

# EXPLORING THE TOOLKIT OF JEAN BOURGAIN

TERENCE TAO

ABSTRACT. Gian-Carlo Rota asserted in [?] that “every mathematician only has a few tricks”. The sheer breadth and ingenuity in the work of Jean Bourgain may at first glance appear to be a counterexample to this maxim. However, as we hope to illustrate in this article, even Bourgain relied frequently on a core set of tools, which formed the base from which problems in many disparate mathematical fields could then be attacked. We discuss a selected number of these tools here, and then perform a case study of how an argument in one of Bourgain’s papers [?] can be interpreted as a sequential application of several of these tools.

## 1. INTRODUCTION

As the other articles in this collection demonstrate, Jean Bourgain achieved breakthroughs in an astounding number of areas across mathematics, many of which would seem at first glance to be quite unrelated to each other. It is almost beyond belief that a single mathematician could have such deep impact on so many different subfields during his or her career. However, if one compares Bourgain’s works in different topics, some common threads begin to emerge. In particular, one discovers that Bourgain had a certain “toolkit” of basic techniques that he was extremely skilled at using in a great variety of situations. While this toolkit is far from sufficient to explain the powerful nature of his work, it does show how he could at least get started on making progress in so many different mathematical problems. In this article we present a selection of Bourgain’s most commonly used tools: quantification of qualitative estimates, dyadic pigeonholing, random translations, and metric entropy and concentration of measure. These tools often did not originate with Bourgain’s work, but he was able to wield them systematically and apply them to a far broader range of problems than had previously been realized. This is far from a complete list of Bourgain’s standard tools - for instance, many of his works also feature systematic use of the uncertainty principle, or the method of probabilistic decoupling - but it is an illustrative sample of that basic toolkit.

Of course, in many cases Bourgain’s work involved much deeper and delicate arguments than just the general-purpose techniques presented here. Nevertheless, knowledge of these basic tools helps place Bourgain’s arguments in a more systematic framework, in which these preliminary techniques are used to isolate the core features of the problem, which were then attacked by the full force of Bourgain’s intellectual firepower. But sometimes these basic methods are already enough to solve non-trivial problems almost on their own. We illustrate this in the final section (Section ??) by tracing through a

single paper [?] of Bourgain's, giving one of the most general results known on the Erdős similarity problem; as we hope to demonstrate, this non-trivial result can be interpreted as essentially being a sequential application of each of the tools listed here in turn.

## 2. NOTATION

We use  $1_E$  to denote the indicator of a set  $E$ , and use  $|E|$  to denote either the cardinality of  $E$  (if it is finite) or its Lebesgue measure (if it is infinite). If  $E$  lies in a vector space, we use  $x + E := \{x + y : y \in E\}$  for the translate of  $E$  by  $x$ ,  $x - E := \{x - y : y \in E\}$  for the reflection of  $E$  across  $x/2$ , and  $\lambda \cdot E := \{\lambda y : y \in E\}$  for the dilation of  $E$  by  $\lambda$ .

We use  $\mathbf{P}(E)$  to denote the probability of a random event  $E$ , and  $\mathbf{E}X$  and  $\mathbf{Var}(X)$  for the mean and variance of a random variable  $X$ .

We will use the asymptotic notation  $X \lesssim Y$ ,  $Y \gtrsim X$ , or  $X = O(Y)$  to denote the bound  $|X| \leq CY$  for an absolute constant  $C$ . If the constant  $C$  depends on parameters we will indicate this by parameters, for instance  $X \lesssim_{p,d} Y$  denotes the bound  $|X| \leq C_{p,d}Y$  for a constant  $C_{p,d}$  depending only on  $p, d$ . It may be noted that in Bourgain's work the implied constant was sometimes omitted completely from the notation. In the words of Heath-Brown, in his Mathematical Reviews summary of Bourgain's paper [?], "Readers are advised not to take some of the assertions made in the course of the proofs too literally, and, in particular, to abandon any preconceived notion as to the meaning of the symbol  $\sim$ ".

## 3. QUANTITATIVE FORMULATION OF QUALITATIVE PROBLEMS

One can roughly divide analysis into "soft analysis" - the study of qualitative properties (continuity, measurability, integrability, etc.) of infinitary objects, and "hard analysis" - the study of quantitative estimation of finitary objects. Bourgain's toolkit lies almost exclusively in the latter category, so when tackling a "soft analysis" problem, often the first step in one of Bourgain's arguments is to locate a more quantitative "hard analysis" estimate that will imply the desired claim, removing almost all the appearances of limits or arbitrarily large and small scales, and instead working with a large but finite number of scales and focusing on estimates that are uniform with respect to several parameters.

For instance, consider the following result of Furstenberg, Katznelson, and Weiss [?]:

**Theorem 3.1** (Furstenberg-Katznelson-Weiss theorem, qualitative version). *Let  $A \subset \mathbb{R}^2$  be a measurable set whose upper density  $\delta := \limsup_{R \rightarrow \infty} \frac{|A \cap B(0, R)|}{|B(0, R)|}$  is positive. Then there exists  $l_0$  such that for all  $l \geq l_0$ , there exist  $x, y \in A$  with  $|x - y| \geq l$ .*

Note that this theorem does not provide any quantitative bound for the length threshold  $l_0$  in terms of the upper density  $\delta$ . Indeed, such a bound is not possible, since if one replaces  $A$  by a rescaled version  $\lambda \cdot A := \{\lambda x : x \in A\}$  then the length threshold  $l_0$

will be replaced by  $\lambda_0$ , while the upper density  $\delta$  remains unchanged. As such, one may be tempted to conclude that Theorem ?? is irredeemably “qualitative” in nature. Nevertheless, in [?], Bourgain gave a new proof of this theorem (as well as several novel generalisations) by first establishing the following quantitative analogue:

**Theorem 3.2** (Furstenberg-Katznelson-Weiss theorem, quantitative version). *Let  $0 < \varepsilon < \frac{1}{2}$ , let  $B \subset [-1, 1]^2$  have measure  $|B| \geq \varepsilon$ , and let  $J = J(\varepsilon)$  be a sufficiently large natural number depending on  $\varepsilon$ . Suppose that  $0 < t_J < \dots < t_1 \leq 1$  are a sequence of scales with  $t_{j+1} \leq t_j/2$  for all  $1 \leq j < J$ . Then for at least one  $1 \leq j \leq J$ , one has*

$$\int_{\mathbb{R}^2} \int_{S^1} 1_B(x) 1_B(x + t_j \omega) \, d\sigma(\omega) dx \gtrsim \varepsilon^2, \quad (3.1)$$

where  $d\sigma$  is surface measure on the unit circle  $S^1$ , normalised to have unit mass.

Note that there is no longer any appearance of limits in Theorem ??, and instead of working with an infinity of scales  $l$ , one now works with a finite number of scales  $t_1, \dots, t_J$ . Furthermore, there is a uniform bound on the number  $J$  of scales involved depending on  $\varepsilon$ ; the arguments in [?] in fact give the explicit dependence  $J = O(\frac{1}{\varepsilon} \log \frac{1}{\varepsilon})$  on  $\varepsilon$ .

