Digital almost nets

Boris Bukh* Ting-Wei Chao*

Department of Mathematical Sciences
Carnegie Mellon University
Pittsburgh, PA 15213, U.S.A.

bbukh@math.cmu.edu, tchao2@andrew.cmu.edu

Submitted: Oct 11, 2021; Accepted: Sep 20, 2022; Published: Nov 18, 2022

Abstract

Digital nets (in base 2) are the subsets of $[0,1]^d$ that contain exactly the expected number of points in every not-too-small dyadic box. We construct finite sets, which we call "almost nets", such that every such dyadic box contains *almost* the expected number of points from the set, but whose size is exponentially smaller than the one of nets. We also establish a lower bound on the size of such almost nets.

Mathematics Subject Classifications: 05D99, 11K38, 52C10, 65D99

1 Introduction

We call a subinterval of $[0,1]$ *basic (in base q)* if it is of the form $\left[\frac{a}{q^k}, \frac{a+1}{q^k}\right)$, for nonnegative integers a and k. A *basic box* is a product of basic intervals, i.e., a set of the form $\prod_{i=1}^{d}\left[\frac{a_i}{q^{k_i}}, \frac{a_i+1}{q^{k_i}}\right)$. If $q=2$, a basic interval is called a dyadic interval, and a basic box is called a dyadic box.

We say that a set $P \subset [0,1]^d$ is a (m,ε)-*almost net in base q* if it is of size $|P| = q^n m$ for some natural number n and

$$(1-\varepsilon)m \leqslant |\beta \cap P| \leqslant (1+\varepsilon)m$$

for every basic box β of volume $\mathrm{vol}(\beta) = q^{-n}$.

In this paper, we are interested in constructions where the parameters m and ε are independent of n. In contrast, since the family of all axis-parallel boxes has finite VC-dimension, one can construct (m,ε)-almost nets with m linear in n by sampling the

*Supported in part by U.S. taxpayers through NSF CAREER grant DMS-1555149 and NSF grant DMS-2154063.

points of P at random, see [8, 6]. The dependence on n is unavoidable for points sampled at random.

The case $\varepsilon = 0$ of the above definition has been well studied. If $\varepsilon = 0$, almost nets are known as *digital* (t, m, s)*-nets* or simply (t, m, s)*-nets* (in base q) in the literature[1]. The adjective 'digital' is due to the fact that basic intervals comprise of numbers with specified initial digits in base q. They are used extensively in discrepancy theory and numerical integration algorithms, and are subject to numerous works, including a book devoted exclusively to them [7]. It is known from [12, Theorem 3] that, for each d, there exist arbitrarily large $(m, 0)$-almost nets with $m \leqslant q^{5d}$, if q is a prime power. On the other hand, m must grow exponentially with d for large enough nets [9] (see also [10] for asymptotic analysis of the bound in [9]).

In contrast to these results, for $\varepsilon > 0$, we construct (m, ε)-almost nets with m being only polynomial in d.

Theorem 1. *For any prime q, any $d \geqslant 2$, and any positive integers m, n satisfying $m \geqslant 400d\log(dq)$, there exists a set $P \subset [0,1]^d$ of size mq^n such that, for any basic box β of volume q^{-n},*

$$\left(1 - 10\sqrt{\frac{d\log(dq)}{m}}\right) m \leqslant |\beta \cap P| \leqslant \left(1 + 10\sqrt{\frac{d\log(dq)}{m}}\right) m. \tag{1}$$

In particular, for every $0 < \varepsilon < 1/2$ and every $d \geqslant 2$, there exist arbitrarily large (m, ε)-almost nets in base q with $m \leqslant 100\varepsilon^{-2} d\log(dq)$.

Furthermore, the set P satisfying (1) can be chosen to be an $\big(M, 0\big)$-net in base q with $M \leqslant d^{4d} q^{6d} m$.

This result has an application in geometric Ramsey theory: A *convex hole* in a finite set $S \subset \mathbb{R}^d$ is a subset $H \subset S$ in convex position and whose convex hull contains no other point of S. An old problem of Valtr [11] asks for the largest $h(d)$ such that every sufficiently large $S \subset \mathbb{R}^d$ in general position contains a hole of size $h(d)$. Using Theorem 1 one can show that $h(d) \leqslant 4^{d+o(d)}$, which is an improvement over the bound of $h(d) \leqslant 2^{7d}$ that can be obtained from (t, m, s)-nets. The details of both bounds are in [5].

