Analysis and Its Applications

Pan Jing Bin

Supervisor: Assistant Professor Subhroshekhar Ghosh

Department of Mathematics

National University of Singapore

Contents

1 Introduction

1.1 Introduction

Orginating out of the study of Fourier series for differential equations, harmonic analysis have since grown into an immense field that encompasses a large part of modern mathematics. The far-reaching impact of harmonic analysis on other branches of mathematics is not only in its theorems and results, but also the tools and techniques developed in the process. In this report, we will explore some of the techniques employed in the study of harmonic analysis through the lens of the Fields medallist Terrance Tao.

In chapter 2, we will provide a general introduction to the princple of amplification and arbitrage, a technique first popularised by Terrance Tao in his online mathematical blog. In chapter 3, we will focus our attention on a particularly special type of amplification technique known as the tensor power trick. In chapter 4, we will switch our attention to one of Tao's papers "Exploring The Toolkit of Jean Bourgain", in which he provides an in-depth discussion on the techniques employed by Jean Bourgain, another renowned harmonic analyst, in his work.

1.2 Notation

For a set E, we use $\mathbb{1}_E$ to denote the indicator function of E. We use $|E|$ to denote either the cardinality of E if E is finite, or the Lebesgue measure of E if E is infinite.

The notation $X \lesssim Y$ denotes $|X| \leq CY$ for some constant C. If this constant depend on parameters, we will use $X \lesssim_{p,d} Y$ to denote $|X| \leq C_{p,d}Y$ for a constant C depending only on p and d. The notation $X \sim Y$ denotes both $X \lesssim Y$ and $Y \lesssim X$. The dependence on parameters is similarly denoted by $X \sim_{p,d} Y$.

2 Amplification and Arbitrage

Inequalities play a fundamental role in many branches of mathematics. This is especially so in the case of harmonic analysis, and a significant part of research in the field is dedicated to the development of tools and techniques to derive lower and upper bounds. The amplification and arbitrage principle is one such family of techniques.

Generally speaking, these techniques are applied to transform weak estimates (usually established as a first attempt to solve a problem) into much stronger ones by leveraging on the various kinds of "symmetries" of the underlying space. In each section, we will discuss the different types of symmetry that can be leveraged on.

2.1 Amplification via phase and dilation

Example 2.1.1. One of the most well-known examples of this technique comes from the proof of the Cauchy-Schwartz inequality

$$|\langle \boldsymbol{v}, \boldsymbol{w}\rangle| \leq \|\boldsymbol{v}\| \, \|\boldsymbol{w}\| \tag{1}$$

for vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ in a complex Hilbert space. Assume that $\boldsymbol{v}$ and $\boldsymbol{w}$ are non-zero (otherwise the inequality is trivial). We first start with the obvious inequality

$$\|\boldsymbol{v} - \boldsymbol{w}\|^2 \geq 0$$

and then expand to obtain

$$\mathrm{Re}\big[\langle \boldsymbol{v}, \boldsymbol{w}\rangle\big] \leq \frac{1}{2}\|\boldsymbol{v}\|^2 + \frac{1}{2}\|\boldsymbol{w}\|^2 .$$

The above inequality is weaker than (1) since the left hand side is smaller and the right hand side is bigger (by AM-GM inequality). However, the right hand side is invariant under a phase change of $e^{i\theta}$. By replacing $\boldsymbol{v}$ with $e^{i\theta}\boldsymbol{v}$, we get

$$\mathrm{Re}\Big[e^{i\theta}\langle \boldsymbol{v}, \boldsymbol{w}\rangle\Big] \leq \frac{1}{2}\|\boldsymbol{v}\|^2 + \frac{1}{2}\|\boldsymbol{w}\|^2 .$$

The imbalance in symmetry means that we now have the freedom in choosing θ. In this case, we choose θ to cancel the phase of $\langle \boldsymbol{v}, \boldsymbol{w}\rangle$ and obtain

$$|\langle \boldsymbol{v}, \boldsymbol{w}\rangle| \leq \frac{1}{2}\|\boldsymbol{v}\|^2 + \frac{1}{2}\|\boldsymbol{w}\|^2 .$$

This inequality is closer to what we want since we have fixed the left hand side. To fix the right hand side, we exploit a different symmetry, namely the homogenisation symmetry $(\boldsymbol{v}, \boldsymbol{w}) \mapsto (\lambda\boldsymbol{v}, \frac{1}{\lambda}\boldsymbol{w})$. We then obtain

$$|\langle \boldsymbol{v}, \boldsymbol{w}\rangle| \leq \frac{\lambda^2}{2}\|\boldsymbol{v}\|^2 + \frac{1}{2\lambda^2}\|\boldsymbol{w}\|^2 .$$

By an elementary calculus argument (or again by the AM-GM inequality), the right hand side is minimised when $\lambda = \sqrt{\|\boldsymbol{w}\| \,/\, \|\boldsymbol{v}\|}$. Thus we get

$$|\langle \boldsymbol{v}, \boldsymbol{w} \rangle| \leq \|\boldsymbol{v}\| \, \|\boldsymbol{w}\|$$

as desired.

2.2 Amplification via linearity

Another powerful type of symmetry that is frequently present is linearity. In the next example, we will illustrate how linearity can be used to obtain depolarised inequalities from polarised ones.

Example 2.2.1. Let X be a normed vector space over $\mathbb{R}$ and let $B(-,-) : X \times X \to \mathbb{R}$ be a symmetric bilinear form. Suppose that one has already proven the polarised inequality

$$|B(f,f)| \leq A \|f\|_X^2$$

for all $f \in X$, where A is an absolute constant. We can amplify this by replacing f with $f + \lambda g$ for arbitrary (non-zero) $f, g \in X$ and $\lambda \in \mathbb{R}$. By the triangle inequality, we obtain

$$|B(f,f) + 2\lambda B(f,g) + \lambda^2 B(g,g)| \leq A(\|f\|_X + |\lambda| \|g\|_X)^2.$$

We optimise in λ by choosing $\lambda = \|f\|_X / \|g\|_X$. Observe that

$$\begin{aligned} 2\lambda |B(f,g)| &\leq A(2\|f\|_X)^2 + |B(f,f)| + \lambda^2 |B(g,g)| \\ &\leq 4A\|f\|_X^2 + A\|f\|_X^2 + \lambda^2 A\|g\|_X^2 \\ &= 6A\|f\|_X^2 . \end{aligned}$$

Dividing by λ throughout gives the depolarised inequality

$$|B(f,g)| \leq 3A\|f\|_X \|g\|_X$$

as desired.

2.3 Amplification via translational invariance

Other than deriving lower and upper bounds, the amplification technique can also be used to rule out certain types of estimates that are not compatible with the symmetries of the underlying space.

Definition 2.3.1. Let $1 \le p, q \le \infty$. For any $\boldsymbol{c} \in \mathbb{R}^n$, define the translation operator $\tau_{\boldsymbol{c}}$ by $(\tau_{\boldsymbol{c}} \circ f)(\boldsymbol{x}) := f(\boldsymbol{x} + \boldsymbol{c})$. A linear operator $T : L^p(\mathbb{R}^n) \to L^q(\mathbb{R}^n)$ is **translation invariant** if

$$T(\tau_{\boldsymbol{c}} \circ f) = \tau_{\boldsymbol{c}} \circ T(f) \quad \text{for all } f \in L^p(\mathbb{R}^n).$$

Example 2.3.2. Let $1 \le p, q \le \infty$ and let $T : L^p(\mathbb{R}^n) \to L^q(\mathbb{R}^n)$ be a non-zero translation invariant operator. We show that any estimate of the form

$$\|(1 + |\boldsymbol{x}|)^\alpha T(f)\|_{L^q(\mathbb{R}^n)} \lesssim_{p,q,\alpha,\beta,n} \left\|(1 + |\boldsymbol{x}|)^\beta f\right\|_{L^p(\mathbb{R}^n)} \tag{2}$$

must satisfy $\alpha \le \beta$.

Suppose that the estimate (2) holds for some parameters p, q, α, β, n. Let $\boldsymbol{x_0} \in \mathbb{R}^n$. Then

$$\left[\int_{\mathbb{R}^n} \left|(1 + |\boldsymbol{x}|)^\alpha T(\tau_{\boldsymbol{x_0}} \circ f)(\boldsymbol{x})\right|^q \, d\mu(\boldsymbol{x})\right]^{\frac{1}{q}} \lesssim_{p,q,\alpha,\beta,n} \left[\int_{\mathbb{R}^n} \left|(1 + |\boldsymbol{x}|)^\beta (\tau_{\boldsymbol{x_0}} \circ f)(\boldsymbol{x})\right|^p \, d\mu(\boldsymbol{x})\right]^{\frac{1}{p}}.$$

By translational invariance of T, we get

$$\left[\int_{\mathbb{R}^n} \left|(1 + |\boldsymbol{x}|)^\alpha T(f)(\boldsymbol{x} + \boldsymbol{x_0})\right|^q \, d\mu(\boldsymbol{x})\right]^{\frac{1}{q}} \lesssim_{p,q,\alpha,\beta,n} \left[\int_{\mathbb{R}^n} \left|(1 + |\boldsymbol{x}|)^\beta f(\boldsymbol{x} + \boldsymbol{x_0})\right|^p \, d\mu(\boldsymbol{x})\right]^{\frac{1}{p}}.$$

After a linear change of variables $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{x_0}$, we get

$$\left[\int_{\mathbb{R}^n} \left|(1 + |\boldsymbol{y} - \boldsymbol{x_0}|)^\alpha T(f)(\boldsymbol{y})\right|^q \, d\mu(\boldsymbol{y})\right]^{\frac{1}{q}} \lesssim_{p,q,\alpha,\beta,n} \left[\int_{\mathbb{R}^n} \left|(1 + |\boldsymbol{y} - \boldsymbol{x_0}|)^\beta f(\boldsymbol{y})\right|^p \, d\mu(\boldsymbol{y})\right]^{\frac{1}{p}}.$$

Hence we have obtained the estimate

$$\|(1 + |\boldsymbol{x} - \boldsymbol{x_0}|)^\alpha T(f)\|_{L^q(\mathbb{R}^n)} \lesssim_{p,q,\alpha,\beta,n} \left\|(1 + |\boldsymbol{x} - \boldsymbol{x_0}|)^\beta f\right\|_{L^p(\mathbb{R}^n)}$$

that is uniform in $\boldsymbol{x_0} \in \mathbb{R}^n$. Now observe that for a fixed function f, as $|\boldsymbol{x_0}| \to \infty$, the left hand side has asymptotic growth $\Theta(|\boldsymbol{x_0}|^\alpha)$ while the right hand side has asymptotic growth $\Theta(|\boldsymbol{x_0}|^\beta)$. This is only possible if $\alpha \le \beta$.

2.4 The general amplification principle

The techniques discussed in the previous sections can be formalised as follows. Consider an inequality of the form

$$A(f) \leq B(f) \tag{3}$$

which holds for all f in some function space X. Here, A and B are functionals of f. Suppose that we have some group G of symmetries $T : X \to X$ that acts on the underlying space X. For example, in the previous section, $G = \{\tau_{\boldsymbol{c}} \ : \ \boldsymbol{c} \in \mathbb{R}^n\}$ is the group of translational operators. We say that A is ***G*-invariant** if

$$A(T(f)) = A(f) \quad \text{ for all } T \in G.$$

Suppose that there is an imbalance of symmetry in the sense that either A is G-invariant and B is not, or vice versa. In the former case, we can amplify (3) to the stronger inequality

$$A(f) \leq \inf_{T \in G} B(T(f))$$

and in the latter case, we can amplify (3) to

$$\sup_{T \in G} A(T(f)) \leq B(f).$$

This gives rise to the following general principle: the most efficient inequalities are those in which the left-hand side and right-hand side enjoy the same kind of symmetries.

A method that is commonly used to prove inequalities in analysis is to chain together several simpler inequalities. Returning back to (3), one may attempt to prove the inequality by chaining two simpler inequalities together:

$$A(f) \leq C(f) \leq B(f).$$

The amplification principle also apply to these situations in the following sense: whenever possible, the symmetries that are present on both sides of the inequality should be maintained. To see why this follows from the general amplification principle, suppose that both A and B are G-invariant but C is not. Then we can amplify both the left-half and right-half to obtain

$$A(f) \leq \inf_{T \in G} C(T(f)) \quad \text{and} \quad \sup_{T \in G} C(T(f)) \leq B(f).$$

In particular, if all the quantities involved are positive (which is often the case in analysis), then we get

$$\frac{\sup_{T \in G} C(T(f))}{\inf_{T \in G} C(T(f))} \leq \frac{B(f)}{A(f)}. \tag{4}$$

In some situations, (4) is a non-trivial strengthening of (3).

Example 2.4.1. Suppose one has proven the triangle inequality

$$|z+w| \leq |z| + |w| \tag{5}$$

for all $z, w \in \mathbb{R}$ and wish to generalise it to $z, w \in \mathbb{C}$. One may first consider using the crude estimate

$$|z+w| = \sqrt{\mathrm{Re}(z+w)^2 + \mathrm{Im}(z+w)^2} \leq |\mathrm{Re}(z+w)| + |\mathrm{Im}(z+w)| \tag{6}$$

and then apply the triangle inequality

$$|\mathrm{Re}(z+w)| \leq |\mathrm{Re}(z)| + |\mathrm{Re}(w)| \quad \text{and} \quad |\mathrm{Im}(z+w)| \leq |\mathrm{Im}(z)| + |\mathrm{Im}(w)|.$$

to both the real and imaginary parts. Using another crude estimate

$$|\mathrm{Re}(z)|, |\mathrm{Im}(z)| \leq |z| \quad \text{and} \quad |\mathrm{Re}(w)|, |\mathrm{Im}(w)| \leq |w| \tag{7}$$

and chaining all the inequalities together gives

$$|z+w| \leq 2|z| + 2|w|.$$

The reason for the "loss" is because while the original inequality (5) is invariant under a phase change $(z, w) \mapsto (e^{i\theta}z, e^{i\theta}w)$, the intermediate expressions (6) and (7) are not.

A complementary approach to preserving symmetry is to "spend" the symmetry at a convenient location. Returning yet again to (3), suppose that both A and B are G-invariant. Further suppose that there exists a special set Y such that for every function f, there exists $T \in G$ such that $T(f) \in Y$. Then it suffices to prove (3) for $f \in Y$ as this together with G-invariance will imply that the same equality holds for all $f \in G(Y) = X$.

Example 2.4.2. Continuing from Example 2.4.1, we may instead use the phase change symmetry $(z, w) \mapsto (e^{i\theta}z, e^{i\theta}w)$ in order to cancel out the phase of $z + w$ and turn it into a non-negative real number. We obtain

$$|z+w| = |e^{i\theta}(z+w)| = e^{i\theta}z + e^{i\theta}w = \mathrm{Re}(e^{i\theta}z) + \mathrm{Re}(e^{i\theta}w) \leq |e^{i\theta}z| + |e^{i\theta}w| = |z| + |w|$$

as desired.