Let us see why Theorem ?? implies Theorem ?. Assume for contradiction that Theorem ? failed, then one can find a set  $A \subset \mathbb{R}^2$  of some positive upper density  $\delta > 0$ , and a sequence of scales  $l_1 < l_2 < \dots$  going to infinity such that for each  $l_j$  there are no  $x, y \in A$  with  $|x - y| = l_j$ . We can sparsify this sequence of scales so that  $l_{j+1} \geq 2l_j$  for all  $j$ . Let  $\varepsilon > 0$  be sufficiently small (depending only on  $\delta$ ), and let  $J$  be as in Theorem ?. As  $A$  has density  $\delta$ , one can find a radius  $R > l_J$  such that  $|A \cap B(0, R)| \gtrsim \delta R^2$ . If we now consider the rescaling  $B := \{x \in [-1, 1]^2 : Rx \in A\}$ , and define the rescaled scales  $0 < t_J < \dots < t_1 \leq 1$  by  $t_j := l_{J+1-j}/R$  for  $j = 1, \dots, J$ , then  $|B| \gtrsim \delta$  and  $t_{j+1} \leq t_j/2$  for all  $1 \leq j < J$ , and for any  $1 \leq j \leq J$  there are no points  $x, y \in A_R$  with  $|x - y| = t_j$ . In particular the left-hand side of (??) vanishes for all  $1 \leq j \leq J$ , and one then contradicts Theorem ? after choosing the parameters appropriately. Note how this argument fails to give any bound on  $l_0$  (as it must), despite being “quantitative” in nature; it gives quantitative bounds on the *number* of genuinely different scales  $l_j$  at which the conclusion of Theorem ? fails, but does not bound the *magnitude* of these scales.

We will sketch the proof of Theorem ? in the next section. For now, we turn to another example of Bourgain’s strategy of attacking qualitative results through quantitative methods. We focus on a specific pointwise ergodic theorem established in [?, Theorem 1]:

**Theorem 3.3** (Bourgain’s pointwise ergodic theorem along squares). *Let  $(X, \mu)$  be a probability space with a measure-preserving transformation  $T : X \rightarrow X$ . Let  $f \in L^r(X, \mu)$  for some  $r > 1$ . Then the averages*

$$A_N f(x) := \frac{1}{N} \sum_{n=1}^N f(T^{n^2} x)$$

converge pointwise as  $N \rightarrow \infty$  for  $\mu$ -almost every  $x \in X$ .

In fact one can replace the squares here by any other polynomial with integer coefficients, but we focus on the squares for sake of concreteness. A standard method to establish almost everywhere convergence results in ergodic theory is to establish an inequality for the associated maximal function

$$Mf(x) := \sup_{N>0} |A_N f(x)|, \quad (3.2)$$

and combine this with an almost everywhere convergence result for functions  $f$  in a dense subclass of the original space  $L^r(X, \mu)$  of interest (e.g.,  $L^\infty(X, \mu)$ ). In [?] a maximal inequality for (??) was established, thus reducing matters to consideration of functions  $f$  in the dense subclass  $L^\infty(X, \mu)$ , but new ideas were needed to handle this subclass. Bourgain achieved this through the following variant of a maximal inequality:

**Theorem 3.4** (Bourgain's variational estimate). *Let  $\lambda > 1$ , let  $\varepsilon > 0$ , and let  $J$  be sufficiently large depending on  $\lambda, \varepsilon$ . Then for any  $1 < N_1 < N_2 < \dots < N_J$ , any  $(X, \mu, T)$  as in Theorem ??, then for any  $f \in L^\infty(X, \mu)$ , one has*

$$\sum_{j=1}^{J-1} \left\| \sup_{N_j \leq N \leq N_{j+1}: N \in Z_\lambda} |A_N f - A_{N_j} f| \right\|_{L^2(X, \mu)} \lesssim \varepsilon J \|f\|_{L^2(X, \mu)}$$

where  $Z_\lambda := \{\lfloor \lambda^n \rfloor : n \in \mathbb{N}\}$ .

Note that  $L^2$  boundedness of the maximal function (??) would suffice to establish Theorem ?? were it not for the additional factor of  $\varepsilon$  on the right-hand side; thus we now sketch how Theorem ?? implies Theorem ???. By the previous discussion we may assume  $f \in L^\infty(X, \mu)$ , and may normalise  $\|f\|_{L^\infty(X, \mu)} = 1$  and take  $f$  to be real-valued. By rounding a natural number  $N$  to the nearest element  $N'$  of  $Z_\lambda$  for some  $\lambda > 1$  we see that  $A_N f(x) = A_{N'} f(x) + O(\lambda - 1)$ . From this it is not difficult to see that we only need to establish pointwise almost everywhere convergence of  $A_N f(x)$ ,  $N \in Z_\lambda$  for any fixed  $\lambda > 1$ . Suppose this claim failed, then there would be a set  $E$  of positive measure in  $X$  and a  $\delta > 0$  such that

$$\limsup_{N \in Z_\lambda} A_N f(x) - \liminf_{N \in Z_\lambda} A_N f(x) \geq \delta$$

for all  $x \in E$ . By standard measure theory arguments one can then recursively construct an infinite sequence  $N_1 < N_2 < \dots$  and a subset  $E'$  of  $E$  of positive measure such that

$$\sup_{N_j \leq N \leq N_{j+1}: N \in Z_\lambda} |A_N f(x) - A_{N_j} f(x)| \gtrsim \delta$$

for all  $j \geq 1$  and  $x \in E'$ . This then can be used to contradict Theorem ?? after selecting  $\varepsilon$  small enough and  $J$  large enough.

The bulk of the paper [?] is occupied with the task of establishing estimates such as Theorem ??, which proceeds by transferring the problem to the integers, applying Fourier-analytic decompositions, and establishing some maximal inequalities of a harmonic analysis flavor. These arguments have proven to be quite influential, and a paradigm for establishing many other pointwise ergodic theorems, but we will not survey these developments here.

## 4. DYADIC PIGEONHOLING

One of the oldest tricks in analysis is that of *dyadic decomposition*: when faced with a sum or integral over a parameter ranging over a wide range of scales, first control the contribution of an individual dyadic scale (such as when the magnitude of the parameter ranges between two fixed consecutive powers  $2^k, 2^{k+1}$  of two), and then sum over all dyadic scales. For instance, we have the Cauchy condensation test: when asked to determine whether a series  $\sum_{n=1}^{\infty} f(n)$  is absolutely convergent, where  $f : \mathbb{N} \rightarrow \mathbb{R}^+$  is non-negative and non-increasing, one can break up the sum dyadically

$$\sum_{n=1}^{\infty} f(n) = \sum_{k=0}^{\infty} \sum_{2^k \leq n < 2^{k+1}} f(n)$$

and then observe that each dyadic component can be easily bounded above and below

$$2^k f(2^{k+1}) \leq \sum_{2^k \leq n < 2^{k+1}} f(n) \leq 2^k f(2^k)$$

at which point one easily sees that the original series  $\sum_{n=1}^{\infty} f(n)$  converges if and only if the condensed sum  $\sum_{k=0}^{\infty} 2^k f(2^k)$  converges. While both sums are infinite, in practice the latter sum is significantly more tractable than the former; for instance any polynomial improvements  $n^{-\varepsilon}$  to bounds for the original sequence  $f(n)$  leads to exponential improvements  $2^{-\varepsilon k}$  in the bounds for the new sequence  $2^k f(2^k)$ .

A surprisingly useful variant of this method was used repeatedly by Bourgain in many problems, in which dyadic decomposition is combined with the pigeonhole principle to locate a single “good” scale in which to run additional arguments. We refer to this combination of dyadic decomposition and the pigeonhole principle as *dyadic pigeonholing*.