The construction behind Theorem 1 is a minor modification on the construction in [4]. Whereas the construction in [4] uses primes in $\mathbb{Z}$, this construction uses irreducible polynomials in $\mathbb{F}_q[x]$. The reason for this change is to make the denominators be powers of the same prime q. Furthermore, because the addition in $\mathbb{F}_q[x]$ satisfies the ultrametric inequality (with respect to the degree), and because we do not need to worry about boxes that are not basic, several details in the new construction are simpler. As such, we do not make any claims about the novelty. Our purpose in writing the present note is to record the details of the construction for its application to convex holes. We also hope that almost nets will find applications in many other areas that currently use the conventional nets.

[1] The parameters t, m, s in the definition of (t, m, s)-nets have different meaning than in the present paper. They correspond to $\log_q m$, $\log_q(mq^n)$ and d respectively in our notation.

We do not know when the bound in Theorem 1 is sharp. The following is the best lower bound we were able to prove. Its dependence on ε is close to optimal, as long as ε is not too small, but the dependence on d is poor. In the special case $\varepsilon = 0$, we recover the lower bound $t = \Omega(s)$ in (t, m, s)-nets via a proof different than those in [9, 10] (keeping in mind that that t and s correspond to $\log_q m$ and d respectively).

Theorem 2. *Assume that there exists an (m, ε)-net $P \subseteq [0,1]^d$ in base q, then the following holds.*

If $\varepsilon \geqslant 1/2\sqrt{d}$, then

$$m = \Omega\big(\frac{\log d}{q^2 \varepsilon^2 \log(1/\varepsilon)}\big).$$

If $1/2\sqrt{d} \geqslant \varepsilon \geqslant e^{-d/8}$, then

$$m = \Omega(q^{-2k-2}\varepsilon^{-2}),$$

where $k = \frac{2\log(1/\varepsilon)}{\log d - \log\log(1/\varepsilon)}$.

In particular, if $\varepsilon = \omega(d^{-t})$ for some constant t, then we have $m = \Omega_{q,t}(1/\varepsilon^2)$.

If $\varepsilon = o(e^{-cd})$ for some constant c such that $0 < c < \min(1/8, 1/q^2)$, then we get an exponential lower bound $m = \Omega\big(q^{-2}e^{c'd}\big)$, where $c' = 2c(1 - 2\log q/\log(1/c))$.

Open problem. It would be interesting to prove a result similar to Theorem 1 which applies to all boxes, not only to basic boxes. It is possible to construct a set P for which $|\beta \cap P|$ is lower-bounded by $(1 - o(1))m$ for all boxes β by taking a union of translates of the set P from Theorem 1 in a manner similar to that in the second part of the proof of [4, Theorem 2]. However, we have been unable to control $|\beta \cap P|$ from above.

Acknowledgment. We are thankful to Ron Holzman for useful discussions, and to two anonymous referees for their help in improving this paper.

2 Proof of Theorem 1

We denote by $\mathbb{F}_q$ the finite field consisting of elements $0, 1, 2, \ldots, q-1$ equipped with the usual mod-q arithmetic. Let $t \stackrel{\text{def}}{=} \lceil 2\log_q d + 2 \rceil$. Since the number of irreducible polynomials of degree t in $\mathbb{F}_q[x]$ is

$$\frac{1}{t}\big(\sum_{i|t} \mu(i) q^{t/i}\big) \geqslant \frac{1}{t}(q^t - q^{t/2+1}) \geqslant d,$$

we may pick d distinct irreducible polynomials $p_1, \ldots, p_d$ of degree t in $\mathbb{F}_q[x]$. We fix some such choice of polynomials for the duration of the proof. We associate each of these d polynomials to the respective coordinate direction. We will be interested in *canonical boxes*, which are the boxes of the form

$$B = \prod_{i=1}^{d} \left[\frac{a_i}{q^{k_i t}}, \frac{a_i + 1}{q^{k_i t}}\right).$$

for some nonnegative integers k_i and $0 \leqslant a_i < q^{k_i t}$, $i = 1, 2, \ldots, d$.

We say that a polynomial $f \in \mathbb{Z}[x]$ is a *basic polynomial* if $\deg f < t$ and all of its coefficients are in $\{0, 1, \ldots, q-1\}$.