3 Tensor Power Trick

A particularly subtle, yet powerful technique for establishing inequalities is the tensor power trick. The general outline for this technique is as follows: suppose that one wants to prove an inequality of the form $X \le Y$, where X and Y are non-negative quantities, but is only able to prove a weaker version $X \le CY$ for some constant C. By replacing all quantities with suitably chosen "tensor powers" of itself, we hope to obtain an inequality of the form $X^n \le CY^n$ for all positive integers n. This technique works best when the constant C is "dimensionally independent" and does not scale with n. By letting $n \to \infty$ and taking n-th roots, we accomplish our original goal.

In this section, we will showcase how the tensor power trick can be applied to many different subfields of mathematics.

3.1 Inequalities in L^p-spaces

Example 3.1.1. (Convexity of L^p-norms) Let (X, Σ, μ) be a measure space and let $f : X \to \mathbb{C}$ be a measurable function satisfying

$$\int_X |f(x)|^p \, d\mu(x) \le 1 \quad \text{and} \quad \int_X |f(x)|^q \, d\mu(x) \le 1$$

for some $0 < p < q < \infty$. We aim to show that

$$\int_X |f(x)|^r \, d\mu(x) \le 1 \quad \text{for all } p < r < q.$$

As a first attempt to solve this problem, we use the observation that $|f(x)|^r \le |f(x)|^p$ when $|f(x)| \le 1$ and $|f(x)|^r \le |f(x)|^q$ when $|f(x)| \ge 1$. Thus the inequality

$$|f(x)|^r \le |f(x)|^p + |f(x)|^q$$

holds for all $x \in X$. Integrating both sides gives

$$\int_X |f(x)|^r \, d\mu(x) \le 2$$

which is off by a factor of 2 from what we desire. We now demonstrate how the tensor power trick can be used to eliminate the extra constant. For each integer $N \in \mathbb{Z}_{\ge 1}$, let (X^N, Σ_N, μ_N) denote the product space (with the product measure) and define $f^{\otimes N} : X^N \to \mathbb{C}$ by

$$f^{\otimes N}(x_1, x_2, \cdots, x_N) := \prod_{i=1}^{N} f(x_i).$$

By Fubini-Tonelli theorem, we get

$$\int_{X^N} |f^{\otimes N}(\boldsymbol{x})|^p \ \mu_N(\boldsymbol{x}) = \int_X \cdots \int_X \prod_{i=1}^N |f(x_i)|^p \ d\mu(x_1) \cdots d\mu(x_N)$$
$$= \prod_{i=1}^N \int_X |f(x_i)|^p \ d\mu(x_i) \leq 1.$$

Similarly, we have

$$\int_{X^N} |f^{\otimes N}(\boldsymbol{x})|^q \ d\mu_N(\boldsymbol{x}) \leq 1.$$

If we now repeat the same argument as before but with $f^{\otimes N}$ instead of f, we get

$$\int_{X^N} |f^{\otimes N}(\boldsymbol{x})|^r \ d\mu_N(\boldsymbol{x}) \leq 2.$$

Applying Fubini-Tonelli theorem yet again yields

$$\left[\int_X |f(x)|^r \ d\mu(x)\right]^N \leq 2.$$

Taking N-th roots and letting $N \to \infty$ gives the desired result.

Corollary 3.1.2. Let (X, Σ, μ) be a measure space and let $f : X \to \mathbb{C}$ be a measurable function. Let $0 < p < r < q < \infty$. We have the inequality

$$\|f\|_{L^r(X,\mu)} \leq \|f\|_{L^p(X,\mu)}^{1-\theta} \|f\|_{L^q(X,\mu)}^{\theta}$$

where θ satisfies $\dfrac{1}{r} = \dfrac{1-\theta}{p} + \dfrac{\theta}{q}$.

From Example 3.1.1, we see that the above inequality is immediate if

$$\|f\|_{L^p(X,\mu)} = \|f\|_{L^q(X,\mu)} = 1. \tag{8}$$

To prove the statement for the general case, we perform suitable rescalings of both the function f and the measure μ to reduce to the case of (8). For any positive scalars $\lambda, \gamma \in \mathbb{R}_{>0}$, we have that

$$\|\lambda f\|_{L^p(X,\gamma\mu)} = \left(\int_X |\lambda f(x)|^p \ d\gamma\mu(x)\right)^{\frac{1}{p}}$$
$$= \lambda \left(\int_X |f(x)|^p \ d\gamma\mu(x)\right)^{\frac{1}{p}}$$
$$= \lambda\gamma^{\frac{1}{p}} \left(\int_X |f(x)|^p \ d\mu(x)\right)^{\frac{1}{p}}$$
$$= \lambda\gamma^{\frac{1}{p}} \|f\|_{L^p(X,\mu)}.$$

Similarly, $\|\lambda f\|_{L^q(X,\gamma\mu)} = \lambda\gamma^{\frac{1}{q}} \|f\|_{L^q(X,\mu)}$. With an appropriate choice of λ and γ:

$$\gamma = \left[\frac{\|f\|_{L^q(X,\mu)}}{\|f\|_{L^p(X,\mu)}}\right]^{\frac{pq}{q-p}} \quad \text{and} \quad \lambda = \|f\|_{L^p(X,\mu)}^{\frac{p}{q-p}} \|f\|_{L^q(X,\mu)}^{\frac{q}{p-q}},$$

we get

$$\|\lambda f\|_{L^p(X,\gamma\mu)} = \|\lambda f\|_{L^q(X,\gamma\mu)} = 1.$$

We then have

$$\|\lambda f\|_{L^r(X,\gamma\mu)} \leq \|\lambda f\|_{L^p(X,\gamma\mu)}^{1-\theta} \|\lambda f\|_{L^q(X,\gamma\mu)}^{\theta}$$

which simplifies to

$$\lambda\gamma^{\frac{1}{r}} \|f\|_{L^r(X,\mu)} \leq \left(\lambda^{1-\theta}\gamma^{\frac{1-\theta}{p}}\right) \|f\|_{L^p(X,\mu)}^{1-\theta} \left(\lambda^{\theta}\gamma^{\frac{\theta}{q}}\right) \|f\|_{L^q(X,\mu)}^{\theta}.$$

By performing some cancellation on both sides using $1/r = (1-\theta)/p + \theta/q$, we obtain the desired result.

Remark 3.1.3. The previous corollary implies that if f is any measurable function having finite p-norm for all $p \in \mathbb{R}_{\geq 1}$, the map

$$p \mapsto \|f\|_{L^{\frac{1}{p}}(X,\mu)}$$

is logarithmically convex on $(0,1]$.

Example 3.1.4. (Hausdorff-Young inequality) Let G be a finite abelian group and let $f : G \to \mathbb{C}$ be a function. Let $\widehat{G}$ denote the group of characters $\chi : G \to S^1$, where we identify S^1 with the unit circle in the complex plane. Define the Fourier transform $\widehat{f} : \widehat{G} \to \mathbb{C}$ by the formula

$$\widehat{f}(\chi) := \frac{1}{|G|}\sum_{x\in G} f(x)\overline{\chi(x)}.$$

We aim to prove the inequality

$$\left(\sum_{\chi\in\widehat{G}} |\widehat{f}(\chi)|^q\right)^{\frac{1}{q}} \leq \left(\frac{1}{|G|}\sum_{x\in G} |f(x)|^p\right)^{\frac{1}{p}}$$

for all p, q satisfying $1 < p < 2$ and $\dfrac{1}{p} + \dfrac{1}{q} = 1$. Firstly, by triangle inequality, we get

$$\sup_{\chi\in\widehat{G}} |\widehat{f}(\chi)| \leq \frac{1}{|G|}\sum_{x\in G} |f(x)| \tag{9}$$

and by Plancherel's theorem, we have

$$\left(\sum_{\chi\in\widehat{G}} |\widehat{f}(\chi)|^2\right)^{\frac{1}{2}} = \left(\frac{1}{|G|}\sum_{x\in G} |f(x)|^2\right)^{\frac{1}{2}}. \tag{10}$$

Suppose that f is supported on a set $A \subseteq G$ and $|f|$ take values in the dyadic interval $[2^m, 2^{m+1})$ for some integer m. Then (9) and (10) gives

$$\sup_{\chi \in \widehat{G}} |\widehat{f}(\chi)| \leq \frac{|A|}{|G|} 2^{m+1} \tag{11}$$

and

$$\left(\sum_{\chi \in \hat{G}} |\widehat{f}(\chi)|^2 \right)^{\frac{1}{2}} \leq \left(\frac{|A|}{|G|} \right)^{\frac{1}{2}} 2^{m+1} \tag{12}$$

respectively. Since $1 < p < 2$, its Hölder conjugate q satisfies $2 < q < \infty$. We get the upper bound

$$\begin{aligned}
\left(\sum_{\chi \in \widehat{G}} |\widehat{f}(\chi)|^q \right)^{\frac{1}{q}} &\leq \left(\sup_{\chi \in \widehat{G}} |\widehat{f}(\chi)|^{q-2} \sum_{\chi \in \widehat{G}} |\widehat{f}(\chi)|^2 \right)^{\frac{1}{q}} \\
&\leq \frac{|A|^{1-\frac{2}{q}}}{|G|^{1-\frac{2}{q}}} 2^{(m+1)(1-\frac{2}{q})} \left[\left(\sum_{\chi \in \widehat{G}} |\widehat{f}(\chi)|^2 \right)^{\frac{1}{2}} \right]^{\frac{2}{q}} \\
&\leq \frac{|A|^{1-\frac{2}{q}}}{|G|^{1-\frac{2}{q}}} 2^{(m+1)(1-\frac{2}{q})} \left(\frac{|A|^{\frac{1}{q}}}{|G|^{\frac{1}{q}}} \right) 2^{(m+1)(\frac{2}{q})} \\
&= \frac{|A|^{1-\frac{1}{q}}}{|G|^{1-\frac{1}{q}}} 2^{m+1} = \frac{|A|^{\frac{1}{p}}}{|G|^{\frac{1}{p}}} 2^{m+1}.
\end{aligned}$$

Hence we obtain

$$\left(\sum_{\chi \in \widehat{G}} |\widehat{f}(\chi)|^q \right)^{\frac{1}{q}} \leq 2 \left(\frac{1}{|G|} \sum_{x \in G} |f(x)|^p \right)^{\frac{1}{p}}. \tag{13}$$

So far, (13) only holds under the assumption that the image of $|f|$ is contained inside a dyadic interval $[2^m, 2^{m+1})$. For a general function f, we proceed as follows: choose integers j and k such that

$$2^k \leq \max_{\substack{x \in G \\ f(x) \neq 0}} |f(x)| < 2^{k+1} \quad \text{and} \quad 2^j \leq \min_{\substack{x \in G \\ f(x) \neq 0}} |f(x)| < 2^{j+1}.$$

Then we decompose f as a sum

$$f = f_j + f_{j+1} + \cdots + f_{k-1} + f_k$$

where for each $x \in G$, either $|f_i(x)| = 0$ or $|f_i(x)|$ lies in the dyadic interval $[2^i, 2^{i+1})$.

By linearity of the Fourier transform and Minkowski's inequality, we get

$$\begin{aligned}\left(\sum_{\chi\in\widehat{G}}|\widehat{f}(\chi)|^q\right)^{\frac{1}{q}} &\leq \sum_{i=j}^{k}\left(\sum_{\chi\in\widehat{G}}|\widehat{f}_i(\chi)|^q\right)^{\frac{1}{q}} \\ &\leq 2\sum_{i=j}^{k}\left(\frac{1}{|G|}\sum_{x\in G}|f_i(x)|^p\right)^{\frac{1}{p}}.\end{aligned}$$

Let $S_i = \left(\frac{1}{|G|}\sum_{x\in G}|f_i(x)|^p\right)^{\frac{1}{p}}$ and let $M = \max\left\{S_j, S_{j+1}, \cdots, S_k\right\}$. Observe that

(i) For each i, we have $S_i \leq 2^{i+1}$;

(ii) $\frac{2^k}{|G|} \leq M$.

By a direct computation (where the logarithm is taken to be in base 2),

$$\begin{aligned}\left(\sum_{\chi\in\widehat{G}}|\widehat{f}(\chi)|^q\right)^{\frac{1}{q}} &\leq 2\left[\sum_{i=j}^{k-\lfloor\log|G|\rfloor} S_i + \sum_{i=k-\lfloor\log|G|\rfloor}^{k} S_i\right] \\ &\leq 2\left[\sum_{i=j}^{k-\lfloor\log|G|\rfloor} 2^{i+1} + \sum_{i=k-\lfloor\log|G|\rfloor}^{k} M\right] \\ &\leq 2^{k+3-\log|G|} + 2(1+\log|G|)M \\ &= 8\left(\frac{2^k}{|G|}\right) + 2(1+\log|G|)M \\ &\leq (10+2\log|G|)M \\ &\leq (10+2\log|G|)\left(\frac{1}{|G|}\sum_{x\in G}|f(x)|^p\right)^{\frac{1}{p}}\end{aligned}$$

where the second last inequality follows from observation (ii). So far, we have established the upper bound

$$\left(\sum_{\chi\in\widehat{G}}|\widehat{f}(\chi)|^q\right)^{\frac{1}{q}} \leq (10+2\log|G|)\left(\frac{1}{|G|}\sum_{x\in G}|f(x)|^p\right)^{\frac{1}{p}} \tag{14}$$

which is off from what we want by a logarithmic factor.