The quantitative result claimed in Theorem ?? is already well suited to a dyadic pigeonholing argument, since the scales  $t_j$  in that argument are already at least dyadically separated, and we can sketch its proof as follows. Using the Fourier transform  $\hat{f}(\xi) := \int_{\mathbb{R}^d} f(x) e^{2\pi i x \cdot \xi} dx$ , one can rewrite the left-hand side of (??) as

$$\int_{\mathbb{R}^2} |\hat{1}_B(\xi)|^2 \hat{\sigma}(t_j \xi) d\xi,$$

where  $\hat{\sigma}(\xi) := \int_{S^1} e^{2\pi i \omega \cdot \xi} d\sigma(\omega)$  is the Fourier transform of the surface measure  $d\sigma$ . One can split this integral into the contribution of the “low frequencies”

$$\int_{|\xi| \leq \delta/t_j} |\hat{1}_B(\xi)|^2 \hat{\sigma}(t_j \xi) d\xi, \quad (4.1)$$

the “medium frequencies”

$$\int_{\delta/t_j \leq |\xi| \leq 1/\delta t_j} |\hat{1}_B(\xi)|^2 \hat{\sigma}(t_j \xi) d\xi, \quad (4.2)$$

and the “high frequencies”

$$\int_{|\xi| > 1/\delta t_j} |\hat{1}_B(\xi)|^2 \hat{\sigma}(t_j \xi) d\xi, \quad (4.3)$$

where  $0 < \delta = \delta(\varepsilon) \leq 1/2$  is a small quantity depending on  $\varepsilon$  to be chosen later. For the contribution (??) of the low frequencies, the factor  $\hat{\sigma}(t_j \xi)$  is close to 1, and it is not difficult to obtain a lower bound on this quantity that is  $\gtrsim |B|^2 \geq \varepsilon^2$  if  $\delta$  is small enough.

For the contribution (??) of the high frequencies, the factor  $\hat{\sigma}(t_j\xi)$  is quite small, and one can show that the contribution of this term is negligible, again for  $\delta$  small enough. The problematic term is the contribution (??), which one can of course upper bound by

$$\int_{\delta/t_j \leq |\xi| \leq 1/\delta t_j} |\hat{1}_B(\xi)|^2 d\xi.$$

By Plancherel's theorem one can upper bound this crudely by

$$\lesssim \int_{\mathbb{R}^2} |\hat{1}_B(\xi)|^2 d\xi = |B|,$$

but this bound is too weak compared to the lower bound of  $\gtrsim |B|^2$  one can obtain for the main term (??). However, note that because of the lacunarity hypothesis  $t_{j+1} \leq t_j/2$ , the annuli  $\{\delta/t_j \leq |\xi| \leq 1/\delta t_j\}$  only overlap with multiplicity  $O(\log \frac{1}{\delta})$ , hence the Plancherel bound actually gives

$$\sum_{j=1}^J \int_{\delta/t_j \leq |\xi| \leq 1/\delta t_j} |\hat{1}_B(\xi)|^2 d\xi \lesssim \log \frac{1}{\delta} |B| \quad (4.4)$$

and hence by the pigeonhole principle we can find a scale  $t_j$  for which

$$\sum_{j=1}^J \int_{\delta/t_j \leq |\xi| \leq 1/\delta t_j} |\hat{1}_B(\xi)|^2 d\xi \lesssim \frac{\log \frac{1}{\delta}}{J} |B|.$$

For  $J$  large enough, one can use this “good” scale to make the contribution (??) of the medium frequencies small compared to that of the high frequencies, and one can conclude the proof of Theorem ??.

Another typical instance of dyadic pigeonholing occurs in [?, Lemma 2.15] when Bourgain establishes new lower bounds on the Hausdorff dimension of Besicovitch sets  $E$  (compact subsets of  $\mathbb{R}^d$  that contain a unit line segment  $\ell_\omega$  in every direction). To lower bound the Hausdorff such a set by  $\alpha$ , one would have to establish a lower bound for the Hausdorff content  $\sum_i r_i^{\alpha-\varepsilon}$  whenever one covers the set  $E$  by small balls  $B(x_i, r_i)$ . By rounding up each  $r_i$ , we can assume without loss of generality that each  $r_i$  is a power of two (dyadic pigeonholing), and replace balls by cubes; we can then group together the cubes of a given size and obtain a covering

$$E \subset \sum_{j \geq j_0} B_j$$

where  $j_0$  is large and  $B_j$  is a union of cubes of sidelength  $2^{-j}$ . To get the required lower bound it would then suffice to show that for at least one of the scales  $j$ , the number of cubes used to form  $B_j$  is  $\gtrsim 2^{j(\alpha-\varepsilon)}$ . But which scale  $j$  to use? Since the  $B_j$  cover each  $\ell_\omega$ , we have

$$\sum_{j \geq j_0} \mathcal{H}^1(\ell_\omega \cap B_j) \geq 1$$

for each of the unit line segments  $\ell_\omega$ , hence on integrating over all directions  $\omega \in S^{d-1}$  using Fubini's theorem we have

$$\sum_{j \geq j_0} \int_{S^{d-1}} \mathcal{H}^1(\ell_\omega \cap B_j) d\omega \gtrsim 1.$$

By the pigeonhole principle, we can then find a scale  $j \geq j_0$  for which

$$\int_{S^{d-1}} \mathcal{H}^1(\ell_\omega \cap B_j) d\omega \gtrsim \frac{1}{j^2}$$

(say). The quantity  $1/j^2$  is quite “large” compared to the scale  $2^{-j}$ , and this estimate asserts (roughly speaking) that “many” of the  $\ell_\omega$  have “large” intersection with the  $B_j$ . Having selected such a good scale  $\delta = 2^{-j}$ , Bourgain was able to proceed to establish new bounds on  $\alpha$  by estimation of an expression now known as the *Keakeya maximal function* associated to this scale; see the companion paper [?] for further discussion of this function and its applications to the Fourier restriction problem.

The dyadic pigeonholing method does not need to explicitly involve powers of two (or other lacunary sequences of scales). One of Bourgain’s earlier uses of the method appears<sup>1</sup> in his work [?, §5] on quantitative versions of a Lipschitz embedding theorem of Ribe [?], where at one point in the argument he has a Lipschitz function  $F : E \rightarrow Y$  from a finite-dimensional normed space  $E$  to a Banach space  $Y$ , and wishes to locate a scale  $t$  at which the Poisson integral  $F * P_t$  of  $F$  has a large directional derivative  $\partial_a(F * P_t)(x)$  at one point  $x \in E$ . This scale  $t$  roughly corresponds to the spatial scale  $2^{-j}$  in the previous discussion. To locate a good scale  $t$ , Bourgain first observes from the triangle inequality that

$$\|\partial_a(F * P_t)\| * P_s(x) \geq \|\partial_a(F * P_{t+s})\|(x) \quad (4.5)$$

and hence the integral quantity  $\int_E \|\partial_a(F * P_t)\|(x) dx$  is non-increasing in  $t$ . On the other hand, the specific construction of the function  $F$  in [?] provided an upper bound on this integral that is uniform in  $t$ . Applying the pigeonhole principle, one can then find a scale  $t > 0$  for which one has

$$\int_E \|\partial_a(F * P_t)\|(x) dx \approx \int_E \|\partial_a(F * P_{(R+1)t})\|(x) dx$$

for some moderately large parameter  $R > 0$ , and where we shall be vague about the precise meaning of the symbol  $\approx$ ; comparing this with (??) and other properties of  $F$  eventually gives the desired lower bound on  $\partial_a(F * P_t)(x)$ . See the companion paper [?] for further discussion of the impact of Bourgain’s “Ribe program”.