For an irreducible polynomial $p \in \mathbb{F}_q[x]$ of degree t and a polynomial $f \in \mathbb{F}_q[x]$, we define the *base-p expansion of f* to be $f = f_0 + f_1 p + \cdots + f_\ell p^\ell$, where each f_i is a basic polynomial. Put $r_p(f) \stackrel{\text{def}}{=} (f_0(1/q) + f_1(1/q)q^{-t} + \cdots + f_\ell(1/q)q^{-\ell t})/q$, where we view the basic polynomials $f_0, f_1, \ldots, f_\ell$ as polynomial functions on $\mathbb{R}$. In other words, if $f_i = \sum_{j<t} c_{i,j} x^j$ with $c_{i,j} \in \{0, 1, \ldots, q-1\}$, then, denoting by C_i the concatenation $c_{i,0}c_{i,1} \ldots c_{i,t-1}$, the base-$q$ expansion of the real number $r_p(f)$ is

$$r_p(f) = 0.C_0 C_1 \ldots C_\ell.$$

Note that $r_p(f) \in [0, 1)$ for every $f \in \mathbb{F}_q[x]$. Define the function $r \colon \mathbb{F}_q[x] \to [0, 1]^d$ by $r(f) \stackrel{\text{def}}{=} \big(r_{p_1}(f), \ldots, r_{p_d}(f)\big)$.

Recall that our aim is to construct a set $P \subset [0, 1]^d$ whose intersection with any basic box of volume q^{-n} has almost the expected number of points.

Definition 3. We say that a box β is *good* if β is a basic box of volume $\operatorname{vol}(\beta) = q^{-n}$. Let B be the smallest canonical box containing β. We call (B, β) a *good pair*.

Note that if (B, β) is a good pair, then $\operatorname{vol}(B) \leqslant q^{-n+dt-1}$. Indeed, every basic interval is contained in an interval of the form $[a/q^{kt}, (a+1)/q^{kt})$ that is at most q^{t-1} times larger, and therefore $\operatorname{vol}(B) \leqslant q^{-n}(q^{t-1})^d \leqslant q^{-n+dt-1}$.

Suppose B is a canonical box. Write it as $B = \prod_i \big[a_i/q^{k_i t}, (a_i+1)/q^{k_i t}\big)$, and consider $r^{-1}(B)$. The set $r^{-1}(B) \subseteq \mathbb{F}_q[x]$ consists of all solutions to the system

$$\begin{aligned} f &\equiv a_1' \pmod{p_1^{k_1}}, \\ f &\equiv a_2' \pmod{p_2^{k_2}}, \\ &\vdots \\ f &\equiv a_d' \pmod{p_d^{k_d}}, \end{aligned}$$

where $a_i' = f_{i,0} + f_{i,1}p_i + \cdots + f_{i,k_i-1}p_i^{k_i-1}$ and $f_{i,0}, f_{i,1}, \ldots, f_{i,k_i-1}$ are the unique basic polynomials satisfying $a_i/q^{k_i t} = (f_{i,0}(1/q) + f_{i,1}(1/q)q^{-t} + \ldots + f_{i,k_i-1}(1/q)q^{-(k_i-1)t})/q$.

By the Chinese Remainder theorem, the set $r^{-1}(B)$ is of the form $A(B) + D(B)\mathbb{F}_q[x]$ where $D(B) \stackrel{\text{def}}{=} p_1^{k_1} p_2^{k_2} \cdots p_d^{k_d}$ and $A(B)$ is the unique element in $r^{-1}(B)$ of degree less than $t(k_1 + \ldots + k_d)$. Note that $\deg D(B) = t(k_1 + \ldots + k_d) = -\log_q(\operatorname{vol}(B))$.