Although the factor now scales with $|G|$, the tensor power trick can still be applied. Define the tensor power $G^{\otimes N}$ and $f^{\otimes N}: G^{\otimes N} \to \mathbb{C}$ by

$$G^{\otimes N} := \underbrace{G \times \cdots \times G}_{\text{n times}} \quad \text{and} \quad f(x_1, \cdots, x_n) := f(x_1) \cdots f(x_n).$$

Since G is abelian, the group of characters of $G^{\otimes N}$ is simply the direct product of the group of characters of G. Thus every character χ of $\widehat{G^{\otimes N}}$ is of the form

$$\chi(x_1, x_2, \cdots, x_n) = \chi_1(x_1)\chi_2(x_2)\cdots\chi_N(x_N)$$

for some $\chi_1, \cdots, \chi_N \in \widehat{G}$. The Fourier transform of $f^{\otimes N}$ is given by

$$\begin{aligned}
\widehat{f^{\otimes N}}(\chi) &= \frac{1}{|G^{\otimes N}|} \sum_{x \in G^{\otimes N}} f^{\otimes N}(x)\overline{\chi(x)} \\
&= \frac{1}{|G|^N} \sum_{x_1 \in G} \cdots \sum_{x_N \in G} f(x_1)\cdots f(x_N)\overline{\chi_1(x_1)}\cdots\overline{\chi_N(x_N)} \\
&= \frac{1}{|G|^N} \left(\sum_{x_1 \in G} f(x_1)\overline{\chi_1(x_1)}\right)\left(\sum_{x_2 \in G} f(x_2)\overline{\chi_2(x_2)}\right)\cdots\left(\sum_{x_N \in G} f(x_N)\overline{\chi_N(x_N)}\right) \\
&= \widehat{f}(\chi_1)\cdots\widehat{f}(\chi_N).
\end{aligned}$$

Thus the left-hand side of equation (14) becomes

$$\begin{aligned}
\left(\sum_{\chi \in \widehat{G^{\otimes N}}} |\widehat{f^{\otimes N}}(\chi)|^q\right)^{\frac{1}{q}} &= \left(\sum_{\chi_1 \in \widehat{G}} \cdots \sum_{\chi_N \in \widehat{G}} |\widehat{f}(\chi_1)|^q \cdots |\widehat{f}(\chi_N)|^q\right)^{\frac{1}{q}} \\
&= \left(\sum_{\chi_1 \in \widehat{G}} |\widehat{f}(\chi_1)|^q \cdots \sum_{\chi_N \in \widehat{G}} |\widehat{f}(\chi_N)|^q\right)^{\frac{1}{q}} \\
&= \left(\sum_{\chi \in \widehat{G}} |\widehat{f}(\chi)|^q\right)^{\frac{N}{q}}.
\end{aligned}$$

On the other hand, the right-hand side of equation (14) becomes

$$\begin{aligned}
\left(\frac{1}{|G^{\otimes N}|} \sum_{x \in G^{\otimes N}} |f^{\otimes N}(x)|^p\right)^{\frac{1}{p}} &= \left(\frac{1}{|G|^N} \sum_{x_1 \in G} \cdots \sum_{x_N \in G} |f(x_1)|^p \cdots |f(x_N)|^p\right)^{\frac{1}{p}} \\
&= \left(\frac{1}{|G|^N} \sum_{x_1 \in G} |f(x_1)|^p \cdots \sum_{x_N \in G} |f(x_N)|^p\right)^{\frac{1}{p}} \\
&= \left(\frac{1}{|G|} \sum_{x \in G} |f(x)|^p\right)^{\frac{N}{p}}.
\end{aligned}$$

To summerise, we have obtained, for all $N \in \mathbb{Z}_{\geq 1}$, the estimate

$$\left(\sum_{\chi \in \widehat{G}} |\widehat{f}(\chi)|^q\right)^{\frac{N}{q}} \leq (10 + 2N \log |G|)\left(\frac{1}{|G|}\sum_{x \in G} |f(x)|^p\right)^{\frac{N}{p}}.$$

Taking N-th roots and letting $N \to \infty$ eliminates the logarithmic dependence (since $N^{\frac{1}{N}} \to 1$) and gives us the desired result.

3.2 Bound on sumsets

While the tensor power trick is mainly used in analysis, it has also found applications in discrete settings. The next example arises in a relatively new branch of mathematics known as additive combinatorics (in fact the name itself has been attributed to Terence Tao and Van H. Vu in their textbook published in 2006).

The central objects of study in additive combinatorics are additive groups and their subsets. A major research area within the field is establishing lower and upper bounds on certain types of subsets.

Definition 3.2.1. Let A and B be subsets of an additive (abelian) group G. Define the **sumset** $A+B$ by

$$A+B := \big\{a+b \ : \ a \in A;\ b \in B\big\}.$$

For $k \in \mathbb{Z}_{\geq 0}$, define the **iterated sumset** kA by

$$kA := \underbrace{A+\cdots+A}_{k \text{ times}}.$$

Example 3.2.2. An important inequality in additive combinatorics, due to Plünnecke, asserts that for any finite non-empty subsets A and B of an additive group G and any positive integer k, we have the bound

$$|kB| \leq \frac{|A+B|^k}{|A|^{k-1}}. \tag{15}$$

Our goal will be to derive the more general inequality

$$|B_1+\cdots+B_k| \leq \frac{|A+B_1|\cdots|A+B_k|}{|A|^{k-1}} \tag{16}$$

from (15). Inequality (15) can be seen as a special case of (16) via $B = B_1 = \cdots = B_k$.

A naive approach to the problem is to apply (15) to $B = B_1 \cup B_2 \cup \cdots \cup B_k$. We get

$$|B_1+\cdots+B_k| \leq |kB| \leq \frac{|A+B|^k}{|A|^{k-1}} \leq \frac{\big(|A+B_1|+\cdots+|A+B_k|\big)^k}{|A|^{k-1}}. \tag{17}$$

However, the right-hand side is way too big. To rectify the issue, we replace G with the larger group $\widetilde{G} = G \times \mathbb{Z}^k$ and replace each set B_i with the larger set

$$\widetilde{B}_i = B_i \times \{e_i, 2e_i, \cdots, N_i e_i\}.$$

for some positive integer N_i whose exact value will be determined later. Here, $\{e_1, \cdots, e_k\}$ is the standard basis for $\mathbb{Z}^k$. To ensure that the addition in the group remains compatible, we also replace A with $\widetilde{A} = A \times \{0\}$.

To choose an element from $\widetilde{B}_1 + \cdots + \widetilde{B}_k$, we first choose an element from $B_1 + \cdots + B_k$ and then, for each j, choose an element from $\{e_j, \cdots, N_j e_j\}$. Thus the left-hand side of equation (17) becomes

$$|\widetilde{B}_1 + \cdots + \widetilde{B}_k| = N_1 \cdots N_k |B_1 + \cdots + B_k|. \tag{18}$$

On the other hand, for each j, we have that $|\widetilde{A} + \widetilde{B}_j| = N_j|A + B_j|$. Thus the right-hand side of (17) becomes

$$\frac{\big(|\widetilde{A} + \widetilde{B}_1| + \cdots + |\widetilde{A} + \widetilde{B}_k|\big)^k}{|\widetilde{A}|^{k-1}} = \frac{\big(N_1|A + B_1| + \cdots + N_k|A + B_k|\big)^k}{|A|^{k-1}}. \tag{19}$$

For each j, we choose $N_j = \dfrac{|A + B_1| \cdots |A + B_k|}{|A + B_j|}$. We have that

$$N_1|A + B_1| = N_2|A + B_2| = \cdots = N_k|A + B_k| = |A + B_1||A + B_2| \cdots |A + B_k|.$$

Putting (18) and (19) together, we get

$$|A + B_1|^{k-1} \cdots |A + B_k|^{k-1} |B_1 + \cdots + B_k| \le k^k \frac{|A + B_1|^k |A + B_2|^k \cdots |A + B_k|^k}{|A|^{k-1}}$$

which simplifies to

$$|B_1 + \cdots + B_k| \le k^k \frac{|A + B_1||A + B_2| \cdots |A + B_k|}{|A|^{k-1}}. \tag{20}$$

Finally, we will remove the constant factor of k^k using the tensor power trick. For each $n \in \mathbb{Z}_{\ge 1}$, we replace the set A and B_j by their cartesian products A^n and B_j^n respectively. The left-hand side of (20) becomes

$$|B_1^n + \cdots + B_k^n| = |B_1 + \cdots + B_k|^n.$$

To see this, observe that making a choice for $B_1^n + \cdots + B_k^n$ is equivalent to making a choice for $B_1 + \cdots + B_k$ a total of n times, once for each one of the n coordinates. Similarly, the right-hand side becomes

$$k^k \frac{|A^n + B_1^n||A^n + B_2^n| \cdots |A^n + B_k^n|}{|A^n|^{k-1}} = k^k \frac{\big(|A + B_1||A + B_2| \cdots |A + B_k|\big)^n}{|A|^{n(k-1)}}.$$

To summerise, we get

$$|B_1 + \cdots + B_k|^n \le k^k \frac{\big(|A + B_1||A + B_2| \cdots |A + B_k|\big)^n}{|A|^{n(k-1)}}$$

Taking limit as $n \to \infty$ once again gives the desired result.

3.3 Cotlar-Knapp-Stein Lemma

Example 3.3.1. Consider a sequence of bounded linear operators $T_1, T_2, \cdots : H \to H$ on a Hilbert space. Suppose that there exists a constant $A \in \mathbb{R}$ such that we have the uniform bound on the operator norm

$$\|T_i\|_{\text{op}} \leq A \tag{21}$$

for all i. This alone is not enough to ensure that the operator norm of the sum $\sum_{i=1}^n T_i$ is uniformly bounded in n since it can be as large as An. However, with the stronger almost-orthogonal assumption:

$$\sum_{j=1}^n \left\|T_i T_j^*\right\|_{\text{op}}^{1/2} \leq A \tag{22}$$

$$\sum_{j=1}^n \|T_i^* T_j\|_{\text{op}}^{1/2} \leq A \tag{23}$$

for all positive integers i and n, then the Cotlar-Knapp-Stein lemma asserts that the series $\sum_{i=1}^\infty T_i$ now converges in the operator norm with the bound

$$\left\|\sum_{i=1}^n T_i\right\|_{\text{op}} \leq A$$

that is uniform in n. The fact that (22) and (23) is a stronger assumption than (21) is due to the following special property of self-adjoint operators:

Lemma 3.3.2. Let $T : H \to H$ be a bounded linear operator on a Hilbert space. Then

$$\|T\|_{\text{op}}^2 = \|T^*T\|_{\text{op}} = \|TT^*\|_{\text{op}}.$$

Proof. It suffices to prove the equality $\|T\|_{\text{op}}^2 = \|T^*T\|_{\text{op}}$. The other inequality can be obtained by replacing T with T^* (and the observation that $\|T\|_{\text{op}} = \|T^*\|_{\text{op}}$).

On one hand, the inequality

$$\|T^*T\|_{\text{op}} \leq \|T^*\|_{\text{op}} \|T\|_{\text{op}} = \|T\|_{\text{op}}^2$$

holds for all bounded linear operators. The other direction follows from

$$\|T\|_{\text{op}}^2 = \sup_{\substack{f \in H \\ \|f\|_H = 1}} \|T(f)\|_H^2 = \sup_{\substack{f \in H \\ \|f\|_H = 1}} \left\langle T^*(T(f)), f \right\rangle_H \leq \sup_{\substack{f \in H \\ \|f\|_H = 1}} \|T^*T(f)\|_H = \|T^*T\|_{\text{op}}$$

where the second last step is due to Cauchy-Schwartz inequality.

Thus (21) can be obtained from (22) by a direct computation

$$\|T_i\|_{\mathrm{op}} = \|T_i T_i^*\|_{\mathrm{op}}^{1/2} \leq \sum_{j=1}^{n} \left\|T_i T_j^*\right\|_{\mathrm{op}}^{1/2} \leq A$$

as long as $n \geq i$.

To prove the main result, we first use Lemma 3.3.2 and expand via the triangle inequality to obtain

$$\begin{aligned}\left\|\sum_{i=1}^{n} T_i\right\|_{\mathrm{op}} &= \left\|\left(\sum_{i=1}^{n} T_i^*\right)\left(\sum_{j=1}^{n} T_j\right)\right\|_{\mathrm{op}}^{1/2} \leq \left(\sum_{i=1}^{n}\sum_{j=1}^{n} \|T_i^* T_j\|_{\mathrm{op}}\right)^{1/2} \\ &\leq \sum_{i=1}^{n}\sum_{j=1}^{n} \|T_i^* T_j\|_{\mathrm{op}}^{1/2} \leq A n^{1/2}.\end{aligned}$$

Here the second last inequality follows from the elementary fact that for all non-negative reals a and b, we have $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$. This is already an improvement over the original bound of An. To improve on this even further, we repeat the process with $(\sum_{i=1}^{n} T_i^*)(\sum_{j=1}^{n} T_j)$ instead of $(\sum_{i=1}^{n} T_i)$. Iterating the process k times gives

$$\begin{aligned}\left\|\sum_{i=1}^{n} T_i\right\|_{\mathrm{op}} &= \left\|\left(\sum_{i_1=1}^{n} T_{i_1}^*\right)\left(\sum_{i_2=1}^{n} T_{i_2}\right)\right\|_{\mathrm{op}}^{1/2} \\ &= \left\|\left(\sum_{i_1=1}^{n} T_{i_1}^*\right)\left(\sum_{i_2=1}^{n} T_{i_2}\right)\left(\sum_{i_3=1}^{n} T_{i_3}^*\right)\left(\sum_{i_4=1}^{n} T_{i_4}\right)\right\|_{\mathrm{op}}^{1/4} \\ &= \cdots \\ &= \left\|\left(\sum_{i_1=1}^{n} T_{i_1}^*\right)\left(\sum_{i_2=1}^{n} T_{i_2}\right)\left(\sum_{i_3=1}^{n} T_{i_3}^*\right)\cdots\left(\sum_{i_{M-1}=1}^{n} T_{i_{M-1}}^*\right)\left(\sum_{i_M=1}^{n} T_{i_M}\right)\right\|_{\mathrm{op}}^{1/M}\end{aligned}$$

where $M = 2^k$. Expanding out gives

$$\left\|\sum_{i=1}^{n} T_i\right\|_{\mathrm{op}} \leq \left(\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\cdots\sum_{i_M=1}^{n} \left\|T_{i_1}^* T_{i_2}\cdots T_{i_{M-1}}^* T_{i_M}\right\|_{\mathrm{op}}\right)^{1/M}. \tag{24}$$

We establish two different bounds on $\left\|T_{i_1}^* T_{i_2}\cdots T_{i_{M-1}}^* T_{i_M}\right\|_{\mathrm{op}}$, namely

$$\left\|T_{i_1}^* T_{i_2}\cdots T_{i_{M-1}}^* T_{i_M}\right\|_{\mathrm{op}} \leq \left\|T_{i_1}^* T_{i_2}\right\|_{\mathrm{op}} \cdots \left\|T_{i_{M-1}}^* T_{i_M}\right\|_{\mathrm{op}} \tag{25}$$

$$\left\|T_{i_1}^* T_{i_2}\cdots T_{i_{M-1}}^* T_{i_M}\right\|_{\mathrm{op}} \leq A^2 \left\|T_{i_2} T_{i_3}^*\right\|_{\mathrm{op}} \cdots \left\|T_{i_{M-2}} T_{i_{M-1}}^*\right\|_{\mathrm{op}}. \tag{26}$$

By taking the geometric mean of (25) and (26), we obtain

$$\left\|T_{i_1}^* T_{i_2} \cdots T_{i_{M-1}}^* T_{i_M}\right\|_{\text{op}} \le A\left\|T_{i_1}^* T_{i_2}\right\|_{\text{op}}^{1/2}\left\|T_{i_2} T_{i_3}^*\right\|_{\text{op}}^{1/2} \cdots \left\|T_{i_{M-1}}^* T_{i_M}\right\|_{\text{op}}^{1/2}.$$

Applying the bound to (24) and collapsing the sum, we get

$$\begin{aligned}
\left\|\sum_{i=1}^{n} T_i\right\|_{\text{op}} &\le \left(A \sum_{i_1=1}^{n}\sum_{i_2=1}^{n} \cdots \sum_{i_M=1}^{n} \left\|T_{i_1}^* T_{i_2}\right\|_{\text{op}}^{1/2}\left\|T_{i_2} T_{i_3}^*\right\|_{\text{op}}^{1/2} \cdots \left\|T_{i_{M-1}}^* T_{i_M}\right\|_{\text{op}}^{1/2}\right)^{1/M} \\
&\le \left(A^2 \sum_{i_1=1}^{n} \cdots \sum_{i_{M-1}=1}^{n} \left\|T_{i_1}^* T_{i_2}\right\|_{\text{op}}^{1/2} \cdots \left\|T_{i_{M-2}} T_{i_{M-1}}^*\right\|_{\text{op}}^{1/2}\right)^{1/M} \\
&\le \cdots \\
&\le \left(A^{M-1} \sum_{i_1=1}^{n}\sum_{i_2=1}^{n} \left\|T_{i_1}^* T_{i_2}\right\|_{\text{op}}^{1/2}\right)^{1/M} \\
&\le \left(A^M \sum_{i_1=1}^{n} 1\right)^{1/M} \\
&= An^{1/M}.
\end{aligned}$$

So far, the above result only holds in the case in which M is a power of 2. Thankfully, the tensor power trick can still be applied as we can still take limit as $M \to \infty$ to eliminate the dependence on n.