Dyadic pigeonholing makes a small but important role in an important result [?] of Bourgain on the energy-critical nonlinear Schrödinger equation, discussed in more detail in Kenig’s article [?]:

**Theorem 4.1** (Global regularity for energy-critical NLS for radial data). *Let  $u_0 \in H^1(\mathbb{R}^3)$  be smooth and spherically symmetric, then there is a unique smooth finite energy global solution  $u : \mathbb{R} \times \mathbb{R}^3 \rightarrow \mathbb{C}$  to the nonlinear Schrödinger equation  $i\partial_t u + \Delta u = |u|^4 u$  with initial data  $u(0, x) = u_0(x)$ .*

At one stage [?, §4] in the (rather intricate) argument, a solution  $u$  is constructed to exhibit a concentration property at a certain time  $t_{j_s}$  and frequency  $N_{j_s}$ , in that (suppressing a parameter  $\eta$  that is not relevant for the current discussion)

$$\|P_{N_{j_s}} u(t_{j_s})\|_{L^2(\mathbb{R}^3)} \gtrsim N_{j_s}^{-1}, \quad (4.6)$$

---

<sup>1</sup>We thank Assaf Naor for this reference.

where  $P_{N_{j_s}}$  is a Fourier projection of Littlewood-Paley type to the frequency region  $\{\xi : |\xi| \sim N_{j_s}\}$ . In Bourgain's argument it is necessary to propagate this lower bound from time  $t_{j_s}$  to a nearby time  $t_{j_r}$ . The NLS equation conserves the full  $L^2$  mass  $\|u(t)\|_{L^2(\mathbb{R}^3)}^2$ , but this cannot be directly used here due to the solution  $u$  potentially having an extremely large amount of  $L^2$  mass at low frequencies. To resolve this, Bourgain applies a Fourier truncation operator  $P_{\geq N}$  restricting to frequencies  $|\xi| \gtrsim N$  for some  $N < N_{j_s}$ , and exploits the approximately conserved nature of the high-frequency portion  $\|P_{\geq N}u(t)\|_{L^2(\mathbb{R}^3)}^2$  of the mass, by a computation of the time derivative

$$\partial_t \|P_{\geq N}u(t)\|_{L^2(\mathbb{R}^3)}^2.$$

There are several components to this time derivative, but the dominant contribution comes from the portion of the solution  $u$  residing at frequency scales  $M$  comparable to  $N$  (intuitively, this reflects the potential exchange of mass in the frequency domain between frequency modes of magnitude just below  $N$ , and frequency modes of magnitude just above  $N$ ). For any given  $N$ , the upper bounds on this time derivative could overwhelm the lower bound in (??); however, by obtaining an estimate for the sum over a range of  $N$  (in a manner analogous to (??)) and applying the pigeonhole principle, one can locate a “good scale”  $N$  for which one has satisfactory control on the derivative of the mass. This idea to use the dyadic pigeonholing method to establish approximate conservation laws has wide application; for instance, it was used by Rodgers and myself recently to resolve the Newman conjecture [?] in analytic number theory.

In some cases one can use more sophisticated tools than the pigeonhole principle to locate a good scale. One example of this arises when establishing Bourgain's quantitative refinement [?] of Roth's theorem [?]:

**Theorem 4.2** (Bourgain-Roth theorem). *Let  $N \geq 10$ , and let  $A \subset \{1, \dots, N\}$  be a set containing no three-term arithmetic progressions. Then  $|A| \lesssim \frac{(\log \log N)^{1/2}}{\log^{1/2} N} N$ .*

A key innovation in this paper was to manipulate *Bohr sets*

$$B(S, \rho) := \{n \in \mathbb{Z} : |n| \leq N; \|n\theta\| \leq \rho \forall \theta \in S\}$$

where  $\rho > 0$  is a “radius”,  $S$  is a collection of frequencies  $\theta \in \mathbb{R}/\mathbb{Z}$ , and  $\|x\|$  denotes the distance of  $x$  to the nearest integer. These sets generalize the long arithmetic progressions that appear prominently in previous work in this area such as [?], and are well adapted to the Fourier-analytic methods used to establish Roth-type theorems. However, a key difficulty arises due to the discontinuous nature of the Bohr sets in  $\rho$ ; in particular, if  $\rho'$  is close to  $\rho$ , there is no *a priori* reason why the cardinality of  $B(S, \rho')$  should be close to that of  $B(S, \rho)$ . However, the cardinality  $|B(S, \rho)|$  is clearly non-decreasing in  $\rho$ , and a simple covering argument gives a doubling bound

$$|B(S, 2\rho)| \lesssim O(1)^{|S|} |B(S, \rho)|.$$

In particular, the distributional derivative  $\frac{d}{d\rho} \log |B(S, \rho)|$  is a measure of total variation  $O(|S|)$  on any dyadic interval  $[\rho_0, 2\rho_0]$ . Combining this with the Hardy-Littlewood maximal inequality, one can conclude that every interval  $[\rho_0, 2\rho_0]$  contains a radius  $\rho$



where the Bohr set  $B(S, \rho)$  is “regular” in the sense that

$$|B(S, \rho')| = \exp\left(O\left(|S| \frac{|\rho' - \rho|}{|\rho|}\right)\right) |B(S, \rho)|$$

for all  $\rho' > 0$ ; see for instance [?, Lemma 4.25] for this version of the construction. Regular Bohr sets are now a standard tool in modern additive combinatorics.

## 5. RANDOM TRANSLATIONS

Consider the interval  $I = [0, 1/N]$  in the unit circle  $\mathbb{R}/\mathbb{Z}$  for some large integer  $N$ . Then  $I$  is much smaller than  $\mathbb{R}/\mathbb{Z}$ , but we can cover  $\mathbb{R}/\mathbb{Z}$  by  $1/|I| = N$  translates  $I + j/N, j = 0, \dots, N-1$  of  $\mathbb{R}/\mathbb{Z}$ . Of course, most subsets  $E$  of  $\mathbb{R}/\mathbb{Z}$  will not have this perfect tiling property. However, by using random translations of  $E$ , one can achieve something fairly close to a perfect tiling:

**Lemma 5.1** (Random translations). *Let  $G = (G, \cdot)$  be a compact group (not necessarily abelian) with Haar probability measure  $\mu$ . Let  $E$  be a measurable subset of  $G$ , and let  $N$  be a natural number. Then there exist translates  $g_1 E, \dots, g_N E$  of  $E$  by some shifts  $g_1, \dots, g_N \in G$  with*

$$\mu(g_1 E \cup \dots \cup g_N E) \geq 1 - (1 - \mu(E))^N.$$

*Proof.* We use the probabilistic method. Let  $g_1, \dots, g_N$  be drawn independently at random from  $G$  using the Haar measure  $\mu$ . Then by the Fubini-Tonelli theorem we have

$$\begin{aligned} \mathbf{E}\mu(g_1 E \cup \dots \cup g_N E) &= \int_G \mathbf{E}1_{g_1 E \cup \dots \cup g_N E}(x) d\mu(x) \\ &= \int_G \mathbf{E}(1 - \prod_{i=1}^N 1_{g_i(G \setminus E)}(x)) d\mu(x) \\ &= \int_G (1 - \prod_{i=1}^N \mathbf{E}1_{g_i(G \setminus E)}(x)) d\mu(x) \\ &= \int_G (1 - \prod_{i=1}^N \mu(G \setminus E)) d\mu(x) \\ &= 1 - (1 - \mu(E))^N \end{aligned}$$

and the claim follows.  $\square$

In particular, if  $\mu(E) \sim 1/N$ , then we can find  $N$  translates  $g_1 E, \dots, g_N E$  of  $E$  whose union has measure  $\sim 1$ , thus these translates behave as if they are disjoint “up to constants”. We observe that the same claim also holds for any homogeneous space  $G/H$  of a compact group  $G$  (with the attendant Haar probability measure), simply by lifting subsets of that homogeneous space back up to  $G$ .