Given a good pair (B, β), define

$$L_B(\beta) \stackrel{\text{def}}{=} \{g \in \mathbb{F}_q[x] : r\big(A(B) + gD(B)\big) \in \beta\}.$$

Claim 4. *The set $\mathcal{L} \stackrel{\text{def}}{=} \{L_B(\beta) : (B, \beta)$ is a good pair$\}$ is of size at most q^{4dt}.*

Proof. Let (B,β) be a good pair. Write B and β in the form

$$B = \prod_{i=1}^{d} \left[\frac{a_i}{q^{k_i t}}, \frac{a_i+1}{q^{k_i t}} \right), \qquad \beta = \prod_{i=1}^{d} \left[\frac{a_i}{q^{k_i t}} + \frac{b_i}{q^{(k_i+1)t}}, \frac{a_i}{q^{k_i t}} + \frac{c_i}{q^{(k_i+1)t}} \right).$$

The condition $r\big(A(B)+gD(B)\big) \in \beta$ is equivalent to

$$\begin{aligned} A(B)+gD(B) &\in a_1' + p_1^{k_1} J_1 \pmod{p_1^{k_1+1}},\\ A(B)+gD(B) &\in a_2' + p_2^{k_2} J_2 \pmod{p_2^{k_2+1}},\\ &\vdots\\ A(B)+gD(B) &\in a_d' + p_d^{k_d} J_d \pmod{p_d^{k_d+1}}, \end{aligned}$$

where the sets J_i consist of the basic polynomials f such that $f(1/q)q^{t-1} \in [b_i, c_i)$.

On the other hand,

$$\begin{aligned} A(B)+gD(B) &\equiv a_1' + (\alpha_1 + g\delta_1)p_1^{k_1} \pmod{p_1^{k_1+1}},\\ A(B)+gD(B) &\equiv a_2' + (\alpha_2 + g\delta_2)p_2^{k_2} \pmod{p_2^{k_2+1}},\\ &\vdots\\ A(B)+gD(B) &\equiv a_d' + (\alpha_d + g\delta_d)p_d^{k_d} \pmod{p_d^{k_d+1}} \end{aligned}$$

for some $\alpha_i, \delta_i \in \mathbb{F}_q[x]/(p_i), i = 1,2,\dots,d$. Since $\dim_{\mathbb{F}_q} \mathbb{F}_q[x]/(p_i) = \deg p_i = t$, there are at most q^{2dt} different choices for $(\alpha_i,\delta_i)_{i=1}^d$. Also, there are at most q^{2dt} different choices for $(b_i,c_i)_{i=1}^d$ satisfying $0 \leqslant b_i < c_i \leqslant q^t$. Since $L_B(\beta)$ is determined by $(\alpha_i,\delta_i,b_i,c_i)_{i=1}^d$, the claim is true. $\square$

To each canonical box B of volume between q^{-n} and $q^{-n+dt-1}$ inclusive we assign a *type*, so that boxes of the same type behave similarly. Formally, let $\mathcal{A}(B)$ be the polynomial obtained from the polynomial $A(B)$ by setting the coefficients of $1, x, x^2, \dots, x^{n-dt-1}$ to zero. Similarly, let $\mathcal{D}(B)$ be the polynomial obtained from $D(B)$ by setting the coefficients of $1, x, x^2, \dots, x^{n-3dt}$ to zero. The type of B is then the pair $\mathcal{T}(B) \stackrel{\text{def}}{=} \big(\mathcal{A}(B), \mathcal{D}(B)\big)$.

Note that, from $q^{-n} \leqslant \operatorname{vol}(B) \leqslant q^{-n+dt-1}$ and $\deg D(B) = -\log_q(\operatorname{vol}(B))$ it follows that

$$n - dt + 1 \leqslant \deg \mathcal{D}(B) \leqslant n. \tag{2}$$

Claim 5. *The number of types is at most q^{4dt}.*

Proof. Since $\deg A(B) < \deg D(B) \leqslant n$, only the dt (resp. $3dt$) leading coefficients of $\mathcal{A}(B)$ (resp. $\mathcal{D}(B)$) may be non-zero. So, there are no more than $q^{dt} \times q^{3dt} = q^{4dt}$ types. $\square$

For a type $\mathcal{T} = (\mathcal{A},\mathcal{D})$, let $\mathcal{Y}(\mathcal{T}) \stackrel{\text{def}}{=} \{\mathcal{A} + g\mathcal{D} : g \in \mathbb{F}_q[x]\}$. Note that if $\mathcal{T} = \mathcal{T}(B)$, then $\mathcal{Y}(\mathcal{T})$ is an approximation to $r^{-1}(B)$. That is to say, the respective elements of $\mathcal{Y}(T)$ and of $r^{-1}(B)$ differ only in low-degree coefficients.

Let I_k denote polynomials of degree less than k in $\mathbb{F}_q[x]$. Our construction will be a union of sets of the form $h + I_{n-dt}$ where $\deg h \leqslant n + dt$.