Finally, to see why the series converge, observe that since (22) and (23) are series with non-negative terms, they are Cauchy. For any $\epsilon > 0$, there exists $N \in \mathbb{Z}_{\ge 1}$ such that

$$n \ge N \implies \sum_{j=N}^{n} \left\|T_i T_j^*\right\|_{\text{op}}^{1/2} < \epsilon \text{ and } \sum_{j=N}^{n} \|T_i^* T_j\|_{\text{op}}^{1/2} < \epsilon.$$

By applying the Cotlar-Knapp-Stein lemma to the sequence $(T_i)_{i=N}^{\infty}$ instead, we get

$$\left\|\sum_{i=N}^{n} T_i\right\|_{\text{op}} \le \epsilon$$

for all $n \in \mathbb{Z}_{\ge N}$. Thus the series $\sum_{i=1}^{\infty} T_i$ is Cauchy and hence converges.

3.4 Entropy Estimates

In probability theory, replacing mathematical objects with higher-dimensional analogues of themselves can also be a powerful way to impose a certain level of regularity due to the various limit theorems.

Example 3.4.1. Let X be a random variable in a probability space $(\Omega, \mathcal{F}, P)$ taking values in the finite set $\mathcal{X} = \{x_1, \cdots, x_k\}$. Let $p(x) = p_X(x)$ denote the probability mass function of X. Define the **entropy** $H(X)$ of X by

$$H(X) := -\sum_{x\in\mathcal{X}} p(x) \ln p(x) \tag{27}$$

with the convention $0 \ln 0 = 0$. Here, we have chosen the natural logarithm (as opposed to base 2) for convenience. For example, if X is uniformly distributed over $\mathcal{X}$, then

$$H(X) = -\sum_{x\in\mathcal{X}} \frac{1}{k} \ln \frac{1}{k} = \ln k. \tag{28}$$

If X is not uniformly distributed, then the entropy formula is more complicated. However, by using the tensor power trick in conjunction with the weak law of large numbers, the formula of $H(X)$ becomes "approximately uniform" in a certain sense. This can then be used to give proofs for various fundamental inequalities in information theory.

Define $X^{\otimes N} = (X_1, \cdots, X_N)$ by taking N independent and identically distributed copies of X. If X is uniformly distributed on k values, then $X^{\otimes N}$ is uniformly distributed on k^N values. By (28), we have the formula

$$H(X^{\otimes N}) = NH(X). \tag{29}$$

However, even when X is not uniformly distributed, (29) still holds. To see this, we start with the trivial base case of $N = 1$ and argue by induction as follows:

$$\begin{aligned}
H(X^{\otimes N}) &= -\sum_{(x_1,\cdots,x_N)\in\mathcal{X}^{\otimes N}} p_{X^{\otimes N}}(x_1, \cdots, x_N) \ln p_{X^{\otimes N}}(x_1, \cdots, x_N) \\
&= -\sum_{x_1\in\mathcal{X}} \cdots \sum_{x_N\in\mathcal{X}} p_{X_1}(x_1) \cdots p_{X_N}(x_N) \ln \left[p_{X_1}(x_1) \cdots p_{X_N}(x_N)\right] \\
&= -\sum_{x_1\in\mathcal{X}} \cdots \sum_{x_N\in\mathcal{X}} p_{X_1}(x_1) \cdots p_{X_N}(x_N) \ln \left[p_{X_1}(x_1) \cdots p_{X_{N-1}}(x_{N-1})\right] \\
&\quad - \sum_{x_1\in\mathcal{X}} \cdots \sum_{x_N\in\mathcal{X}} p_{X_1}(x_1) \cdots p_{X_N}(x_N) \ln p_{X_N}(x_N) \\
&= -\sum_{x_1\in\mathcal{X}} \cdots \sum_{x_{N-1}\in\mathcal{X}} p_{X_1}(x_1) \cdots p_{X_{N-1}}(x_{N-1}) \ln \left[p_{X_1}(x_1) \cdots p_{X_{N-1}}(x_{N-1})\right] \\
&\quad - \sum_{x_N\in\mathcal{X}} p_{X_N}(x_N) \ln p_{X_N}(x_N) \\
&= (N-1)H(X) + H(X) = NH(X).
\end{aligned}$$

The main ingredient that makes this approximation possible is the weak law of large numbers. More formally, for any $\delta, \epsilon \in \mathbb{R}_{>0}$, we have the following bound

$$P\left(\left|\frac{\mathbb{1}_{\{X_1=x_j\}} + \mathbb{1}_{\{X_2=x_2\}} + \cdots + \mathbb{1}_{\{X_N=x_j\}}}{N} - p(x_j)\right| \geq \delta\right) < \epsilon \tag{30}$$

for all sufficiently large integers N and $x_j \in \mathcal{X}$ (where we have used the fact that $\mathcal{X}$ is finite).

This implies that with probability at least $1-\epsilon$, each alphabet x_j will be attained by $p(x_j)N + \gamma_j N$ of the N trials for some $|\gamma_j| < \delta$. We first make two observations:

(i) $\sum_{j=1}^{k} \gamma_j = 0$;

(ii) For all j, we have $\ln\big(p(x_j) + \gamma_j\big) - \ln p(x_j) = O(\delta)$.

For each such configuration $\big(p(x_1)N + \gamma_1 N, \cdots, p(x_k)N + \gamma_k N\big)$, the number of possible combinations is given by

$$\binom{N}{p(x_1)N + \gamma_1 N}\binom{N - p(x_1)N - \gamma_1 N}{p(x_2)N + \gamma_2 N} \cdots \binom{p(x_k)N + \gamma_k N}{p(x_k)N + \gamma_k N}.$$

This is a telescoping product that evaluates to

$$C := \frac{N!}{(p(x_1)N + \gamma_1 N)!(p(x_2)N + \gamma_2 N)! \cdots (p(x_k)N + \gamma_k N)!}.$$

Taking logarithm on both sides and applying Stirling's approximation formula

$$\ln(n!) = n \ln n - n + \Theta(\ln n)$$

we obtain

$$\begin{aligned}
\ln C &= \ln N! - \sum_{j=1}^{k} \ln\big(p(x_j)N + \gamma_j N\big)! \\
&= N \ln N - N - \sum_{j=1}^{k} \big(p(x_j)N + \gamma_j N\big) \ln\big(p(x_j)N + \gamma_j N\big) \\
&+ \sum_{j=1}^{k} \big(p(x_j)N + \gamma_j N\big) + \Theta(\ln N).
\end{aligned}$$

Using observations (i) and (ii), the third and fourth term evaluates to

$$
\begin{aligned}
&-\sum_{j=1}^{k}\big(p(x_j)N+\gamma_j N\big)\ln\big(p(x_j)N+\gamma_j N\big)+\sum_{j=1}^{k}\big(p(x_j)N+\gamma_j N\big)\\
&=-\sum_{j=1}^{k}\big(p(x_j)N+\gamma_j N\big)\ln N-N\sum_{j=1}^{k}p(x_j)\ln\big(p(x_j)+\gamma_j\big)-N\sum_{j=1}^{k}\gamma_j\ln\big(p(x_j)+\gamma_j\big)+N\\
&=-N\ln N+NH(X)+N+O(\delta)N.
\end{aligned}
$$

Hence

$$\ln C = NH(X)+O(\delta)N+\Theta(\ln N)$$

and so the number of possible combinations is given by

$$C=e^{N(H(X)+o(1))}.$$

Here, we have used the observation that both $O(\delta)N$ and $\Theta(\ln N)$ can be combined into the expression $o(1)N$.

On the other hand, each configuration occurs with probability

$$p(x_1)^{p(x_1)N+\gamma_1 N}p(x_2)^{p(x_2)N+\gamma_2 N}\cdots p(x_k)^{p(x_k)N+\gamma_k N}.$$

Again, by taking logarithm on both sides, we obtain

$$
\begin{aligned}
\sum_{j=1}^{k}\big(p(x_j)N+\gamma_j N\big)\ln p(x_j) &= N\sum_{j=1}^{N}p(x_j)\ln p(x_j)+O(\delta)N\\
&=-NH(X)+O(\delta)N.
\end{aligned}
$$

Thus each configuration appears with probability $e^{-N(H(X)+o(1))}$.

We now demostrate how the approximately uniform distribution of $X^{\otimes N}$ can be used to give an alternative proof for the subadditivity inequality

$$H(X,Y)\leq H(X)+H(Y)$$

in information theory.

On one hand, for any $\epsilon\in\mathbb{R}_{>0}$, by choosing N sufficiently large, we have that the random variables $X^{\otimes N}$ and $Y^{\otimes N}$ take on $e^{N(H(X)+o(1))}$ and $e^{N(H(Y)+o(1))}$ values respectively with probability at least $1-\epsilon$. Thus the random variable

$$(X,Y)^{\otimes N}=(X^{\otimes N},Y^{\otimes N})$$

will take on at most $e^{N(H(X)+H(Y)+o(1))}$ values. On the other hand, by applying the tensor power trick to the random variable (X,Y) instead, we see that $(X,Y)^{\otimes N}$ takes on $e^{N(H(X,Y)+o(1))}$ values with probability $1-\epsilon$. By taking N-th roots and taking limit as $N\to 0$, we see that this is only possible if $H(X,Y)\leq H(X)+H(Y)$.

3.5 Kabatjanskii-Levenstein Bound

In this section, we investigate the problem of how many "almost orthogonal" unit vectors one can place into the Euclidean space $\mathbb{R}^n$. If we require that $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m$ are exactly orthogonal, then elementary linear algebra dictates that $m \leq n$. However, if we relax the orthogonality conidition just a little by imposing a bound on the magnitude of the dot product $|\langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle|$ instead, then it turns out that we can pack a lot more unit vectors into $\mathbb{R}^n$.

Lemma 3.5.1. Let $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m$ be unit vectors in $\mathbb{R}^n$ such that

$$\left|\langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle\right| \leq \frac{1}{2n^{1/2}} \quad \text{for all } 1 \leq i < j \leq n.$$

Then $m < 2n$.

Proof. Suppose that $m \geq 2n$. We take the first $2n$ vectors $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{2n}$ to form the Gram matrix $\boldsymbol{G} \in \mathrm{Mat}_{2n\times 2n}(\mathbb{R})$ defined by

$$\boldsymbol{G} := \big(\langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle\big)_{1\leq i,j\leq 2n} = \boldsymbol{A}^T\boldsymbol{A}$$

where

$$\boldsymbol{A} := \begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_{2n} \end{pmatrix} \in \mathrm{Mat}_{n\times 2n}(\mathbb{R}).$$

Since $\boldsymbol{G}$ is symmetric with rank at most n, the matrix $\boldsymbol{G} - \boldsymbol{I}_{2n}$ has eigenvalue of -1 with multiplicity at least n. Recall that the Frobenius norm is defined by

$$\|\boldsymbol{B}\|_F := \sqrt{\sum_{1\leq i,j\leq 2n} |b_{i,j}|^2} = \sqrt{\mathrm{tr}(\boldsymbol{B}^T\boldsymbol{B})}.$$

As the diagonal entries of $\boldsymbol{G} - \boldsymbol{I}_{2n}$ are 0, the square of the Frobenius norm of $\boldsymbol{G} - \boldsymbol{I}$ is given by

$$\|\boldsymbol{G} - \boldsymbol{I}\|_F^2 = \sum_{\substack{1\leq i,j\leq 2n \\ i\neq j}} \left|\langle v_i, v_j \rangle\right|^2 \leq (4n^2 - 2n)\frac{1}{4n} = n - \frac{1}{2}. \tag{31}$$

On the other hand,

$$\|\boldsymbol{G} - \boldsymbol{I}\|_F^2 = \lambda_1^2 + \cdots + \lambda_{2n}^2 \geq n \tag{32}$$

where $\lambda_1, \cdots \lambda_{2n}$ are the eigenvalues of $\boldsymbol{G} - \boldsymbol{I}$. Inequality (31) is a direct contradiction to inequality (32) and this completes the proof of the lemma.

Next, we will like to apply the tensor power trick to obtain a more general version of Lemma 3.5.1.

Proposition 3.5.2. Let $k \in \mathbb{Z}_{\geq 1}$ and let $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m \in \mathbb{R}^n$ be unit vectors such that

$$|\langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle| \leq 2^{-\frac{1}{k}} \binom{n+k-1}{k}^{-\frac{1}{2k}} \qquad \text{for all } 1 \leq i < j \leq n.$$

Then $m < 2\binom{n+k-1}{k}$.

Remark 3.5.3. Lemma 3.5.1 corresponds to the case $k = 1$.