Lemma ?? allows one in many cases to reduce the analysis of “small” subsets of a compact group  $G$  (or a homogenous space  $G/H$  of  $G$ ) to the analysis of “large” sets,

particularly if the problem in question enjoys some sort of translation symmetry with respect to the group  $G$ . This idea was for instance famously exploited by Stein [?] in his maximum principle equating almost everywhere convergence results for translation-invariant operators with weak-type  $(p, p)$  maximal inequalities. In [?, §6], Bourgain noted that these techniques could also be combined with the factorization theory of Pisier, Nikishin, and Maurey [?] (which Bourgain had previously used for instance in [?]), although it has subsequently been realized that the arguments can be formulated without explicit reference to that theory. Specifically, in the context of restriction estimates for the sphere, Bourgain observed

**Proposition 5.2.** *Suppose that  $d \geq 2$  and  $1 < p < 2$  is such that one has the restriction estimate*

$$\|\hat{f}\|_{L^1(S^{d-1}, d\sigma)} \lesssim_{p,d} \|f\|_{L^p(\mathbb{R}^d)} \quad (5.1)$$

for all Schwartz functions  $f: \mathbb{R}^d \rightarrow \mathbb{C}$ , where  $\sigma$  is normalized surface measure on the sphere  $S^{d-1}$ . Then one can automatically improve this to the stronger estimate

$$\|\hat{f}\|_{L^{p,\infty}(S^{d-1}, d\sigma)} \lesssim_{p,d} \|f\|_{L^p(\mathbb{R}^d)}.$$

*Proof.* (Sketch) We can normalize  $\|f\|_{L^p(\mathbb{R}^d)} = 1$ . Let  $\lambda > 0$ , and let  $E \subset S^{d-1}$  denote the level set

$$E := \{\omega \in S^{d-1} : |\hat{f}(\omega)| \geq \lambda\}.$$

Our task is to show that  $\sigma(E) \lesssim_{p,d} \lambda^{-p}$ . A direct application of (??) only gives the estimate  $\sigma(E) \lesssim_{p,d} \lambda^{-1}$ , which is inferior when  $\lambda$  is large. However, if we let  $N$  be an integer with  $N \sim 1/\sigma(E)$ , then by Lemma ?? (applied to the homogeneous space  $S^{d-1} \equiv SO(d)/SO(d-1)$ ) one can find rotations  $R_1 E, \dots, R_N E$  of  $E$  with  $\sigma(\bigcup_{i=1}^N R_i E) \sim 1$ . If one then considers the random sum

$$F(x) := \sum_{i=1}^N \epsilon_i F(R_i x)$$

where the  $\epsilon_i$  are independent random signs  $\{-1, +1\}$  (or random gaussian variables), a routine application of Khintchine's inequality reveals that with positive probability, one has

$$\|F\|_{L^p(\mathbb{R}^d)} \lesssim_p N^{1/p}$$

and

$$\sigma(\{\omega \in S^{d-1} : |\hat{F}(\omega)| \gtrsim \lambda\}) \gtrsim \sigma(\bigcup_{i=1}^N R_i E) \gtrsim 1$$

and hence by (??)

$$\lambda \lesssim_{p,d} N^{1/p}$$

which gives the required estimate  $\sigma(E) \lesssim_{p,d} \lambda^{-p}$ .  $\square$

Variations of this argument also appear at several other locations in [?].

## 6. METRIC ENTROPY AND CONCENTRATION OF MEASURE

If  $X$  is a random variable (which for sake of discussion we take to be real-valued) with finite second moment, so that the mean  $\mathbf{E}X$  and the variance  $\mathbf{Var}(X)$  are both finite, then Chebyshev's inequality asserts that

$$\mathbf{P}(|X - \mathbf{E}X| \geq \lambda \sqrt{\mathbf{Var}(X)}) \leq \frac{1}{\lambda^2}$$

for any  $\lambda > 0$ , thus there the random variable  $X$  exhibits some *concentration of measure* to the interval  $[\mathbf{E}X - \sqrt{\mathbf{Var}(X)}, \mathbf{E}X + \sqrt{\mathbf{Var}(X)}]$ , in the sense that the probability of lying far outside this interval drops at a polynomial rate to the (normalized) distance to this interval). In many situations (particularly if  $X$  is somehow “influenced” by many “independent sources of randomness”), the decay is in fact far stronger than the  $1/\lambda^2$  decay; exponential or even Gaussian type decay can often be obtained. For instance, if  $X$  is a Gaussian variable, then we have

$$\mathbf{P}(|X - \mathbf{E}X| \geq \lambda \sqrt{\mathbf{Var}(X)}) \lesssim \exp(-c\lambda^2) \quad (6.1)$$

for all  $\lambda > 0$  and some absolute constant  $c > 0$ . Or, if  $X = \sum_{i=1}^n X_i$  is the sum of independent random variables  $X_i$  that each lie in some interval  $[a_i, b_i]$ , then the classical Hoeffding inequality gives

$$\mathbf{P}\left(|X - \mathbf{E}X| \geq \lambda \sqrt{\sum_{i=1}^n (b_i - a_i)^2}\right) \lesssim \exp(-c\lambda^2).$$

Many further *concentration of measure* inequalities of this type are available; see for instance the text [?] for a systematic discussion.

One can combine these sorts of concentration of measure inequalities to control the large deviations of suprema  $\sup_{t \in T} X_t$  of a random process  $(X_t)_{t \in T}$ , where  $t$  is a parameter ranging in some index set  $T$ . If for instance  $T$  is finite, then from the union bound we have

$$\mathbf{P}(\sup_{t \in T} X_t > \lambda) = \mathbf{P}(\bigvee_{t \in T} (X_t > \lambda)) \leq \sum_{t \in T} \mathbf{P}(X_t > \lambda)$$

for any  $\lambda > 0$ . For  $\lambda$  large, one can hope to use concentration of measure inequalities to obtain exponentially strong bounds on each individual probability  $\mathbf{P}(X_t > \lambda)$ ; if  $T$  is not too large (e.g., subexponential in cardinality) then this method can lead to non-trivial large deviation bounds on the supremum  $\sup_{t \in T} X_t$ . However, often in many cases of interest  $T$  is too large for this method to be directly used; for instance,  $t$  could be a continuous parameter, in which case  $T$  is likely to be uncountably infinite. But in many applications  $T$  has the structure of a totally bounded metric space  $(T, d)$ , thus for each scale  $\varepsilon > 0$  there exists some finite subset  $T_\varepsilon$  of  $T$  with the property that every element of  $T$  lies within  $\varepsilon$  of some element of  $T_\varepsilon$ . The minimal possible cardinality of  $T_\varepsilon$  is known as the *metric entropy* of  $T$  at scale  $\varepsilon$  and we will denote it by  $N(T, d; \varepsilon)$ . Applying these nets for each  $\varepsilon = 2^{-n}$ , one can assign to each element  $t \in T$  a chain  $t_0, t_1, \dots$  converging to  $t$  with  $t_n \in T_{2^{-n}}$  and  $d(t_n, t_{n+1}) \leq 2^{-n+1}$  for all  $n$ . If  $X_t$  depends continuously on  $t$ , the triangle inequality then gives the bound

$$X_t \leq |X_{t_0}| + \sum_{n=0}^{\infty} |X_{t_n} - X_{t_{n+1}}|$$

and hence

$$\sup_{t \in T} X_t \leq \sup_{t \in T_1} |X_t| + \sum_{n=0}^{\infty} \sup_{t \in T_{2^{-n}}; t' \in T_{2^{-n-1}}; d(t, t') \leq 2^{-n+1}} |X_t - X_{t'}|. \quad (6.2)$$

It is then possible to obtain good large deviation bounds on the uncountable supremum  $\sup_{t \in T} X_t$  by using the previous strategy to obtain large deviation bounds on the finite suprema  $\sup_{t \in T_1} |X_t|$  and  $\sup_{t \in T_{2^{-n}}; t' \in T_{2^{-n-1}}; d(t, t') \leq 2^{-n+1}} |X_t - X_{t'}|$ , which one then combines using crude tools such as the union bound. We refer to this as the *chaining argument*; it is particularly effective when one has good bounds on the metric entropies  $N(T, d; 2^{-n})$ . A prototype application of the chaining argument is *Dudley's inequality* [?]