We first prove that there is no difference in how the sets $\mathcal{Y}(\mathcal{T})$ and $r^{-1}(B)$ intersect $h + I_{n-dt}$.

Claim 6. *Suppose $\mathcal{T}(B) = \big(\mathcal{A}(B), \mathcal{D}(B)\big)$. Then for any polynomial $h \in I_{n+dt}$ and any polynomial g, $\mathcal{A}(B) + g\mathcal{D}(B) \in h + I_{n-dt}$ if and only if $A(B) + gD(B) \in h + I_{n-dt}$.*

Proof. If $\mathcal{A}(B)+g\mathcal{D}(B) \in h+I_{n-dt}$, then $\deg(\mathcal{A}(B)+g\mathcal{D}(B)) < n+dt$. Since $\deg \mathcal{A}(B) < n$ and $\deg \mathcal{D}(B) \geqslant n - dt$, it follows that $\deg g < 2dt$. From the definition of $\mathcal{A}(B)$ and $\mathcal{D}(B)$, the coefficients of $x^{n-dt}, x^{n-dt+1}, \dots$ in $\mathcal{A}(B)+g\mathcal{D}(B)$ are the same as the respective coefficients in $A(B) + gD(B)$. The opposite direction is similar. $\square$

For a type $\mathcal{T}$ and $L \in \mathcal{L}$ that satisfy $\mathcal{T} = \mathcal{T}(B)$ and $L = L_B(\beta)$ for some good pair (B, β), define

$$\mathcal{Y}_\mathcal{T}(L) \stackrel{\text{def}}{=} \{\mathcal{A} + g\mathcal{D} : g \in L\}.$$

With this definition, $\mathcal{Y}_\mathcal{T}(L)$ is the approximation to $r^{-1}(\beta)$ induced by the approximation $\mathcal{Y}(\mathcal{T})$ to $r^{-1}(B)$.

Claim 7. *The set $\overline{\mathcal{Y}}_\mathcal{T}(L) \stackrel{\text{def}}{=} \mathcal{Y}_\mathcal{T}(L) \cap I_{n+dt}$ is of size exactly q^{dt}.*

Proof. Let (B, β) be a good pair such that $\mathcal{T} = \mathcal{T}(B)$ and $L = L_B(\beta)$. From the previous claim, we know that the size of $\overline{\mathcal{Y}}_\mathcal{T}(L)$ is the same as the size of $r^{-1}(\beta) \cap I_{n+dt}$. By the Chinese remainder theorem, each of the canonical boxes of volume $q^{-(\lfloor n/t \rfloor + d)t}$ contains equally many points from $r(I_{n+dt})$. Since $n \leqslant (\lfloor n/t \rfloor + d)t$, the number of points in $\beta \cap r(I_{n+dt})$ is equal to $q^{n+dt} \operatorname{vol}(\beta) = q^{dt}$. $\square$

Claim 8. *Let h be chosen uniformly from I_{n+dt}. Then $|\overline{\mathcal{Y}}_\mathcal{T}(L) \cap (h + I_{n-dt})|$ is 1 with probability q^{-dt} and is 0 otherwise.*

Proof. Let $u \in \overline{\mathcal{Y}}_\mathcal{T}(L)$ be arbitrary. Clearly $\Pr[u \in h + I_{n-dt}] = q^{-2dt}$. The events of the form $u \in h + I_{n-dt}$ are mutually disjoint as u ranges over $\overline{\mathcal{Y}}_\mathcal{T}(L)$. Indeed, suppose $\mathcal{T} = (\mathcal{A}, \mathcal{D})$ and $u, u' \in \overline{\mathcal{Y}}_\mathcal{T}(L)$ are such that $u, u' \in h + I_{n-dt}$ for some $h \in I_{n+dt}$. We may write $u = \mathcal{A} + g\mathcal{D}$ and $u' = \mathcal{A} + g'\mathcal{D}$. Then $u - u' = (g - g')\mathcal{D} \in I_{n-dt}$. Since $\deg \mathcal{D}(B) \geqslant n - dt$, this implies that $g = g'$ and hence $u = u'$.