Proof. Recall that the inner product on $\mathbb{R}^n$ (and more generally on any Hilbert space) can be naturally extended to the k-th tensor power $(\mathbb{R}^n)^{\otimes k}$ via

$$\left\langle \boldsymbol{v}_1 \otimes \boldsymbol{v}_2 \otimes \cdots \otimes \boldsymbol{v}_k, \boldsymbol{w}_1 \otimes \boldsymbol{w}_2 \otimes \cdots \otimes \boldsymbol{w}_k \right\rangle_{(\mathbb{R}^n)^{\otimes k}} := \langle \boldsymbol{v}_1, \boldsymbol{w}_1 \rangle_{\mathbb{R}^n} \langle \boldsymbol{v}_2, \boldsymbol{w}_2 \rangle_{\mathbb{R}^n} \cdots \langle \boldsymbol{v}_k, \boldsymbol{w}_k \rangle_{\mathbb{R}^n}.$$

As in turns out, applying the tensor power trick to $(\mathbb{R}^n)^{\otimes k}$ directly is not very helpful. Instead, we work with subspace of $(\mathbb{R}^n)^{\otimes k}$ of symmetric tensors instead. For a pure tensor, define its **symmetric component** by

$$\mathrm{Sym}(\boldsymbol{v}_1 \otimes \boldsymbol{v}_2 \otimes \cdots \otimes \boldsymbol{v}_k) := \frac{1}{k!} \sum_{\sigma \in S_k} \boldsymbol{v}_{\sigma(1)} \otimes \boldsymbol{v}_{\sigma(2)} \otimes \cdots \otimes \boldsymbol{v}_{\sigma(k)}.$$

We then define the subspace of symmetric tensors by

$$\mathrm{Sym}^k(\mathbb{R}^n) := \mathrm{span}_{\mathbb{R}} \Big\{ \mathrm{Sym}(\boldsymbol{v}_1 \otimes \boldsymbol{v}_2 \otimes \cdots \otimes \boldsymbol{v}_k) \ : \ \boldsymbol{v}_1, \cdots, \boldsymbol{v}_k \in \mathbb{R}^n \Big\} \subseteq (\mathbb{R}^n)^{\otimes k}.$$

If $\{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n\}$ is the standard basis for $\mathbb{R}^n$, then a basis for $\mathrm{Sym}^k(\mathbb{R}^n)$ is given by

$$\big\{ \mathrm{Sym}(\boldsymbol{e}_{i_1} \otimes \cdots \otimes \boldsymbol{e}_{i_k}) \ : \ 1 \leq i_1 \leq i_2 \leq \cdots \leq i_k \leq n \big\}.$$

Thus $\dim\left(\mathrm{Sym}^k(\mathbb{R}^n)\right) = \binom{n+k-1}{k}$.

While the dimension of $\mathrm{Sym}^k(\mathbb{R}^n)$ is significantly smaller than that of $(\mathbb{R}^n)^{\otimes k}$, the inner product remains unchanged. For each i, define $\boldsymbol{v}_i^{\otimes k} = \underbrace{\boldsymbol{v}_i \otimes \cdots \otimes \boldsymbol{v}_i}_{k\text{-times}}$. Observe that

$$\left| \langle \boldsymbol{v}_i^{\otimes k}, \boldsymbol{v}_j^{\otimes k} \rangle_{(\mathbb{R}^n)^{\otimes k}} \right| = \left| \langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle_{\mathbb{R}^n} \right|^k \leq \frac{1}{2} \binom{n+k-1}{k}^{-\frac{1}{2}}.$$

By applying Lemma 3.5.1 to the vectors $\boldsymbol{v}_1^{\otimes k}, \boldsymbol{v}_2^{\otimes k}, \cdots, \boldsymbol{v}_m^{\otimes k}$ in the vector space $\mathrm{Sym}^k(\mathbb{R}^n)$ instead, we get the desired conclusion.

4 Exploring The Toolkit of Jean Bourgain

Jean Bourgain is renowned for making important contributions to an immense number of subfields of mathematics, some of which are seemingly unrelated to each other at first glance. However, upon closer inspection of Bourgain's work, it turns out that even he relied on a core set of basic tools to tackle the different types of problems. In this section, we will study some of these techniques and how they can be applied to problems in different areas of analysis.

4.1 Quantitative Formulation of Qualitative Problems

In the early 20th century, analysis became increasingly divided into two different categories: "hard" analysis and "soft" analysis. While "hard" analysis is mostly concerned with the quantitative aspects of analysis such as the measure of bounded sets and the norm of integrable functions, "soft analysis" is mostly interested in the qualitative aspects such as the measurability of sets and the continuity of functions.

In general, these two categories use different sets of tools, tackle different kinds of problems and even ask different types of questions. Hence it is common for modern analysts to specialise only in one of the two types. Bourgain himself specialises almost exclusively in "hard analysis". Thus when tackling a "soft analysis" problem, the first step in Bourgain's argument is usually to locate a quantitative estimate that will imply the desired claim.

Example 4.1.1. In the first example, we will look at how Bourgain transforms a qualitative problem in geometric measure theory into a quantitative problem in harmonic analysis.

Theorem 4.1.2. (Furstenberg-Katznelson-Weiss theorem, qualitative version) Let A be a measurable subset of $\mathbb{R}^2$ whose upper density

$$\delta := \limsup_{R\to\infty} \frac{|A \cap B(0,R)|}{|B(0,R)|}$$

is positive. Then there exists $\ell_0 \in \mathbb{R}$ such that for all $\ell \geq \ell_0$, there exists $x, y \in A$ with $|x - y| = \ell$.

Observe that this is purely an existence theorem and does not provide any quantitative bound for the length threshold ℓ_0 in terms of the upper density δ. In fact, no such estimate is possible since if we replace the set A by a rescaled version

$$\lambda \cdot A := \{\lambda x : x \in A\}$$

for a positive scalar λ, then the new length threshold becomes $\lambda\ell_0$. On the other hand, the upper density of the new set remains unchanged since

$$\begin{aligned}\limsup_{R\to\infty} \frac{|\lambda A \cap B(0,R)|}{|B(0,R)|} &= \limsup_{R\to\infty} \frac{|\lambda A \cap B(0,\lambda R)|}{|B(0,\lambda R)|} \\ &= \limsup_{R\to\infty} \frac{|\lambda|^2 |A \cap B(0,R)|}{|\lambda|^2 |B(0,R)|} \\ &= \delta.\end{aligned}$$

Thus at first glace it may seem as if Theorem (4.1.2) is irredeemably qualitative in nature. Nevertheless, Bourgain managed to give a new proof of this theorem by first establishing the following quantitative analogue:

Theorem 4.1.3. (Furstenberg-Katznelson-Weiss theorem, quantitative version). Let $0 < \epsilon < \frac{1}{2}$, let $B \subseteq [-1,1]^2$ have measure $|B| \geq \epsilon$ and let $J = J(\epsilon)$ be a sufficiently large natural number depending on ϵ. Suppose that $0 < t_J < \cdots < t_1 \leq 1$ is a sequence of scales with $t_{j+1} \leq t_j/2$ for all $1 \leq j < J$. Then for at least one scale $t_j \in \{t_1, \cdots, t_J\}$, we have

$$\int_{\mathbb{R}^2} \int_{S^1} \mathbb{1}_B(x) \mathbb{1}_B(x + t_j\omega)\ d\sigma(\omega) dx \gtrsim \epsilon^2 \tag{33}$$

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

Observe that there is no longer any mention of qualitative aspects such as limits. Instead, one now works with a finite number of scales $t_1, \cdots, t_J$. Furthermore, establishing an upper bound for J in terms of ϵ is now possible. As we shall see in the next section, the proof provided by Bourgain gave an explicit dependence $J = O(\frac{1}{\epsilon}\log\frac{1}{\epsilon})$.

The key observation that connects Theorem 4.1.3 to Theorem 4.1.2 is that if the integral (33) is non-vanishing for the scale t_j, then there exists $x \in \mathbb{R}^2$ such that the inner integral

$$\int_{S^1} \mathbb{1}_B(x) \mathbb{1}_B(x + t_j\omega)\ d\sigma(\omega)$$

is also non-vanishing. This implies the existence of $\omega \in S^1$ such that both x and $x + t_j\omega$ lie in B. We have hence found our two distinct points of distance t_j apart. The bulk of the remainder of the proof then involves rescaling the set A to obtain a suitable set B.

More formally, we argue by contradiction as follows. Suppose that Theorem 4.1.2 failed. Then there exists a measurable subset $A \subseteq \mathbb{R}^2$ having positive upper density δ and an increasing sequence of scales $\ell_1 < \ell_2 < \cdots$ going to infinty such that for each ℓ_j, there exists no $x, y \in A$ such that $|x - y| = \ell_j$. We now proceed to "align" the hypotheses of Theorem 4.1.2 with that of Theorem 4.1.3.

Firstly, we sparsify the sequence of scales so that $\ell_{j+1} \geq 2\ell_j$ for all j. Fix a sufficiently small constant ϵ (whose value depends only on δ) and let J be as in Theorem 4.1.3. As A has density δ, there exists a radius $R \in (\ell_J, \ell_{J+1})$ such that $|A \cap B(0,R)| \gtrsim \delta R^2$.

We now rescale the set A. Define

$$B := \{x \in [-1,1]^2 \ : \ Rx \in A\} = \frac{1}{R} \cdot A \cap [-1,1]^2$$

and define the sequence $0 < t_J < \cdots < t_1 \leq 1$ by

$$t_j := \frac{\ell_{J+1-j}}{R}, \quad j = 1, 2, \cdots, J.$$

Then $|B| \gtrsim \delta$ and by choosing $\epsilon \sim \delta$, we have $|B| \geq \epsilon$. Furthermore, $t_{j+1} \leq t_j/2$ for all $1 \leq j < J$. Thus the hypotheses of Theorem 4.1.3 are now satisfied.

But by construction, for all scales $t_j \in \{t_1, \cdots, t_J\}$, there exists no pair of distinct points $x, y \in B$ such that $|x - y| = t_j$. This is because if such a pair were to exist, then $Rx, Ry \in A$ and $|Rx - Ry| = \ell_{J+1-j}$. Thus the integral (33) is identically zero and we have obtained the desired contradiction.

Example 4.1.4. We postpone the proof of Theorem 4.1.3 to the next section and instead look at another example of how quantitative methods can be used to establish qualitative results.

Definition 4.1.5. Let (X, μ) be a probability space. A measurable function $T : X \to X$ is **measure-preserving** if for all measurable subsets A of X, we have

$$\mu\big(T^{-1}(A)\big) = \mu(A).$$

Theorem 4.1.6. (Bourgain's pointwise ergodic theorem along squares). Let (X, μ) be a probability space with a measure-preserving transformation $T : X \to X$. Then for any $r \in \mathbb{R}_{>1}$ and for any $f \in L^r(X, \mu)$, the averages

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

converge almost surely as $N \to \infty$.

Using standard arguments in ergodic theory, Bourgain was able to reduce the above problem to the case where f lies in the dense subclass $L^\infty(X, \mu)$ of $L^r(X, \mu)$. However, new techniques were needed to handle the problem for this subclass. This is achieved via the following variational estimate:

Theorem 4.1.7. (Bourgain's variational estimate) Let $\lambda > 1$, let $\epsilon > 0$ and let J be a sufficiently large integer depending on λ and ϵ. Then for any $1 < N_1 < N_2 < \cdots < N_J$, any (X, μ, T) as in Theorem 4.1.6 and any $f \in L^\infty(X, \mu)$, we have the estimate

$$\sum_{j=1}^{J-1} \left\| \sup_{N_j \leq N \leq N_{j+1} : N \in Z_\lambda} \left| A_N f - A_{N_j} f \right| \right\|_{L^2(X,\mu)} \lesssim \epsilon J \, \|f\|_{L^2(X,\mu)}, \tag{34}$$

where $Z_\lambda := \{\lfloor \lambda^n \rfloor : n \in \mathbb{Z}_{\geq 0}\}$.

We now show why Theorem 4.1.7 implies Theorem 4.1.6. We may take $f \in L^\infty(X,\mu)$ and normalise $\|f\|_{L^\infty(X,\mu)} = 1$. By discarding a subset of measure 0, we may further assume that $|f| \leq 1$. We first argue that it suffices to show almost sure convergence for the subsequence $(A_N f)_{N\in Z_\lambda}$ instead.

Lemma 4.1.8. Suppose that for all $\lambda \in \mathbb{R}_{>1}$, the subsequence $(A_N f)_{N\in Z_\lambda}$ converges almost surely. Then the original sequence $(A_N f)_{N=1}^\infty$ also converges almost surely.

Proof. The set S defined by

$$S := \bigcap_{k=1}^{\infty} \left\{ x \in X \ : \ \big(A_N f(x)\big)_{N\in Z_\lambda} \text{ converges for } \lambda = 1 + \frac{1}{k} \right\}$$

has probability 1. We will show that $(A_N(f))_{N=1}^\infty$ converges pointwise on S. Suppose not. There exists $y \in S$ such that the sequence $(A_N f(y))_{N=1}^\infty$ is not Cauchy. This means that we can find $\epsilon \in \mathbb{R}_{>0}$ such that for all $m \in \mathbb{Z}_{\geq 1}$, there exists $m' \in \mathbb{Z}_{>m}$ such that

$$\big|A_m f(y) - A_{m'} f(y)\big| > \epsilon. \tag{35}$$

Choose $\lambda = 1 + \dfrac{1}{k}$ with k large enough so that $\dfrac{6}{k} < \epsilon$.

Round up m and m' to the nearest integers $N_m, N_{m'} \in Z_\lambda$ respectively. Then

$$\begin{aligned}
\big|A_{N_m} f(y) - A_m f(y)\big| &= \left| \frac{1}{N_m} \sum_{n=1}^{N_m} f\big(T^{n^2} y\big) - \frac{1}{m} \sum_{n=1}^{m} f\big(T^{n^2} y\big) \right| \\
&= \left| \frac{1}{N_m} \sum_{n=m+1}^{N_m} f(T^{n^2} y) + \frac{m - N_m}{mN_m} \sum_{n=1}^{m} f(T^{n^2} y) \right| \\
&\leq \frac{1}{N_m} \sum_{n=m+1}^{N_m} \big|f(T^{n^2} y)\big| + \frac{N_m - m}{mN_m} \sum_{n=1}^{m} \big|f(T^{n^2} y)\big| \\
&\leq \frac{N_m - m}{N_m} + \frac{N_m - m}{N_m} \\
&\leq \frac{2(\lambda m - m)}{N_m} \leq 2(\lambda - 1) < \frac{\epsilon}{3}.
\end{aligned}$$

where we have used the fact that $|f| \leq 1$ on S and $N_m \leq \lambda m$. Similarly, we have the bound $|A_{N_{m'}} f(y) - A_{m'}(y)| < \epsilon/3$. Finally, by choosing m sufficiently large so that $|A_{N_m} f(y) - A_{N_{m'}}(y)| < \epsilon/3$, we have that

$$\begin{aligned}
&|A_m f(y) - A_{m'} f(y)| \\
&\leq |A_m f(y) - A_{N_m} f(y)| + |A_{N_m} f(y) - A_{N_{m'}} f(y)| + |A_{N_{m'}} f(y) - A_{m'} f(y)| \\
&< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon
\end{aligned}$$

which is a contradiction to (35). This completes the proof of the lemma.