$$\mathbf{E} \sup_{t \in T} X_t \lesssim \sum_{n \in \mathbb{Z}} 2^{-n} \sqrt{\log N(T, d; 2^{-n})}$$

whenever  $(X_t)_{t \in T}$  is a (mean zero) Gaussian process with  $T$  equipped with the metric (or more precisely, pseudo-metric)

$$d(t, t') := \sqrt{\mathbf{Var}(X_t - X_{t'})}.$$

This inequality can be readily proven by combining (??) with (??) and the union bound; see for instance [?, §2] for a clear treatment. However, the chaining argument is substantially more general and flexible than this, with many early applications of the chaining method due to Bourgain. Perhaps the most well known is the following result:

**Theorem 6.1** (Random sets of orthonormal systems have the  $\Lambda(p)$  property). *Let  $\phi_1, \dots, \phi_n$  be a system of bounded orthonormal functions on a probability space  $(X, \mu)$ , let  $2 < p < \infty$ , and let  $S \subset \{1, \dots, n\}$  be a random set with each  $i = 1, \dots, n$  lying in  $S$  with an independent probability of  $n^{2/p-1}$ . Then with probability  $\sim 1$ , one has the “ $\Lambda(p)$  inequality”*

$$\left\| \sum_{i \in S} a_i \phi_i \right\|_{L^p(X, \mu)} \lesssim_p \left( \sum_{i \in S} |a_i|^2 \right)^{1/2}$$

for all real or complex numbers  $a_i$ .

This theorem famously resolved a long-standing problem in harmonic analysis, namely whether it was possible to produce an (infinite) family of plane waves  $x \mapsto e^{2\pi i n x}$  on the unit circle  $\mathbb{R}/\mathbb{Z}$  which obeyed the  $\Lambda(p)$  inequality but not the  $\Lambda(q)$  inequality for a given choice of  $2 < p < q < \infty$ .

The proof of Theorem ?? is quite complicated and we only give an extremely oversimplified sketch here. The main difficulty here is one needs to control a random uncountable supremum

$$K := \sup_{|a| \leq 1} \left\| \sum_{i=1}^n 1_{i \in S} a_i \phi_i \right\|_{L^p(X, \mu)} \quad (6.3)$$

where  $a = (a_1, \dots, a_n)$  ranges over vectors of norm at most 1. Raising this expression to the power  $p$ , we obtain

$$\sup_{|a| \leq 1} \int_X \sum_{i=1}^n 1_{i \in S} a_i \phi_i \sum_{j=1}^n 1_{j \in S} \bar{a}_j \bar{\phi}_j \left| \sum_{k=1}^n 1_{k \in S} a_k \phi_k \right|^{p-2} d\mu. \quad (6.4)$$

The set  $S$  appears here three times, but by randomly decomposing  $S$  into three subsets  $S_1, S_2, S_3$ , and also judiciously splitting  $a$  into three components  $a, b, c$ , it turns out that

one can reduce the task of bounding this expression into that of bounding<sup>2</sup> “decoupled” analogues of (??) such as

$$\sup_{|a|,|b|,|c|\leq 1} \int_X \sum_{i=1}^n 1_{i\in S_1} a_i \phi_i \sum_{j=1}^n 1_{j\in S_2} \bar{b}_j \bar{\phi}_j \left| \sum_{k=1}^n 1_{k\in S_3} c_k \phi_k \right|^{p-2} d\mu.$$

In fact by using some further dyadic decompositions (in the spirit of Section ??) one can make further restrictions on the support and pointwise magnitudes of  $a, b, c$ ; a typical such restriction to keep in mind is that there is some  $1 \leq m \leq n$  for each of the  $a, b, c$  are supported on sets of cardinality at most  $m$  and are pointwise bounded by  $O(m^{-1/2})$ . One then applies a variant of Dudley’s inequality to control the supremum in  $a$ , reducing matters to controlling metric entropies of a collection of functions of the form

$$\left\{ \sum_{j=1}^n 1_{j\in S_2} b_j \phi_j \left| \sum_{k=1}^n 1_{k\in S_3} c_k \phi_k \right|^{p-2} : |b|, |c| \leq 1 \right\}$$

(with additional constraints on the support of  $b, c$  that we do not detail here). These are controlled in turn by elementary inequalities (such as Hölder’s inequality) as well as a variant of the random variable  $K$  defined in (??), as well as some metric entropy estimates of Bourgain, Lindenstrauss, and Milman [?].

The chaining method was later streamlined into the *generic chaining* method of Talagrand, in which the metric balls  $\{t' : d(t, t') \leq \varepsilon\}$  that implicitly appear in the chaining argument are allowed to be weighted by an arbitrary measure on  $T$  known as a *majorizing measure*, leading to estimates that are essentially optimal in many situations, and can be used for instance to give a simplified proof of Theorem ?? with a stronger conclusion, and which avoids the use of decoupling methods; see [?] for details.

There are many other works of Bourgain (and coauthors) in which metric entropy and chaining arguments are used to bound large deviations of supremum type quantities. Here is a small sample:

- The paper [?] concerns the approximation theory of zonotopes (finite sums of intervals in a normed vector space), showing that all zonoids (limits of zonotopes) can be efficiently approximated by “low complexity” zonotopes. A key step is to use a chaining argument to show that an  $L^1$  norm  $\|x\|_{L^1(X, \mu)}$  can be efficiently approximated by an empirical sample  $\frac{1}{N_0} \sum_{i=1}^{N_0} |x(i)|$  uniformly for all  $x$  in a certain convex body, relying heavily on metric entropy estimates on convex bodies such as the dual Sudakov inequality [?].
- In [?], a metric entropy and concentration of measure argument is used to show that Montgomery’s large values conjecture [?] in analytic number theory on the distribution of large values of a Dirichlet polynomial  $\sum_{n \sim M} a_n n^{it} = \sum_{n \sim M} a_n e^{it \log n}$  is almost surely true if one replaces the frequencies  $\log n$  by a suitable random set.

---

<sup>2</sup>This sort of trick to decouple probabilistic expressions of dependent random variables into probabilistic expressions of independent random variables is another useful member of Bourgain’s toolkit that is now widely used in probability theory. We will not discuss this decoupling trick in further detail here, but see for instance [?].

- In [?], a metric entropy and concentration of measure argument (as well as the decoupling trick) is used to show if  $K$  is a symmetric convex body of unit volume whose moment of inertia  $\int_K yy^T dy$  is normalised to be a constant multiple  $LI$  of the identity matrix, then this moment of inertia can be closely approximated by that of a surprisingly small number of randomly chosen points from  $K$ ; this has applications to random matrix theory and statistics, in particular in allowing one to compare a covariance matrix with an empirical sample of that matrix [?].
- In [?], a chaining argument combined with dyadic decomposition is used to show that a randomly selected collection  $S$  of columns in a bounded orthonormal system obeys the “restricted isometry property”, of being an approximate isometry when restricted to any  $m$  rows, as long as  $S$  is only slightly larger than  $m$ , improving quantitatively over previous bounds [?, ?] in the area.

## 7. PUTTING IT ALL TOGETHER: A CASE STUDY

Let us say that a subset  $S$  of the reals has *property (E)* if, whenever every measurable subset  $A$  of  $\mathbb{R}$  of positive measure contains an affine image  $x + tS := \{x + ty : y \in S\}$  of  $S$  for some  $x \in \mathbb{R}$  and  $t \neq 0$ . An easy application of the Lebesgue density theorem shows that every finite set of reals has property (E). The following question of Erdős is still unsolved:

**Problem 7.1** (Erdős similarity problem). *Does there exist an infinite set  $S$  with property (E)?*

One of the strongest general negative results in this direction is by Bourgain [?]:

**Theorem 7.2** (Triple sumsets fail (E), qualitative version). *Let  $S_1, S_2, S_3$  be infinite subsets of  $\mathbb{R}$ . Then the sumset  $S_1 + S_2 + S_3 := \{s_1 + s_2 + s_3 : s_1 \in S_1, s_2 \in S_2, s_3 \in S_3\}$  fails property (E).*

The corresponding question for double sumsets  $S_1 + S_2$  remains open; it will be clear shortly why it is necessary in Bourgain’s arguments to have at least three summands.