In the combination with Claim 7, this implies that

$$\Pr\big[\, |\overline{\mathcal{Y}}_\mathcal{T}(L) \cap (h + I_{n-dt})| = 1\big] = q^{-2dt} q^{dt} = q^{-dt}. \qquad \square$$

Sample $q^{dt}m$ elements uniformly at random from I_{n+dt}, independently from one another. Let H be the resulting multiset, and consider the multiset

$$H + I_{n-dt} \stackrel{\text{def}}{=} \{h + f : h \in H, f \in I_{n-dt}\}.$$

For a type $\mathcal{T}$ and $L \in \mathcal{L}$ that satisfy $\mathcal{T} = \mathcal{T}(B)$ and $L = L_B(\beta)$ for some good pair (B, β), define the random variable $N_{\mathcal{T},L} \stackrel{\text{def}}{=} |\overline{\mathcal{Y}}_\mathcal{T}(L) \cap (H + I_{n-dt})|$. This random variable is distributed according to the binomial distribution $\operatorname{Binom}(q^{dt}m, q^{-dt})$.

Let $\varepsilon = \sqrt{33dt \log q/m}$. Note that $\varepsilon < \sqrt{33d(2 \log d + 3 \log q)/m} < 10\sqrt{d \log(dq)/m}$, and in particular $\varepsilon < 1/2$. Hence, $\varepsilon^2/2 - \varepsilon^3/2 \geqslant \varepsilon^2/4$. By the tail bounds for the binomial distribution [3, Theorems A.1.11 and A.1.13] we obtain

$$\Pr\big[\, N_{\mathcal{T},L} - m > \varepsilon m\big] < e^{-(\varepsilon^2/2-\varepsilon^3/2)m} < q^{-8dt}/2,$$

$$\Pr\big[\, N_{\mathcal{T},L} - m < -\varepsilon m\big] < e^{-\varepsilon^2 m/2} < q^{-8dt}/2.$$

From Claims 4 and 5, and the union bound it then follows that there exists a choice of H such that $N_{\mathcal{T},L}$ is bounded between $(1-\varepsilon)m$ and $(1+\varepsilon)m$ whenever $\mathcal{T} = \mathcal{T}(B)$, $L = L_B(\beta)$ and (B, β) is a good pair. By Claim 6, this implies that the number of points in any good box β of volume q^{-n}, the size $\beta \cap r(H + I_{n-dt})$ is bounded between $(1-\varepsilon)m$ and $(1+\varepsilon)m$.

Hence the multiset $r(H + I_{n-dt})$ in $[0,1)^d$ is of size exactly mq^n and satisfies (1). Since the r-image of every set of the form $h + I_{n-dt}$, for $h \in \mathbb{F}_q[x]$, is a $(q^{dt}, 0)$-net, it follows that $r(H + I_{n-dt})$ is a $(M, 0)$-net with $M = q^{2dt}m \leqslant d^{4d}q^{6d}m$.

To obtain a set satisfying the same conclusion, we may perturb the points of the multiset $r(H + I_{n-dt})$ slightly to ensure distinctness.

3 Proof of Theorem 2

We shall derive Theorem 2 from the following lemma.

Lemma 9. *For any positive integers n, d, q and positive real numbers m, ε with $n \geqslant d \geqslant 2$, $q \geqslant 2$, and $\varepsilon < 1/4$, if there exists an (m, ε)-almost net $P \subseteq [0,1]^d$ in base q of size $q^n m$, then*

$$m = \Omega\Big(\frac{\log(\binom{d}{k})}{q^{2k}\varepsilon^2 \log(1/\varepsilon)}\Big),$$

for any integer k such that $1 \leqslant k \leqslant d/2$ and

$$2\varepsilon \geqslant \binom{d}{k}^{-1/2} \tag{3}$$

holds.

Proof. Let $\mathcal{B}$ be the box $[0, 1/q^{n-2k}) \times [0,1)^{d-1}$. For any point $v = (v_1, \ldots, v_d) \in \mathcal{B}$, write its coordinates in base q as $v_\ell = (0.v_{\ell,1}v_{\ell,2}\ldots)_q$. Noting that the first $n-2k$ base-q digits of v_1 are zero, we let $X_1(v)$ be the first non-trivial digit of v_1, i.e., $X_1(v) \overset{\text{def}}{=} v_{1,n-2k+1}$. Similarly, let $X_\ell(v) \overset{\text{def}}{=} v_{\ell,1}$ for $\ell \geqslant 2$.