With the lemma in hand, instead of showing almost sure convergence for the original sequence $(A_N f(y))_{N=1}^{\infty}$, it suffices to show almost sure convergence for the subsequence $(A_N f)_{N\in Z_\lambda}$ instead for an arbitrarily chosen $\lambda > 1$. Assume that almost sure convergence for this subsequence fails for some λ. Define the set

$$\Omega := \left\{ x \in X \ : \ \limsup_{N\in Z_\lambda} A_N f(x) = \liminf_{N\in Z_\lambda} A_N f(x) \right\}.$$

Then Ω^c has positive measure. By writing

$$\Omega^c = \bigcup_{n=1}^{\infty} \left\{ x \in X \ : \ \limsup_{N\in Z_\lambda} A_N f(x) - \liminf_{N\in Z_\lambda} A_N f(x) \geq \frac{1}{n} \right\},$$

we see that there exists a set E of positive measure δ and a positive number δ' such that

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

for all $x \in E$. We now inductively construct an increasing sequence of integers $(N_k)_{k=1}^{\infty} \subseteq Z_\lambda$ and a sequence of decreasing subsets $E = E_1 \supseteq E_2 \supseteq \cdots$ such that the following properties are satisfied:

(i) For all positive integers k, we have $\mu(E_{k+1}) \geq \mu(E_k) - \dfrac{\delta}{2^{k+1}}$;

(ii) For all positive integers k and $x \in E_{k+1}$, we have

$$\sup_{N_k \leq N \leq N_{k+1} : N\in Z_\lambda} \left| A_N f(x) - A_{N_k} f(x) \right| \geq \frac{\delta'}{2}.$$

For the base step, simply define $N_1 = 1$ and $E_1 = E$.

In the inductive step, suppose that N_k and E_k has been defined. Since E_k is a subset of E, we may use triangle inequality to write

$$E_k = \bigcup_{N=N_k+1}^{\infty} \left\{ x \in E_k \ : \ \sup_{N_k \leq j \leq N : j\in Z_\lambda} \left| A_j f(x) - A_{N_k} f(x) \right| \geq \frac{\delta'}{2} \right\}.$$

Thus there exists $Z \in \mathbb{Z}_{\geq N_k+1}$ such that

$$\mu\left(\bigcup_{N=N_k+1}^{Z} \left\{ x \in E_k \ : \ \sup_{N_k \leq j \leq N : j\in Z_\lambda} \left| A_j f(x) - A_{N_k} f(x) \right| \geq \frac{\delta'}{2} \right\} \right) \geq \mu(E_k) - \frac{\delta}{2^{k+1}}.$$

Define N_{k+1} to be the least integer in Z_λ that is greater than Z. Thus condition (i) is maintained.

Define

$$E_{k+1} := \left\{ x \in E_k \ : \ \sup_{N_k \le N \le N_{k+1} : N \in Z_\lambda} \left| A_N f(x) - A_{N_k} f(x) \right| \ge \frac{\delta'}{2} \right\}.$$

Then property (ii) is maintained. This completes the induction step.

Finally, if we take

$$\widetilde{E} := \bigcap_{k=1}^{\infty} E_k,$$

then property (i) forces $\mu(\widetilde{E}) \ge \delta/2 > 0$. Furthermore, for all $k \in \mathbb{Z}_{\ge 1}$ and $x \in \widetilde{E}$,

$$\sup_{N_k \le N \le N_{k+1} : N \in Z_\lambda} \left| A_N f(x) - A_{N_k} f(x) \right| \ge \frac{\delta'}{2}$$

by property (ii).

In particular, by considering the restriction of the integral to $\widetilde{E}$, the left-hand side of (34) enjoys the lower bound

$$\sum_{j=1}^{J-1} \left\| \sup_{N_j \le N \le N_{j+1} : N \in Z_\lambda} \left| A_N f - A_{N_j} f \right| \right\|_{L^2(X,\mu)} \ge \sum_{j=1}^{J-1} \frac{\delta'}{2} \cdot \mu(\widetilde{E})^{1/2} \ge \frac{J \delta^{1/2} \delta'}{4}$$

while the right hand side is ϵJ. Our desired contradiction can then be obtained by choosing ϵ sufficiently small (the dependence of J on ϵ can be mitigated due to cancellation on both sides).

4.2 Dyadic Pigeonholing

A classical technique in analysis is that of dyadic decomposition, where one partitions the domain of interest into a number of smaller segments that ranges between two consecutive powers 2^k and 2^{k+1} of two. Bourgain took this technique one step further by combining it with the pigeonhole principle in order to locate a single "good" segment in which to run additional arguments. The combination of dyadic decomposition with the pigeonhole principle is known as **dyadic pigeonholing**.

Example 4.2.1. As a first example of this technique, we will sketch Bourgain's proof of Theorem 4.1.3. The theorem is restated here for convenience.

(Furstenberg-Katznelson-Weiss theorem, quantitative version). Let $0 < \epsilon < \frac{1}{2}$, let $B \subseteq [-1,1]^2$ have measure $|B| \geq \epsilon$ and let $J = J(\epsilon)$ be a sufficiently large natural number depending on ϵ. Suppose that $0 < t_J < \cdots < t_1 \leq 1$ is a sequence of scales with $t_{j+1} \leq t_j/2$ for all $1 \leq j < J$. Then for at least one scale $t_j \in \{t_1, \cdots, t_J\}$, we have

$$\int_{\mathbb{R}^2} \int_{S^1} \mathbb{1}_B(x)\mathbb{1}_B(x + t_j\omega)\ d\sigma(\omega)dx \gtrsim \epsilon^2. \tag{36}$$

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

We will adopt the following definition of the Fourier transform:

$$\widehat{f}(\xi) := \int_{\mathbb{R}^n} f(x)e^{-2\pi i x\cdot\xi}\ dx.$$

By Fubini-Tonelli's theorem and Plancherel's theorem, we have

$$\begin{aligned}
\int_{\mathbb{R}^2} \int_{S^1} \mathbb{1}_B(x)\mathbb{1}_B(x + t_j\omega)\ d\sigma(\omega)dx &= \int_{S^1} \int_{\mathbb{R}^2} \mathbb{1}_B(x)\mathbb{1}_B(x + t_j\omega)\ dxd\sigma(\omega) \\
&= \int_{S^1} \int_{\mathbb{R}^2} |\widehat{\mathbb{1}}_B(\xi)|^2 e^{2\pi i t_j\omega\cdot\xi}\ d\xi d\sigma(\omega) \\
&= \int_{\mathbb{R}^2} |\widehat{\mathbb{1}}_B(\xi)|^2 \widehat{\sigma}(t_j\xi)\ d\xi
\end{aligned}$$

where $\widehat{\sigma}(\xi) := \displaystyle\int_{S^1} e^{2\pi i\omega\cdot\xi}\ d\sigma(\omega)$. Note that $\widehat{\sigma}$ is continuous in ξ and purely real since

$$\begin{aligned}
\operatorname{Im}\left(\int_{S^1} e^{2\pi i\omega\cdot\xi}\ d\sigma(\omega)\right) &= \operatorname{Im}\left(\frac{1}{2\pi}\int_0^{2\pi} e^{2\pi i\|\xi\|\cos\theta}\ d\theta\right) \\
&= \frac{1}{2\pi}\int_0^{2\pi} \sin\big(2\pi\,\|\xi\|\cos\theta\big)\ d\theta = 0.
\end{aligned}$$

For each fixed scale t_j, we decompose this integral into

$$\int_{|\xi|\leq\frac{\delta}{t_j}} |\widehat{\mathbb{1}}_B(\xi)|^2\widehat{\sigma}(t_j\xi)\ d\xi + \int_{\frac{\delta}{t_j}\leq|\xi|\leq\frac{1}{\delta t_j}} |\widehat{\mathbb{1}}_B(\xi)|^2\widehat{\sigma}(t_j\xi)\ d\xi + \int_{\frac{1}{\delta t_j}\leq|\xi|} |\widehat{\mathbb{1}}_B(\xi)|^2\widehat{\sigma}(t_j\xi)\ d\xi$$

where $\delta > 0$ is a small constant whose exact value will be determined later.

Our strategy will be to show that the first term, which is the contribution of the low frequencies, achieves the lower bound of $\gtrsim \epsilon^2$ while the contribution from the other two terms are "negligible" (hence preventing unwanted cancellation).

Using the continuity of $\widehat{\sigma}$, we may choose δ sufficiently small so that $\widehat{\sigma}(t_j\xi) \gtrsim 1$ for all $|\xi| \leq \delta/t_j$. Then

$$\int_{|\xi|\leq\frac{\delta}{t_j}} |\widehat{\mathbb{1}}_B(\xi)|^2\widehat{\sigma}(t_j\xi)\ d\xi \gtrsim \int_{|\xi|\leq\frac{\delta}{t_j}} |\widehat{\mathbb{1}}_B(\xi)|^2\ d\xi \geq \int_{|\xi|\leq\delta} |\widehat{\mathbb{1}}_B(\xi)|^2\ d\xi.$$

By a direct computation,

$$\widehat{\mathbb{1}}_B(\xi) = \int_{\mathbb{R}^2} \mathbb{1}_B(x)e^{-2\pi ix\cdot\xi}\ dx = \int_B e^{-2\pi ix\cdot\xi}\ dx.$$

Since $\widehat{\mathbb{1}}_B(0) = |B|$, by continuity of $\widehat{\mathbb{1}}_B(\xi)$, we have

$$|\widehat{\mathbb{1}}_B(\xi)|^2 \gtrsim |B|^2$$

for $|\xi| \leq \delta$ (again if δ is small enough). Thus the contribution of the low frequencies satisfy

$$\int_{|\xi|\leq\frac{\delta}{t_j}} |\widehat{\mathbb{1}}_B(\xi)|^2\widehat{\sigma}(t_j\xi)\ d\xi \gtrsim \delta^2|B|^2 \geq \delta^2\epsilon^2.$$

For the last term, which is the contribution of the high frequencies, the factor $\widehat{\sigma}$ is small. Using Fourier-theoretic arguments, it can be shown that the last term decays faster as $\delta \to 0$ as compared to the first term. Hence contribution is negligible compared to the low frequencies

The middle term, however, is particularly tricky. Naively, one can use the Plancherel's theorem to upper bound the middle term by

$$\int_{\frac{\delta}{t_j}\leq|\xi|\leq\frac{1}{\delta t_j}} |\widehat{\mathbb{1}}_B(\xi)|^2\widehat{\sigma}(t_j\xi)\ d\xi \leq \int_{\mathbb{R}^2} |\widehat{\mathbb{1}}_B(\xi)|^2\ d\xi = \int_{\mathbb{R}^2} |\mathbb{1}_B(x)|^2\ dx = |B|$$

However, this is far too weak as compared to the lower bound of $\gtrsim |B|^2$ for the low frequencies. Thankfully, because of the lacunarity hypothesis $t_{j+1} \leq t_j/2$, the annuli

$$A(t_j) := \left\{\xi \in \mathbb{R}^2\ :\ \frac{\delta}{t_j} \leq |\xi| \leq \frac{1}{\delta t_j}\right\}$$

overlap with multiplicity at most $2\log_2 1/\delta + 1$. To see this, first fix a point $\xi_0 \in \mathbb{R}^2$ and let $\{A(t_j), A(t_{j+1}), \cdots, A(t_{j+\ell})\}$ denote the complete set of annuli that contains ξ_0. We have

$$\frac{\delta}{t_{j+\ell}} \leq |\xi_0| \leq \frac{1}{\delta t_j}.$$

Since $t_{j+\ell} \leq \frac{t_j}{2^\ell}$, we obtain

$$2^\ell \delta^2 \leq 1 \implies \ell \leq 2 \log_2 \frac{1}{\delta}.$$

Control over the overlap means that if we now sum over all possible scales $t_1, \cdots, t_J$, we get

$$\sum_{j=1}^{J} \int_{\frac{\delta}{t_j} \leq |\xi| \leq \frac{1}{\delta t_j}} |\widehat{\mathbb{1}}_B(\xi)|^2 \widehat{\sigma}(t_j \xi)\ d\xi \lesssim |B| \log_2 \frac{1}{\delta}. \tag{37}$$

Crucially, the bound (37) does not depend on J. By the pigeonhole principle, there exists a scale t_j for which

$$\int_{\frac{\delta}{t_j} \leq |\xi| \leq \frac{1}{\delta t_j}} |\widehat{\mathbb{1}}_B(\xi)|^2 \widehat{\sigma}(t_j \xi)\ d\xi \lesssim \frac{|B|}{J} \log_2 \frac{1}{\delta}.$$

Hence if J is sufficiently large, the contribution of the middle term becomes negligible as well and we obtain our desired conclusion.

Example 4.2.2. Another example of Dyadic pigeonholing occurs in Bourgain's study of a variant of the Kakeya needle problem, which asks for how "small" Besicovitch sets (subsets of Euclidean space $\mathbb{R}^n$ that contains a unit line segment in every direction) can be. One metric that can be used to measure the "size" of such sets is the Hausdorff dimension.

Definition 4.2.3. Let $\mathbb{R}^n$ denote the usual n-dimensional Euclidean space. For any $\delta > 0$ and $S \subseteq \mathbb{R}^n$, define the **Hausdorff content** of S by

$$H_\delta^n(S) := \inf \left\{ \sum_{i=1}^{\infty} (\text{diam } U_i)^n : S \subseteq \bigcup_{i=1}^{\infty} U_i,\ \text{diam}(U_i) < \delta \right\}$$

where the infimum is taken over all countable covers U_i of S. Define

$$H^n(S) := \lim_{\delta \to 0} H_\delta^n(S).$$

The map H^n is an outer measure and its restriction to the σ-field of Carathéodory-measurable sets (via the Carathéodory's extension theorem) is the ***n*-dimensional Hausdorff measure**.

In particular, Bourgain's work involves establishing lower bounds on the Hausdorff dimension of Besicovitch sets. To lower bound the Hausdorff dimension of a Besicovitch set E by some $\alpha \in \mathbb{R}$, in one of the intermediate steps, he first attempts to establish a lower bound, for all $\epsilon > 0$, on the Hausdorff content

$$\sum_{i \in I} r_i^{\alpha - \epsilon}$$

whenever one covers the set E by small balls $B(x_i, r_i)$. By rounding up if necessary, we may assume that each r_i is a power of two. Furthermore, we may replace the balls of radius r_i by cubes of sidelength r_i instead (since their measures differ by a uniform constant). By grouping together cubes of a given size, we obtain a covering

$$E \subseteq \bigcup_{k=k_0}^{\infty} B_k$$

where k_0 is a large integer and B_k is the union of cubes of sidelength 2^{-k}. To get the required lower bound, Bourgain aims to show that there exists at least one scale k such that the number of cubes used to form B_k is $\gtrsim 2^{k(\alpha-\epsilon)}$.

To find the correct scale, the technique of dyadic decomposition is applied as follows. For each direction $\omega \in S^{n-1}$, the set E contains a unit line segment ℓ_ω in that direction. As the union of the B_k's cover E, they also contain each ℓ_ω. Thus we have

$$1 = \mathcal{H}^1(\ell_\omega) = \mathcal{H}^1(\ell_\omega \cap E) \leq \sum_{k=k_0}^{\infty} \mathcal{H}^1(\ell_\omega \cap B_k).$$

By integrating over all directions, we have that

$$1 \sim \int_{S^{n-1}} 1 \; d\omega \leq \int_{S^{n-1}} \sum_{k=k_0}^{\infty} \mathcal{H}^1(\ell_\omega \cap B_k) \; d\omega = \sum_{k=k_0}^{\infty} \left[\int_{S^{n-1}} \mathcal{H}^1(\ell_\omega \cap B_k) \; d\omega \right]$$

where the non-negativity of the function is used to justify the swapping of the order of integration and summation by Fubini's theorem. Since the above sum is bounded below by a constant, we may compare it a convergent series (say $\sum_{k=k_0}^{\infty} 1/k^2$) to get

$$\sum_{k=k_0}^{\infty} \left[\int_{S^{n-1}} \mathcal{H}^1(\ell_\omega \cap B_k) \; d\omega \right] \gtrsim \sum_{k=k_0}^{\infty} \frac{1}{k^2}.$$

By dyadic pigeonholing, there exists at least one scale k such that

$$\int_{S^{n-1}} \mathcal{H}^1(\ell_\omega \cap B_k) \; d\omega \gtrsim \frac{1}{k^2}.$$

Using this scale k, Bourgain was able to proceed to establish new bounds on α by estimating an expression now known as the Kakeya maximal function.