As we shall see, the proof of this result can largely be described as an application of several of the tools discussed in this paper. The first step is to convert the problem to a quantitative one, as per Section ?? . The quantitative formulation is as follows:

**Theorem 7.3** (Triple sumsets fail (E), quantitative version). *Let  $S_1, S_2, S_3$  be bounded infinite subsets of  $\mathbb{R}$  containing 0 as an adherent point. Then there does not exist a constant  $C$  for which one has the bound*

$$\int_{(\mathbb{R}/\mathbb{Z})^J} \inf_{1 < t < 2} \sup_{x' \in x + tS_0 v} |f(x')| dx \leq C \int_{(\mathbb{R}/\mathbb{Z})^J} |f(x)| dx$$

*for all tori  $(\mathbb{R}/\mathbb{Z})^J$ , all continuous  $f : (\mathbb{R}/\mathbb{Z})^J \rightarrow \mathbb{R}$ , all vectors  $v \in \mathbb{R}^J$ , and all finite subsets  $S_0$  of  $S_1 + S_2 + S_3$ .*

Let us now sketch why Theorem ?? implies Theorem ??. (The converse implication is also true; see [?, §2].) Since unbounded sets clearly fail property (E), we may assume without loss of generality in Theorem ?? that  $S_1, S_2, S_3$  are bounded; by Bolzano-Weierstrass and translation, we may assume that each of the  $S_i$  contain 0 as an adherent point.

Let  $\delta > 0$  be sufficiently small, and let  $M > 0$ . From Theorem ?? and rescaling, one can find a torus  $(\mathbb{R}/\mathbb{Z})^J$ , a vector  $v \in \mathbb{R}^J$ , a finite subset  $S_0$  of  $S_1 + S_2 + S_3$  and a continuous function  $f$  (which we can take to be non-negative) such that

$$\int_{(\mathbb{R}/\mathbb{Z})^J} F(x) dx > M \int_{(\mathbb{R}/\mathbb{Z})^J} f(x) dx$$

where  $F(x) := \inf_{\delta < t < 2\delta} \sup_{x' \in x + tS_0 v} f(x')$ . Applying dyadic pigeonholing (and writing  $\int_{(\mathbb{R}/\mathbb{Z})^J} F dx = \int_0^\infty |\{F \geq \lambda\}| d\lambda$  and  $\int_{(\mathbb{R}/\mathbb{Z})^J} f dx = \int_0^\infty |\{f > \lambda\}| d\lambda$ ), as per Section ??, we can then find a threshold  $\lambda > 0$  such that

$$|\{F \geq \lambda\}| > M |\{f > \lambda\}|.$$

If we write  $A := \{f > \lambda\}$  and

$$A_1 := \bigcap_{\delta < t < 2\delta} \bigcup_{y \in S_0} (A - tyv)$$

then we have  $|A_1| \geq M|A|$ . The set  $A_1$  could be small compared with  $(\mathbb{R}/\mathbb{Z})^J$ ; however by using the random translations trick as per Section ??, we can find an open set  $B \subset (\mathbb{R}/\mathbb{Z})^J$  (a union of finitely many translates of  $A$ ) of arbitrarily small measure such that the set

$$B_1 := \bigcap_{\delta < t < 2\delta} \bigcup_{y \in S_0} (B - tyv)$$

has measure arbitrarily close to 1. By construction, the set  $B_1 \setminus B$  does not contain any set of the form  $x + t(S_1 + S_2 + S_3)v$  with  $x \in \mathbb{R}/\mathbb{Z}$  and  $\delta < t < 2\delta$  (here we use the fact that  $B$  is open and 0 is an adherent point of  $S_1 + S_2 + S_3$ ). If one then considers the set  $\{y \in [0, 1] : x + tyv \in B_1 \setminus B\}$  for a randomly chosen  $x \in (\mathbb{R}/\mathbb{Z})^J$ , one can then find a subset of  $[0, 1]$  of measure arbitrarily close to 1 that does not contain any set of the form  $x + tS$  with  $x \in \mathbb{R}$  and  $\delta < t < 2\delta$ . Taking intersections over all small dyadic choices of  $\delta$  (and then restricting to a small interval to eliminate large scales) we can then establish Theorem ??.

Now we sketch the proof of Theorem ??. We have to find a continuous function  $f : (\mathbb{R}/\mathbb{Z})^J \rightarrow \mathbb{R}$  and a vector  $v \in \mathbb{R}^J$  for which

$$\int_{(\mathbb{R}/\mathbb{Z})^J} \inf_{1 < t < 2} \sup_{x' \in x + tS_0 v} |f(x')| dx \gg \int_{(\mathbb{R}/\mathbb{Z})^J} |f(x)| dx$$

for some finite  $S_0 \subset S_1 + S_2 + S_3$ , where we informally use  $X \gg Y$  to denote the claim that  $X$  is much larger than  $Y$ . The next idea is to take advantage of large deviations as per Section ??. To do this one needs to select the vector  $v$  so that the collection of dilates  $tS_0 v \bmod \mathbb{Z}^J$ ,  $1 \leq t \leq 2$  have “low entropy”. The construction is as follows. Let  $J := 3J_0$  for a large  $J_0$ , then as  $S_1, S_2, S_3$  all have 0 as an adherent point we can find non-zero real numbers  $s_{i,j} \in S_i$  for  $i = 1, 2, 3$  and  $j = 1, \dots, J_0$  with the relative size relation

$$|s_{1,1}| \gg \dots \gg |s_{1,J_0}| \gg |s_{2,1}| \gg \dots \gg |s_{2,J_0}| \gg |s_{3,1}| \gg \dots \gg |s_{3,J_0}| > 0.$$

We then let  $v \in \mathbb{R}^J$  be the vector

$$v := \left( \frac{1}{10s_{i,j}} \right)_{i=1,2,3; j=1,\dots,J_0}.$$

If we set  $S_0 := \{s_{1,j_1} + s_{2,j_2} + s_{3,j_3} : 1 \leq j_1, j_2, j_3 \leq J_0\}$ , then  $S_0$  is a finite subset of  $S_1 + S_2 + S_3$ . A routine calculation shows that for any  $1 \leq t \leq 2$ , the set  $tS_0v \bmod \mathbb{Z}^J$  consists of  $J_0^3$  points separated from each other (in the  $\ell^\infty$  metric on  $(\mathbb{R}/\mathbb{Z})^J$ ) by  $\gtrsim 1$ . If this set had no further structure, one would then expect the “entropy” of such sets to be exponential in  $J \times J_0^3 \sim J^4$ . However the arithmetic structure gives this set significantly lower entropy. Indeed, observe that each of the  $J$  coordinates of each of the  $J_0^3$  elements of  $tS_0v \bmod \mathbb{Z}^J$  are sums of three quantities of the form  $\frac{s_{i',j'}}{10s_{i,j}} \bmod 1$  for  $i, i' = 1, \dots, 3$ ,  $j, j' = 1, \dots, J_0$ . Thus the set  $tS_0v \bmod \mathbb{Z}^J$  is completely described by  $O(J^2)$  parameters, so the metric entropy of these sets (using the Hausdorff metric and the  $\ell^\infty$  norm on  $(\mathbb{R}/\mathbb{Z})^J$ ) at any given metric scale  $0 < \delta < 1$  is  $O(1/\delta)^{O(J^2)}$ . In particular, this entropy is subexponential compared to the cardinality  $J_0^3$  of  $S_0$ . (It is at this point that it is essential that we have at least three summands in the set  $S_1 + S_2 + S_3$ .)