The proof idea is to use almost independence of functions $X_1, \ldots, X_d$ for a randomly chosen point of $\mathcal{B}$. However, we do not directly appeal to the known bound on the size of probability spaces supporting almost independent random variables (see e.g. [1, 2]) because those bounds are formulated for $\{0,1\}$-valued random variables, whereas $X_1, \ldots, X_d$ take q distinct values.

Let $S \stackrel{\text{def}}{=} P \cap \mathcal{B}$, and $t \stackrel{\text{def}}{=} |S|$. Since P is an (m, ε)-almost net, it follows that t is between $q^{2k}(1-\varepsilon)m$ and $q^{2k}(1+\varepsilon)m$. Assume $v^1, \dots, v^t$ are all the points in S.

For $x \in \mathbb{R}$, let $e_q(x) \stackrel{\text{def}}{=} \exp(2\pi i x/q)$ where $i = \sqrt{-1}$. Let U be a $\binom{d}{k}$-by-t matrix, where the rows are indexed by $\binom{[d]}{k}$ and the columns are indexed by $[t]$. The general entry of U is

$$U_{J,\ell} \stackrel{\text{def}}{=} e_q\big(\sum_{j \in J} X_j(v^\ell)\big).$$

Also, define $A \stackrel{\text{def}}{=} \frac{1}{t} UU^*$.

Claim 10. *The diagonal terms in A are all 1. The off-diagonal terms are, in absolute value, bounded above by 2ε.*

Proof. The general term of A is given by

$$A_{J_1,J_2} = \tfrac{1}{t} \sum_{\ell=1}^{t} e_q\Big(\sum_{j \in J_1} X_j(v^\ell) - \sum_{j \in J_2} X_j(v^\ell)\Big).$$

If $J_1 = J_2$, this is clearly 1.

Suppose $J_1 \neq J_2$. Note that, for any $\alpha = (\alpha_j)_{j \in J_1 \Delta J_2}$ with $\alpha_j \in \{0, 1, \dots, q-1\}$, the set

$$\{v \in \mathcal{B} : X_j(v) = \alpha_j \text{ for } j \in J_1 \Delta J_2\}$$

is a basic box of volume $q^{-n+2k-|J_1 \Delta J_2|}$. Thus, for any $\tau \in \{0, 1, \dots, q-1\}$, the region

$$B_\tau \stackrel{\text{def}}{=} \Big\{v \in \mathcal{B} : \sum_{j \in J_1} X_j(v) - \sum_{j \in J_2} X_j(v) \equiv \tau \pmod{q}\Big\}$$

can be partitioned into $q^{|J_1 \Delta J_2|-1}$ many basic boxes of volume $q^{-n+2k-|J_1 \Delta J_2|}$ each. Since we have $-n + 2k - |J_1 \Delta J_2| \geqslant -n$, it follows that the number of ℓ such that $v^\ell \in B_\tau$ is bounded between $q^{2k-1}(1-\varepsilon)m$ and $q^{2k-1}(1+\varepsilon)m$. Thus,

$$\begin{aligned} |A_{J_1,J_2}| &= \frac{1}{t}\left|\sum_{\tau=1}^{q} |B_\tau \cap S| e_q(\tau)\right| \\ &\leqslant \frac{1}{t}\left|\sum_{\tau=1}^{q} q^{2k-1} m e_q(\tau)\right| + \frac{1}{t}\sum_{\tau=1}^{q} |\varepsilon q^{2k-1} m e_q(\tau)| \\ &= \frac{\varepsilon q^{2k} m}{t} \leqslant 2\varepsilon. \end{aligned}$$

□

We apply [1, Theorem 2.1] to the matrix $(A + \bar{A})/2$. We obtain that, if $\binom{d}{k}^{-1/2} \leqslant 2\varepsilon < 1/2$, then $2q^{2k}(1+\varepsilon)m \geqslant 2\,\mathrm{rank}(A) \geqslant \mathrm{rank}\big((A+\bar{A})/2\big) = \Omega\big(\frac{\log(\binom{d}{k})}{\varepsilon^2 \log(1/\varepsilon)}\big)$. Therefore,

$$m = \Omega\big(\frac{\log(\binom{d}{k})}{q^{2k}\varepsilon^2 \log(1/\varepsilon)}\big).$$