Example 4.2.4. In some cases, Bourgain used more sophisticated tools than the pigeonhole principle to locate a good scale. One such example is in the proof of the following theorem in additive combinatorics.

Theorem 4.2.5. (Bourgain-Roth theorem) Let $N \in \mathbb{Z}_{\geq 10}$ and let $A \subseteq \{1, \cdots, N\}$ 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 the paper is the introduction of **Bohr sets**, now a standard tool in the field of additive combinatorics.

Definition 4.2.6. Fix $N \in \mathbb{Z}_{\geq 1}$. Let $Z = \{0, 1, \cdots, N-1\}$. For $k \in \mathbb{R}$, define

$$\|k\|_{\mathbb{R}/\mathbb{Z}} := \min_{z \in \mathbb{Z}} \big|k + z\big|.$$

Let $S \subseteq \mathbb{R}/\mathbb{Z}$ be a finite set of frequencies and let $r \in [0, 1]$ be a "radius". Define the **Bohr set** $B_N(S, r)$ by

$$B_N(S, r) := \big\{n \in Z \ : \ \|n\xi\|_{\mathbb{R}/\mathbb{Z}} < r \text{ for all } \xi \in S\big\}.$$

Remark 4.2.7. An immediate consequence of the finiteness of S is that $|B_N(S, r)|$ is left-continuous in the variable r.

Remark 4.2.8. If $r > \dfrac{1}{2}$, then trivially $B_N(S, r) = Z$.

We first derive some elementary bounds on the cardinality of Bohr sets.

Proposition 4.2.9. (Size estimate)

$$|B_N(S, r)| \geq r^{|S|} N.$$

Proof. We use the probabilistic method. For each $\xi \in S$, draw an element θ_ξ from $\mathbb{R}/\mathbb{Z}$ independently and uniformly at random. Then for each integer $n \in Z$, we have that

$$P\left(\|\xi \cdot n - \theta_\xi\|_{\mathbb{R}/\mathbb{Z}} < r/2\right) = r.$$

By independence,

$$P\left(\|\xi \cdot n - \theta_\xi\|_{\mathbb{R}/\mathbb{Z}} < r/2 \ \text{ for all } \xi \in S\right) = r^{|S|}.$$

Summing over all integers n gives

$$\mathbb{E}\bigg[\Big|\big\{n \in Z; \ \|\xi \cdot n - \theta_\xi\|_{\mathbb{R}/\mathbb{Z}} < r/2 \text{ for all } \xi \in S\big\}\Big|\bigg] = r^{|S|} N.$$

Hence there exists a choice $\{\theta'_\xi \ : \ \xi \in S\}$ such that

$$\Big|\big\{n \in Z; \ \big\|\xi \cdot n - \theta'_\xi\big\|_{\mathbb{R}/\mathbb{Z}} < r/2 \text{ for all } \xi \in S\big\}\Big| \geq r^{|S|} N. \tag{38}$$

Note that if two integers n_1 and n_2 lie in the above set, then their difference $n_1 - n_2$ lies in the set $B_N(S, r)$. In other words, if the elements of the above set is $\{n_1, n_2, \cdots, n_k\}$, then

$$\{0, n_1 - n_2, n_1 - n_3, \cdots, n_1 - n_k\} \subseteq B_N(S, r) \tag{39}$$

where addition is performed modulo N. It then follows that $|B_N(S, r)| \geq r^{|S|} N$.

Proposition 4.2.10. (Doubling estimate)

$$|B_N(S,2r)| \leq 4^{|S|}|B_N(S,r)|.$$

By left continuity of the map $r \mapsto |B_N(S,r)|$, it suffices to show that

$$|B_N(S,2r-\epsilon)| \leq 4^{|S|}|B_N(S,r)|$$

for all $\epsilon > 0$. The ϵ-margin allows us to cover the set

$$\left\{\theta \in \mathbb{R}/\mathbb{Z} \ : \ \|\theta\|_{\mathbb{R}/\mathbb{Z}} < 2r-\epsilon\right\}$$

by 4 sets of the form

$$\left\{\theta \in \mathbb{R}/\mathbb{Z} \ : \ \|\theta-\theta_j\|_{\mathbb{R}/\mathbb{Z}} < r/2\right\}.$$

for appropriately chosen points $\theta_1, \theta_2, \theta_3, \theta_4 \in \mathbb{R}/\mathbb{Z}$.

First suppose that $S = \{\xi\}$ consists of just a single element. Then

$$\begin{aligned}
\left|B_N(S,2r-\epsilon)\right| &= \left|\{n \in Z \ : \ \|n\xi\|_{\mathbb{R}/\mathbb{Z}} < 2r-\epsilon\}\right| \\
&\leq \left|\bigcup_{j=1}^{4}\left\{n \in Z \ : \ \|n\xi-\theta_j\|_{\mathbb{R}/\mathbb{Z}} < r/2\right\}\right| \\
&\leq \sum_{j=1}^{4}\left|\left\{n \in Z \ : \ \|n\xi-\theta_j\|_{\mathbb{R}/\mathbb{Z}} < r/2\right\}\right| \\
&= 4|B_N(S,r)|
\end{aligned}$$

where the last line follows from a similar argument as in (39) in the previous proposition.

We now generalise the argument to the case $S = \{\xi_1, \cdots, \xi_k\}$. Observe that

$$\begin{aligned}
\left|B_N(S,2r-\epsilon)\right| &= \left|\bigcap_{i=1}^{k}\{n \in Z \ : \ \|n\xi_i\|_{\mathbb{R}/\mathbb{Z}} < 2r-\epsilon\}\right| \\
&\leq \left|\bigcap_{i=1}^{k}\bigcup_{j=1}^{4}\left\{n \in Z \ : \ \|n\xi_i-\theta_j\|_{\mathbb{R}/\mathbb{Z}} < r/2\right\}\right|.
\end{aligned}$$

By applying the distributive law, we have

$$\begin{aligned}
\left|B_N(S,2r-\epsilon)\right| &\leq \left|\bigcup_{1\leq j_1\cdots,j_k\leq 4}\bigcap_{i=1}^{k}\left\{n \in Z \ : \ \|n\xi_i-\theta_{j_i}\|_{\mathbb{R}/\mathbb{Z}} < r/2\right\}\right| \\
&\leq \sum_{1\leq j_1,\cdots,j_k\leq 4}\left|\bigcap_{i=1}^{k}\left\{n \in Z \ : \ \|n\xi_i-\theta_{j_i}\|_{\mathbb{R}/\mathbb{Z}} < r/2\right\}\right| \\
&\leq 4^{|S|}|B_N(S,r)|
\end{aligned}$$

as desired.

However, a key difficulty arises from the fact that Bohr sets are in general not right-continuous in r. Thus the notion of regular Bohr sets was introduced.

Definition 4.2.11. The Bohr set $B_N(S,r)$ is **regular** if there exists a constant C such that for all $r' > 0$, we have the upper bound

$$\left|B_N(S,r')\right| \leq \exp\left(C \cdot |S| \frac{|r'-r|}{|r|}\right) \left|B_N(S,r)\right|. \tag{40}$$

As it turns out, every Bohr set is "close" to a regular one. The next theorem makes the study of Bohr sets especially well suited to dyadic pigeonholing arguments.

Theorem 4.2.12. Let $S \subseteq \mathbb{R}/\mathbb{Z}$ be a finite set and let $0 < \epsilon < 1$. Then there exists $r \in [\epsilon, 2\epsilon]$ such that $B_N(S,r)$ is regular.

Proof. Note that (40) is equivalent to the condition that

$$\frac{1}{|S|} \log_2 \frac{|B_N(S,r')|}{|B_N(S,r)|} \leq \widetilde{C} \cdot \frac{|r'-r|}{|r|} \tag{41}$$

for some constant $\widetilde{C}$.

We will first show that the doubling estimate (Proposition 4.2.10) implies that Bohr sets already enjoy a certain level of regularity when the distance $|r'-r|$ is large. Assume without loss of generality that $r' > r$ and $r'/r \geq 2^{0.1}$ (the choice of 0.1 is arbitrary).

Let $k \in \mathbb{Z}_{\geq 1}$ be such that $r' \in [2^{k-1}r, 2^k r)$. Then

$$\frac{1}{|S|} \log_2 \frac{|B_N(S,r')|}{|B_N(S,r)|} \leq \frac{1}{|S|} \log_2 \frac{|B_N(S,2^k r)|}{|B_N(S,r)|} \leq \frac{1}{|S|} \log_2 4^{k|S|} = 2k.$$

If $k=1$, then we simply choose $\widetilde{C} \in \mathbb{R}_{\geq 8}$ (in fact the explicit choice of $\widetilde{C} = 40$ later also works here) large enough to get the bound

$$2(1) \leq \widetilde{C}(2^{0.1}-1) = \frac{\widetilde{C}|2^{0.1}r - r|}{|r|} \leq \frac{\widetilde{C}|r'-r|}{|r|}$$

directly and if $k \geq 2$, then

$$2k \leq 8(2^{k-1}-1) \leq \frac{\widetilde{C}|2^{k-1}r - r|}{|r|} \leq \frac{\widetilde{C}|r'-r|}{|r|}.$$

We now tackle the case in which $2^{-0.1} \leq \left|\frac{r'}{r}\right| \leq 2^{0.1}$. Define $f : [0,1] \to \mathbb{R}$ by

$$f(a) = \frac{1}{|S|} \log_2 \left|B_N(S, 2^a \epsilon)\right|.$$

Since f is monotone, it is measurable. We aim to find $a \in [0.1, 0.9]$ such that (41) holds for all $r' > 0$ satisfying

$$2^{-0.1} < \left| \frac{r'}{2^a \epsilon} \right| < 2^{0.1}. \tag{42}$$

The conclusion then follows by choosing $r = 2^a \epsilon$. By writing $r' = 2^{\widetilde{a}} \epsilon$ for some $\widetilde{a} \in \mathbb{R}$, we see that (42) is equivalent to $|\widetilde{a} - a| < 0.1$, Furthermore,

$$\frac{|r' - r|}{|r|} = \frac{|2^{\widetilde{a}} - 2^a|}{|2^a|} \leq |2^{\widetilde{a}-a} - 1| \leq |\widetilde{a} - a|.$$

where we have used the condition $|\widetilde{a} - a| < 0.1$ to obtain the last inequality.

To summarise, it suffices to show that there exists $a \in [0.1, 0.9]$ such that there exists a constant $\widetilde{C}$ so that we have the Lipschitz property

$$|f(\widetilde{a}) - f(a)| \leq \widetilde{C} |\widetilde{a} - a|$$

for any $\widetilde{a}$ with $|\widetilde{a} - a| < 0.1$. We claim that the existence of such an a is guaranteed for $\widetilde{C} = 40$. Suppose not. Then such a point a cannot be found for $\widetilde{C} = 40$. This means that for each $a \in [0.1, 0.9]$, there exists an interval I_a of length at most 0.1, either of the form $[a', a]$ or $[a, a']$ where $a' \in [0, 1]$, such that

$$|f(a') - f(a)| > 40|a - a'|.$$

These intervals cover $[0.1, 0.9]$, which has measure 0.8. By the Vitali covering lemma, there exists a countable subcollection of disjoint intervals $(I_k)_{k \in J}$ such that

$$\sum_{k \in J} |I_k| \geq 0.8/5.$$

But then by the doubling estimate, we have

$$2 \geq \int_0^1 1 \; df(x) \geq \sum_{k \in J} \int_{I_k} 1 \; df(x) \geq \sum_{k \in J} \int_{I_k} 40 \; dx \geq \frac{0.8}{5} \times 40 = 6.4$$

which is a contradiction.

4.3 Random Translations

Consider the interval $I = [0, 1/N]$ in the unit circle $\mathbb{R}/\mathbb{Z}$. If N is an integer, then we can cover $\mathbb{R}/\mathbb{Z}$ by N copies of I by considering the translates $I + j/N$ for $j = \{0, \cdots, N-1\}$. Unfortunately, most subsets E of $\mathbb{R}/\mathbb{Z}$ do not have this perfect tiling property. However, by using random translations of E, we can achieve something fairly close to a perfect tiling.

Lemma 4.3.1. (Random translations) Let G be a compact group and let μ denote its normalised Haar (probability) measure. Let E be a measurable subset of G and let N be a positive integer. Then there exist $g_1, \cdots, g_N \in G$ such that the translates $g_1 E, \cdots, g_n E$ of E satisfy

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

Proof. We use the probabilistic method. If $g_1, \cdots, g_N \in G$ are drawn independently and uniformly at random from G using the probability measure μ, then by Fubini-Tonelli theorem, we have

$$\mathbb{E}\big[\mu(g_1 E \cup \cdots \cup g_N E)\big] = \int_G \mathbb{E}\big[\mathbb{1}_{g_1 E \cup \cdots \cup g_N E}(x)\big]\, d\mu(x).$$

Since $g_1, \cdots, g_N$ are i.i.d uniform, we have that

$$\begin{aligned}
\int_G \mathbb{E}\big[\mathbb{1}_{g_1 E \cup \cdots \cup g_N E}(x)\big]\, d\mu(x) &= \int_G \mathbb{E}\Big[1 - \prod_{i=1}^{N} \mathbb{1}_{g_i(G \setminus E)}(x)\Big]\, d\mu(x) \\
&= \int_G 1 - \prod_{i=1}^{N} \mathbb{E}\Big[\mathbb{1}_{g_i(G \setminus E)}(x)\Big]\, d\mu(x) \\
&= \int_G 1 - \prod_{i=1}^{N} \mu(G \setminus E)\ d\mu(x) \\
&= 1 - (1 - \mu(E))^N.
\end{aligned}$$

This completes the proof.

Lemma 4.3.1 allows us to reduce the analysis of "small" sets to the analysis of "large" sets by covering the entire space (usually assumed to be compact) with random translates of the small sets. This technique works particularly well when the properties of the set that we are concerned with are preserved by translations. Before we demonstrate this technique with an example, we first need the following inequality.