Now we sketch how to construct the function  $f$ . Let  $\varepsilon > 0$ , and suppose  $J$  is large depending on  $\varepsilon$ . Partition  $(\mathbb{R}/\mathbb{Z})^J$  into cubes of sidelength (say)  $\frac{1}{100J}$ , and let  $E$  be the union of a random collection of these cubes, with each cube selected in  $E$  with an independent probability of  $\varepsilon$ . We set  $f$  to be the indicator function  $1_E$  (we ignore for this sketch the requirement that  $f$  be continuous, as this can be addressed by a standard mollification). Then  $\int_{(\mathbb{R}/\mathbb{Z})^J} |f(x)| dx \sim \varepsilon$  with high probability. On the other hand, for any given  $x \in (\mathbb{R}/\mathbb{Z})^J$  and  $1 < t < 2$ , we will have  $\sup_{x' \in x + tS_0v} |f(x')| = 1$  with probability at least  $1 - \exp(-\varepsilon J_0^3)$ . Using the metric entropy bound, one can then show that  $\inf_{1 < t < 2} \sup_{x' \in x + tS_0v} |f(x')| = 1$  with probability  $1 - O(O(J)^{O(J^2)} \exp(-\varepsilon J_0^3))$ , which is comparable to 1 if  $J$  is large enough. This gives the claim.

## REFERENCES

- [1] K. Ball, *The Legacy of Jean Bourgain in Geometric Functional Analysis*, preprint.
- [2] J. Bourgain, *A Szemerédi type theorem for sets of positive density in  $\mathbb{R}^k$* , Israel J. Math. **54** (1986), no. 3, 307–316.
- [3] J. Bourgain, *Remarks on the extension of Lipschitz maps defined on discrete sets and uniform homeomorphisms. Geometrical aspects of functional analysis (1985/86)*, 157–167, Lecture Notes in Math., 1267, Springer, Berlin, 1987.
- [4] J. Bourgain, *Construction of sets of positive measure not containing an affine image of a given infinite structures*, Israel J. Math. **60** (1987), no. 3, 333–344.
- [5] J. Bourgain, *Pointwise ergodic theorems for arithmetic sets*, With an appendix by the author, Harry Furstenberg, Yitzhak Katznelson and Donald S. Ornstein. Inst. Hautes Études Sci. Publ. Math. No. 69 (1989), 5–45.
- [6] J. Bourgain, *Besicovitch type maximal operators and applications to Fourier analysis*, Geom. Funct. Anal. **1** (1991), no. 2, 147–187.
- [7] J. Bourgain, *On the distribution of Dirichlet sums*, J. Anal. Math. **60** (1993), 21–32.
- [8] J. Bourgain, *Remarks on Halasz-Montgomery type inequalities*, Geometric aspects of functional analysis (Israel, 1992–1994), 25–39, Oper. Theory Adv. Appl., 77, Birkhäuser, Basel, 1995.
- [9] J. Bourgain, *Global wellposedness of defocusing critical nonlinear Schrödinger equation in the radial case*, J. Amer. Math. Soc. **12** (1999), no. 1, 145–171.



- [10] J. Bourgain, *Random points in isotropic convex sets*, Convex geometric analysis (Berkeley, CA, 1996), 53–58, Math. Sci. Res. Inst. Publ., 34, Cambridge Univ. Press, Cambridge, 1999.
- [11] J. Bourgain, *On triples in arithmetic progression*, Geom. Funct. Anal. **9** (1999), no. 5, 968–984.
- [12] J. Bourgain, *An improved estimate in the restricted isometry problem*, Geometric aspects of functional analysis, 65–70, Lecture Notes in Math., 2116, Springer, Cham, 2014.
- [13] J. Bourgain, J. Lindenstrauss, V. Milman, *Approximation of zonoids by zonotopes*, Acta Math. 162 (1989), no. 1-2, 73–141.
- [14] J. Bourgain, L. Tzafriri, *Invertibility of “large” submatrices with applications to the geometry of Banach spaces and harmonic analysis*, Israel J. Math. 57 (1987), no. 2, 137–224.
- [15] J. Bourgain, L. Tzafriri, *On a problem of Kadison and Singer*, J. Reine Angew. Math. 420 (1991), 1–43.
- [16] E. Candès, T. Tao, *Near-optimal signal recovery from random projections: universal encoding strategies?*, IEEE Trans. Inform. Theory **52** (2006), no. 12, 5406–5425.
- [17] A. Cauchy, *Anciens Exercices*, vol. 2 (1827), p. 221.
- [18] V. de la Peña, E. Giné, *Decoupling, From dependence to independence. Randomly stopped processes. U-statistics and processes. Martingales and beyond. Probability and its Applications* (New York). Springer-Verlag, New York, 1999.
- [19] C. Demeter, *Bourgain’s work in Fourier restriction*, preprint.
- [20] R. Dudley, *The sizes of compact subsets of Hilbert space and continuity of Gaussian processes*, Journal of Functional Analysis **1** (1967), 290–330.
- [21] H. Furstenberg, Y. Katznelson, B. Weiss, *Ergodic theory and configurations in sets of positive density*, Mathematics of Ramsey theory, 184–198, Algorithms Combin., 5, Springer, Berlin, 1990.
- [22] C. Kenig, *On the work of Jean Bourgain in nonlinear dispersive equations*, preprint.
- [23] M. Ledoux, *The concentration of measure phenomenon*. Mathematical Surveys and Monographs, 89. American Mathematical Society, Providence, RI, 2001.
- [24] H. Montgomery, *Topics in multiplicative number theory*. Lecture Notes in Mathematics, Vol. 227. Springer-Verlag, Berlin-New York, 1971.
- [25] A. Pajor, N. Tomczak-Jaegermann, *Subspaces of small codimension of finite dimensional Banach spaces*, Proc. Amer. Math. Soc. **97** (1986), 634–642.
- [26] G. Pisier, *Factorization of operators through  $L^{p,\infty}$  or  $L^{p,1}$  and noncommutative generalizations*, Math. Ann. 276 (1986), no. 1, 105–136.
- [27] M. Ribe, *On uniformly homeomorphic normed spaces*, Ark. Mat. **14** (1976), no. 2, 237–244.
- [28] B. Rodgers, T. Tao, *The de Bruijn-Newman constant is non-negative*, Forum Math. Pi **8** (2020), e6, 62 pp.
- [29] G-C. Rota, *Ten lessons I wish I had been taught*, Notices Amer. Math. Soc. 44 (1997), no. 1, 22–25.
- [30] E. M. Stein, *On limits of sequences of operators*, Ann. of Math. (2) **74** (1961), 140–170.
- [31] K. F. Roth, *On certain sets of integers*, J. London Math. Soc. **28** (1953), 104–109.
- [32] M. Rudelson, R. Vershynin, *On sparse reconstruction from Fourier and Gaussian measurements*, Comm. Pure Appl. Math. **61** (2008), no. 8, 1025–1045.
- [33] M. Talagrand, *Majorizing measures: the generic chaining*, Ann. Probab. 24 (1996), no. 3, 1049–1103.
- [34] M. Talagrand, *The generic chaining. Upper and lower bounds of stochastic processes*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2005.
- [35] T. Tao, V. Vu, *Additive combinatorics*. Cambridge Studies in Advanced Mathematics, 105. Cambridge University Press, Cambridge, 2006.
- [36] R. Vershynin, *How close is the sample covariance matrix to the actual covariance matrix?*, J. Theoret. Probab. **25** (2012), no. 3, 655–686.

UCLA DEPARTMENT OF MATHEMATICS, LOS ANGELES, CA 90095-1555.

Email address: tao@math.ucla.edu