n \in \{1, \dots, L\}$ satisfying (C.2) in a different way. Divide $\{1, \dots, L\}$ into $\leqslant 1 + L/q$ intervals of length $q$. In each interval, $\frac{an}{q} (\mathrm{mod}\ 1)$ ranges over each rational (mod 1) with denominator $q$ precisely once. At most $2q(\delta_1 + 1/4q) + 1 < 2(\delta_1 q + 2)$ of these rationals $x$ satisfy $\|x\| \leqslant \delta_1 + \frac{1}{4q}$. Thus, the total number of $n \in \{1, \dots, L\}$ satisfying (C.2) is bounded above by $2(L/q + 1)(\delta_1 q + 2) = 2\delta_1 L + 2\delta_1 q + \frac{4L}{q} + 4$. It follows that

$$2\delta_1 L + 2\delta_1 q + \frac{4L}{q} + 4 \geqslant \delta_2 L. \tag{C.3}$$

Using $\delta_2 \geqslant 32\delta_1$, $q \leqslant 4L$ and $L \geqslant 16/\delta_2$, one may check that the first, second and fourth terms on the left are each at most $\delta_2 L/4$. Therefore, (C.3) forces us to conclude that $4L/q > \delta_2 L/4$, and therefore, $q \leqslant 16/\delta_2$, which is a bound on $q$ of the required strength.

Now we obtain the claimed bound on $\|\alpha q\|$. Note that, by the assumptions and the inequality on $q$ just established, we have $\delta_1 \leqslant \delta_2/32 \leqslant 1/2q$, and so if $n \in S$ then, by (C.2), we have $\|an/q\| < 1/q$, which implies that $q|n$. That is, all elements of $S$ are divisible by $q$. It follows from this and the definition of $\theta$ that if $n \in S$ then $\|\theta n\| = \|\alpha n\| \leqslant \delta_1$. However, since (by (C.1)) we have $|\theta| \leqslant 1/4Lq$, for $n \in \{1, \dots, L\}$, we have $\|\theta n\| = |\theta n|$. Therefore,

$$|\theta n| \leqslant \delta_1 \tag{C.4}$$

for all $n \in S$. Finally, recall that $S$ consists of multiples of $q$ and that $|S| \geqslant \delta_2 L$; therefore, there is some $n \in S$ with $|n| \geqslant \delta_2 q L$. Using this $n$, (C.4) implies that $|\theta| \leqslant \delta_1/q\delta_2 L$, and so finally $\|\alpha q\| \leqslant |\theta q| \leqslant \delta_1/\delta_2 L$. This concludes the proof. □

## Acknowledgements

I thank Zach Hunter and Sarah Peluse for comments on the first version of the manuscript, and the two referees for a careful reading of the paper.

## Conflicts of interest

None.



https://doi.org/10.1112/S0010437X24007723 Published online by Cambridge University Press

Financial support

The author gratefully acknowledges the support of the Simons Foundation (Simons Investigator grant 376201).

Journal information

*Compositio Mathematica* is owned by the Foundation Compositio Mathematica and published by the London Mathematical Society in partnership with Cambridge University Press. All surplus income from the publication of *Compositio Mathematica* is returned to mathematics and higher education through the charitable activities of the Foundation, the London Mathematical Society and Cambridge University Press.

References

Big21 K. D. Biggs, *Efficient congruencing in ellipsephic sets: the quadratic case*, Acta Arith. **200** (2021), 331–348.

Big23 K. D. Biggs, *Efficient congruencing in ellipsephic sets: the general case*, Int. J. Number Theory **19** (2023), 169–197.

BB23 K. D. Biggs and J. Brandes, *A minimalist version of the circle method and Diophantine problems over thin sets*, Preprint (2023), arXiv:2304.07891.

Bon70 A. Bonami, *Étude des coefficients Fourier des fonctions de $L^p(G)$*, Annales de l'Institut Fourier **20** (1970), 335–402.

Bro88 G. Brown, *Some inequalities that arise in measure theory*, J. Austral. Math. Soc. **45** (1988), 83–94.

BKMP88 G. Brown, M. S. Keane, W. Moran and C. E. M. Pierce, *An inequality, with applications to Cantor measures and normal numbers*, Mathematika **35** (1988), 87–94.

BM83 G. Brown and W. Moran, *Raikov systems and radicals in convolution measure algebras*, J. London Math. Soc. **28** (1983), 531–542.

CTV06 K. P. Costello, T. C. Tao and V. H. Vu, *Random symmetric matrices are almost surely nonsingular*, Duke Math. J. **135** (2006), 395–413.

DS16 S. Drappeau and X. Shao, *Weyl sums, mean value estimates, and Waring's problem with friable numbers*, Acta Arith. **176** (2016), 249–299.

Gow07 W. T. Gowers, *Hypergraph regularity and the multidimensional Szemerédi theorem*, Ann. Math. **166** (2007), 897–946.

GK22 W. T. Gowers and T. Karam, *Equidistribution of high-rank polynomials with variables restricted to subsets of $\boldsymbol{F}_p$*, Preprint (2022), arXiv:2209.04932.

GT10 B. J. Green and T. C. Tao, *Linear equations in primes*, Ann. Math. **171** (2010), 1753–1850.

HS85 D. Hajela and P. Seymour, *Counting points in hypercubes and convolution measure algebras*, Combinatorica **5** (1985), 205–214.

KS20 N. Kirshner and A. Samorodnitsky, *Samorodnitsky on $\ell^4 : \ell^2$ ratio of functions with restricted Fourier support*, J. Combin. Theory Ser. A **172** (2020), 105202.

KT05 A. V. Kumchev and D. I. Tolev, *An invitation to additive prime number theory*, Serdica Math. J. **31** (2005), 1–74.

LLS85 H. J. Landau, B. F. Logan and L. A. Shepp, *An inequality conjectured by Hajela and Seymour arising in combinatorial geometry*, Combinatorica **5** (1985), 337–342.

NS89 M. B. Nathanson and A. Sárközy, *Sumsets containing long arithmetic progressions and powers of 2*, Acta Arith. **54** (1989), 147–154.

Sal21 J. Salmensuu, *A density version of Waring's problem*, Acta Arith. **199** (2021), 383–412.

TT05 J. M. Thuswaldner and R. F. Tichy, *Waring's problem with digital restrictions*, Isr. J. Math. **149** (2005), 317–344.

Vau97 R. Vaughan, *The Hardy–Littlewood method*, Cambridge Tracts in Mathematics, vol. 125, second edition (Cambridge University Press, Cambridge, 1997).

Vu00 V. H. Vu, *On a refinement of Waring's problem*, Duke Math. J. **105** (2000), 107–134.
Woo77 D. R. Woodall, *A theorem on cubes*, Mathematika **24** (1977), 60–62.
Woo03a T. D. Wooley, *On Vu's thin basis theorem in Waring's problem*, Duke Math. J. **120** (2003), 1–34.
Woo03b T. D. Wooley, *On Diophantine inequalities: Freeman's asymptotic formulae*, in *Proc. of the session in analytic number theory and diophantine equations*, Bonner Math. Schriften, eds D. R. Heath-Brown and B. Z. Moroz (Bonn Mathematical Publications Universität Bonn, Mathematisches Institut, Bonn, 2003), vol. 360.

Ben Green ben.green@maths.ox.ac.uk
Mathematical Institute, Andrew Wiles Building, Woodstock Rd, Oxford OX2 6GG, UK



https://doi.org/10.1112/S0010437X24007723 Published online by Cambridge University Press