Theorem 4.3.2. (Khintchine inequality) Let (X, μ) be a measure space. Let $\epsilon_1, \cdots, \epsilon_n$ be independent random signs each taking values in $\{+1, -1\}$ with equal probability $1/2$. Let $1 < p < \infty$. Then for all complex-valued functions $f_1, \cdots, f_n \in L^p(X, \mu)$, we have that

$$\mathbb{E}\left[\left\|\sum_{j=1}^{n} \epsilon_j f_j\right\|_{L^p(X,\mu)}^{p}\right] \sim_p \left\|\left(\sum_{j=1}^{n} |f_j|^2\right)^{\frac{1}{2}}\right\|_{L^p(X,\mu)}^{p}.$$

Proof. We first prove that for any complex numbers $z_1, \cdots, z_n$, the estimate

$$\mathbb{E}\left[\left|\sum_{j=1}^{n} \epsilon_j z_j\right|^{p}\right] \sim_p \left(\sum_{j=1}^{n} |z_j|^2\right)^{p/2} \tag{43}$$

holds. Once (43)is proved, by integrating both sides and using Fubini-Tonelli theorem to pull the expectation outside of the integral, we obtain the desired conclusion.

By splitting into the real and imaginary parts, we may assume that each z_i is real. By normalising, we further assume

$$\sum_{j=1}^{n} |z_j|^2 = 1$$

and seek to show both

$$\mathbb{E}\left[\left|\sum_{j=1}^{n} \epsilon_j z_j\right|^{p}\right] \lesssim_p 1 \tag{44}$$

and

$$\mathbb{E}\left[\left|\sum_{j=1}^{n} \epsilon_j z_j\right|^{p}\right] \gtrsim_p 1. \tag{45}$$

When $p = 2$, the equality

$$\mathbb{E}\left[\left|\sum_{j=1}^{n} \epsilon_j z_j\right|^{2}\right] = \sum_{1 \le i,j \le n} \mathbb{E}\big[\epsilon_i \epsilon_j\big] z_i z_j = \sum_{i=1}^{n} |z_j|^2 = 1$$

is achieved hence both (44) and (45) holds. By Hölder's inequality, (44) holds for for all $1 < p \le 2$ and (45) holds for all $2 \le p < \infty$. To handle the remaining cases, we use the exponential moment method. We first show (44) for $2 < p < \infty$. By applying Markov's inequality with the function $x \mapsto e^{tx}$, we see that for any $\alpha, t \in \mathbb{R}_{>0}$, we have

$$P\left(\sum_{j=1}^{n} \epsilon_j z_j \ge \alpha\right) \le \exp(-t\alpha)\mathbb{E}\left[\exp\left(t\sum_{j=1}^{n} \epsilon_j z_j\right)\right] = \exp(-t\alpha)\prod_{j=1}^{n} \mathbb{E}\big[\exp(t\epsilon_j z_j)\big].$$

A direct computation of the expectation gives

$$\mathbb{E}\big[\exp(t\epsilon_j z_j)\big] = \frac{\exp(tz_j) + \exp(-tz_j)}{2} = \cosh(tz_j).$$

By comparing Taylor series, we see that $\cosh(x) \le \exp(x^2/2)$ for any $x \in \mathbb{R}$. Hence

$$\begin{aligned} P\left(\sum_{j=1}^n \epsilon_j z_j \ge \alpha\right) &\le \exp(-t\alpha) \prod_{j=1}^n \exp\left(\frac{t^2 z_j^2}{2}\right) \\ &\le \exp(-t\alpha) \exp\left(\frac{t^2}{2}\right) \end{aligned}$$

where the last line follows from $\sum_{j=1}^n z_j^2 = 1$. Choosing $t = \alpha$ then gives

$$P\left(\sum_{j=1}^n \epsilon_j z_j \ge \alpha\right) \le \exp\left(-\frac{\alpha^2}{2}\right).$$

Since the random variable $\sum_{j=1}^n \epsilon_j z_j$ is symmetric around the origin, we conclude that

$$P\left(\left|\sum_{j=1}^n \epsilon_j z_j\right| \ge \alpha\right) \le 2\exp\left(-\frac{\alpha^2}{2}\right).$$

Finally, using the fact that for any random variable Y on a probability space (Ω, P),

$$\begin{aligned} \int_0^\infty p\alpha^{p-1} P(|Y| \ge \alpha)\ d\alpha &= \int_0^\infty p\alpha^{p-1} \int_\Omega \mathbb{1}_{\{\omega' \,:\, |Y(\omega')| \ge \alpha\}}(\omega)\ dP(\omega)\ d\alpha \\ &= \int_\Omega \int_0^{|Y(\omega)|} p\alpha^{p-1}\ d\alpha\ dP(\omega) \\ &= \int_\Omega |Y(\omega)|^p\ dP(\omega) = \mathbb{E}\big[|Y|^p\big], \end{aligned} \tag{46}$$

we have that

$$\mathbb{E}\left[\left|\sum_{j=1}^n \epsilon_j z_j\right|^p\right] \le 2\int_0^\infty p\alpha^{p-1} \exp\left(-\frac{\alpha^2}{2}\right)\ d\alpha \lesssim_p 1$$

since the integral is finite. Thus (44) holds for all $2 < p < \infty$. Finally, (45) for $1 < p < 2$ can be derived from 44 for $2 < p < \infty$ using Hölder's inequality

$$1 = \mathbb{E}\left[\left|\sum_{j=1}^n \epsilon_j z_j\right|^2\right] \le \mathbb{E}\left[\left|\sum_{j=1}^n \epsilon_j z_j\right|^p\right]^{1/p} \mathbb{E}\left[\left|\sum_{j=1}^n \epsilon_j z_j\right|^q\right]^{1/q}$$

and the fact that the Hölder conjugate q of p satisfies $2 < q < \infty$. This completes the proof of the theorem.

Example 4.3.3. Consider a smooth compact hypersurface S in $\mathbb{R}^n$, with the surface measure $d\sigma$. Introduced by Elias Stein in the 1970s, the restriction problem asks: under what conditions do we have estimates (possibly depending on parameters) of the form

$$\|\widehat{f}\|_{L^q(S,d\sigma)} \lesssim \|f\|_{L^p(\mathbb{R}^n)}\,.$$

In the context of the unit sphere S^{n-1}, Bourgain was able to leverage on the group structure of S^{n-1} to improve on existing estimates.

Definition 4.3.4. (Weak L^p-spaces) Let (X,μ) be a measure space and let $f : X \to \mathbb{C}$ be a measurable function. The **distribution function** of f is defined to be the function $d_f : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ given by

$$d_f(\alpha) := \mu\Big(\big\{x \in X \ : \ |f(x)| \geq \alpha\big\}\Big).$$

For $1 \leq p < \infty$, the space **weak $\boldsymbol{L^p}$** (denoted by $L^{p,\infty}(X,\mu)$) is defined to be the set of all measurable functions $f : X \to \mathbb{C}$ such that

$$\|f\|_{L^{p,\infty}(X,\mu)} := \inf\left\{C \in \mathbb{R}_{>0} \ : \ d_f(\alpha) \leq \frac{C^p}{\alpha^p} \text{ for all } \alpha > 0\right\}.$$

is finite.

Remark 4.3.5. If $f \in L^p(X,\mu)$, then a direct application of Chebyshev's inequality gives the bound

$$\alpha^p d_f(\alpha) \leq \int_{\{x \ : \ |f(x)|\geq\alpha\}} |f(x)|^p \ d\mu(x) \leq \|f\|^p_{L^p(X,\mu)}\,.$$

It then follows that $\|f\|_{L^{p,\infty}(X,\mu)} \leq \|f\|_{L^p(X,\mu)}$ and so the weak L^p space is indeed larger than the usual L^p space.

Remark 4.3.6. For any $1 \leq p < q < \infty$, the "layered cake representation" (using a similar argument as in (46)) shows that for any $f \in L^{q,\infty}(X,\mu)$, we have the bound

$$\|f\|^p_{L^p(X,\mu)} = \int_0^\infty p\alpha^{p-1} d_f(\alpha) \ d\alpha \leq \|f\|^q_{L^{q,\infty}(X,\mu)} \int_0^\infty p\alpha^{p-q-1} \ d\alpha \lesssim \|f\|^q_{L^{q,\infty}(X,\mu)}$$

since the integral converges.

Proposition 4.3.7. (Restriction estimates on the sphere) Suppose that $n \geq 2$ and $1 < p < 2$ is such that one has the restriction estimate

$$\|\widehat{f}\|_{L^1(S^{n-1},d\sigma)} \lesssim_{p,n} \|f\|_{L^p(\mathbb{R}^n)} \tag{47}$$

for all Schwartz functions $f : \mathbb{R}^n \to \mathbb{C}$, where σ is the normalized surface measure on the sphere S^{n-1}. We will demostrate that one can automatically improve this to the stronger (by the previous remark) estimate

$$\|\widehat{f}\|_{L^{p,\infty}(S^{n-1},d\sigma)} \lesssim_{p,n} \|f\|_{L^p(\mathbb{R}^n)}\,. \tag{48}$$

We first normalise $\|f\|_{L^p(\mathbb{R}^n)} = 1$. For each $\lambda \in \mathbb{R}_{>0}$, let E_λ denote the level set

$$E_\lambda := \left\{ \boldsymbol{\omega} \in S^{n-1} \ : \ |\widehat{f}(\boldsymbol{\omega})| \geq \lambda \right\}.$$

The estimate (48) is then equivalent to $\sigma(E_\lambda) \lesssim_{p,n} \lambda^{-p}$. Applying (47) directly to

$$\|\cdot\|_{L^{1,\infty}(S^{n-1},d\sigma)} \leq \|\cdot\|_{L^1(S^{n-1},d\sigma)}$$

only gives the estimate

$$\sigma(E_\lambda) \lesssim_{p,n} \lambda^{-1}$$

which is inferior when λ is large. However, by rounding up $1/\sigma(E_\lambda)$ to the nearest integer N_λ and then applying Lemma 4.3.1 to S^{n-1} (viewed as a compact group via the identification $S^{n-1} \cong SO(n)/SO(n-1)$), there exist rotations $\boldsymbol{R}_1, \cdots, \boldsymbol{R}_{N_\lambda}$ such that

$$\sigma\left(\bigcup_{k=1}^{N_\lambda} \boldsymbol{R}_k E_\lambda \right) \sim 1.$$

Define

$$F(\boldsymbol{x}) := \sum_{k=1}^{N_\lambda} \epsilon_k f(\boldsymbol{R}_k \boldsymbol{x})$$

where the ϵ_i's are independent random signs taking values in $\{+1, -1\}$ with equal probability $1/2$. By Khintchine's inequality (Theorem 4.3.2), we have that

$$\mathbb{E}\Big[\|F\|_{L^p(\mathbb{R}^n)}^p \Big] \sim_p \int_{\mathbb{R}^n} \left| \sum_{k=1}^{N_\lambda} \big| f(\boldsymbol{R}_k \boldsymbol{x}) \big|^2 \right|^{\frac{p}{2}} d\mu(\boldsymbol{x}) \leq \sum_{k=1}^{N_\lambda} \int_{\mathbb{R}^n} \big| f(\boldsymbol{R}_k \boldsymbol{x}) \big|^p \, d\mu(\boldsymbol{x}) = N_\lambda.$$

Here, the second last equality follows from the fact that for any $r \in (0,1)$, we have $(x+y)^r \leq x^r + y^r$ (and the fact that $p/2 < 1$). This implies that there exist signs $s_1, \cdots, s_{N_\lambda} \in \{-1, 1\}$ such that the function G defined by

$$G(\boldsymbol{x}) := \sum_{k=1}^{N_\lambda} s_k f(\boldsymbol{R}_k \boldsymbol{x})$$

satisfies

$$\|G\|_{L^p(\mathbb{R}^n)} \lesssim_p N_\lambda^{1/p}.$$

Let $f_k(\boldsymbol{x})$ denote $f(\boldsymbol{R}_k \boldsymbol{x})$. Then by using the fact that each $\boldsymbol{R}_k$ is orthogonal, we have

$$\widehat{f_k}(\boldsymbol{\xi}) = \frac{1}{|\det \boldsymbol{R}_k|} \widehat{f}\big((\boldsymbol{R}_k^{-1})^T \boldsymbol{\xi} \big) = \widehat{f}(\boldsymbol{R}_k \boldsymbol{\xi}).$$

Thus

$$\left\{ \boldsymbol{\omega} \in S^{n-1} \ : \ |\widehat{f_k}(\boldsymbol{\omega})| \geq \lambda \right\} = \boldsymbol{R}_k^{-1} E_\lambda.$$

Hence we have

$$\sigma\Big(\big\{\boldsymbol{\omega} \in S^{n-1} \ : \ |\widehat{G}(\boldsymbol{\omega})| \gtrsim \lambda\big\}\Big) \gtrsim \sigma\left(\bigcup_{k=1}^{N_\lambda} \boldsymbol{R}_k E_\lambda\right) \gtrsim 1.$$

Now if we apply (47) to the function G, we have

$$1 \lesssim \sigma\Big(\big\{\boldsymbol{\omega} \in S^{n-1} \ : \ |\widehat{G}(\boldsymbol{\omega})| \gtrsim \lambda\big\}\Big) \lesssim_{p,n} \frac{\|G\|_{L^p(\mathbb{R}^n)}}{\lambda} \lesssim_p \frac{N_\lambda^{1/p}}{\lambda}$$

which gives the required estimate of $\sigma(E_\lambda) \lesssim_{p,n} \lambda^{-p}$.

References

Bourga'in, J. (1991). Besicovitch type maximal operators and applications to fourier analysis. *Geometric & Functional Analysis GAFA*, *1*(2), 147–187.

Tao, T. (2007, Sep). *Amplification, arbitrage, and the tensor power trick.* Retrieved from `https://terrytao.wordpress.com/2007/09/05/amplification-arbitrage-and-the-tensor-power-trick/`

Tao, T. (2008, Aug). *Tricks wiki article: The tensor power trick.* Retrieved from `https://terrytao.wordpress.com/2008/08/25/tricks-wiki-article-the-tensor-product-trick/`

Tao, T. (2013, Jul). *A cheap version of the kabatjanskii-levenstein bound for almost orthogonal vectors.* Retrieved from `https://terrytao.wordpress.com/2013/07/18/a-cheap-version-of-the-kabatjanskii-levenstein-bound-for-almost-orthogonal-vectors/`

Tao, T. (2017, Aug). *An addendum to "amplification, arbitrage, and the tensor power trick".* Retrieved from `https://terrytao.wordpress.com/2017/08/22/an-addendum-to-arbitrage-amplification-and-the-tensor-power-trick/`

Tao, T. (2020, Mar). *247b, notes 1: Restriction theory.* Retrieved from `https://terrytao.wordpress.com/2020/03/29/247b-notes-1-restriction-theory/`

Tao, T. (2021). Exploring the toolkit of jean bourgain. *Bulletin of the American Mathematical Society*, *58*(2), 155–171.

Tao, T., & Vu, V. H. (2006). *Additive combinatorics* (Vol. 105). Cambridge University Press.