□

The right hand side of lemma 9 is a decreasing function of k for $k \in [1, d/2]$. Therefore, we shall pick k as small as possible. If $\varepsilon \geqslant 1/2\sqrt{d}$, then we may set $k = 1$ and get

$$m = \Omega\big(\frac{\log d}{q^2\varepsilon^2\log(1/\varepsilon)}\big).$$

If $1/2\sqrt{d} \geqslant \varepsilon \geqslant e^{-d/8}$, then we may set $k = \frac{2\log(1/\varepsilon)}{\log d - \log\log(1/\varepsilon)}$. From the assumption on ε, we have $k \leqslant \log(1/\varepsilon)$. Therefore,

$$\begin{aligned}\binom{d}{k} &\geqslant (d/k)^k \\ &\geqslant \exp\big(\frac{2\log(1/\varepsilon)}{\log d - \log\log(1/\varepsilon)}(\log d - \log k)\big) \\ &\geqslant \exp\big(\frac{2\log(1/\varepsilon)}{\log d - \log\log(1/\varepsilon)}(\log d - \log\log(1/\varepsilon))\big) \\ &\geqslant \frac{1}{\varepsilon^2},\end{aligned}$$

and so (3) holds. Hence, we may apply Lemma 9 with $\lceil k \rceil$ in place of k and obtain

$$m = \Omega(q^{-2k-2}\varepsilon^{-2}).$$

In particular, if $\varepsilon = \omega(d^{-t})$ for some constant t, then k is also a constant, and so $m = \Omega_{q,t}(1/\varepsilon^2)$ in this case.

If $\varepsilon = o(e^{-cd})$ for some constant c such that $0 < c < \min(1/8, 1/q^2)$, then the (m, ε)-net is also an (m, e^{-cd})-net, when d is large enough. We may apply the result above with e^{-cd} in place of ε. In this case, the calculations above yield $k = 2cd/\log(1/c)$, and we get $m = \Omega\big(q^{-2}e^{c'd}\big)$ where $c' = 2c\big(1 - 2\log q/\log(1/c)\big)$.

References

[1] Noga Alon. Perturbed identity matrices have high rank: proof and applications. *Combinatorics, Probability and Computing*, 18(1–2):3–15, 2009.

[2] Noga Alon, Alexandr Andoni, Tali Kaufman, Kevin Matulef, Ronitt Rubinfeld, and Ning Xie. Testing k-wise and almost k-wise independence. In *STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, pages 496–505. ACM, New York, 2007.

[3] Noga Alon and Joel H. Spencer. *The Probabilistic Method*. Wiley Publishing, 4th edition, 2016.

[4] Boris Bukh and Ting-Wei Chao. Empty axis-parallel boxes. *Int. Math. Res. Not. IMRN*, (18):13811–13828, 2022. arXiv:2009.05820.

[5] Boris Bukh, Ting-Wei Chao, and Ron Holzman. On convex holes in d-dimensional point sets. *Combin. Probab. Comput.*, 31(1):101–108, 2022. arXiv:2007.08972.

[6] Mónika Csikós and Nabil H. Mustafa. Optimal approximations made easy. *Inform. Process. Lett.*, 176:Paper No. 106250, 6, 2022. arXiv:2008.08970.

[7] Josef Dick and Friedrich Pillichshammer. *Digital nets and sequences*. Cambridge University Press, Cambridge, 2010. Discrepancy theory and quasi-Monte Carlo integration.

[8] Sariel Har-Peled and Micha Sharir. Relative (p, ϵ)-approximations in geometry. *Discrete Comput. Geom.*, 45(3):462–496, 2011. arXiv:0909.0717.

[9] William J. Martin and Terry I. Visentin. A dual Plotkin bound for (T, M, S)-nets. *IEEE Trans. Inform. Theory*, 53(1):411–415, 2007.

[10] Rudolf Schürer. A new lower bound on the t-parameter of (t, s)-sequences. In *Monte Carlo and quasi-Monte Carlo methods 2006*, pages 623–632. Springer, Berlin, 2008.

[11] Pavel Valtr. Sets in $\mathbf{R}^d$ with no large empty convex subsets. *Discrete Math.*, 108(1–3):115–124, 1992. Topological, algebraical and combinatorial structures. Frolík's memorial volume.

[12] Chao Ping Xing and Harald Niederreiter. A construction of low-discrepancy sequences using global function fields. *Acta Arith.*, 73(1):87–102, 1995.