A short overview on Vinogradov's Mean Value Theorem and Decoupling Inequalities

Nirmal Krishnan

June 2023

1 Introduction

This chapter is based mainly on [1]. One of the most famous problems in number theory is the one of Waring: For every $k \in \mathbb{N}$, can every natural number N be written as

$$N = x_1^k + \ldots + x_s^k,$$

where $x_1, \ldots, x_s$ are nonnegative integers and s is some natural number dependent on k.

The existence of such an s was proved by Hilbert, but his proof was unfortunately ineffective; it gave no bounds or estimates on s as a function of k. Writing $g(k)$ for the least such s, it is known that $g(2) = 4, g(3) = 9$ and so on, up to atleast $g(6) = 73$.

Another target of study in this corner of mathematics is to study $r_{s,k}(N)$, the number of representations of N as the sum of s kth powers, and in our case specifically its asymptotics. Using their circle method, the idea of Hardy and Littlewood is to write

$$r_{s,k}(N) = \int_0^1 g_k(\alpha, X)^s e(-N\alpha)\,\mathrm{d}\alpha,$$

where

$$g_k(\alpha, X) = \sum_{1 \leq x \leq X} e(\alpha x^k),$$

and $X \approx N^{1/k}$. So to understand the asymptotics of $r_{s,k}$, we want to understand the behaviour of g_k. To this end, we follow Vinogradov, who noted that

$$\int_0^1 |g_k(\alpha, X)|^{2s}\,\mathrm{d}\alpha$$

counts the number of solutions of the equation

$$x_1^k + \ldots + x_s^k = x_{s+1}^k + \ldots + x_{2s}^k$$

which follows from writing $|g_k(\alpha, X)|^{2s} = g_k(\alpha, X)^s \cdot \overline{g_k(\alpha, X)}^s$ and noting that when integrating the only exponentials which live will be the ones where the exponents cancel each other out, i.e. the tuples $(x_1, \ldots, x_{2s})$ for which $x_1^k + \ldots + x_s^k - x_{s+1}^k - \ldots - x_{2s}^k = 0$. We do not stop to examine this function g_k, but instead define one more auxillary function $J_{k,s}(X)$ which asymptotics we will study. Define $J_{k,s}(X)$ to be number of intgral solutions to the system of k equations

$$x_1^\ell + \ldots + x_s^\ell = x_{s+1}^\ell + \ldots + x_{2s}^\ell, \quad (1 \leq \ell \leq k)$$

with $1 \leq x_i \leq X$. Note first that for $J_{k,s}$ we have the identity

$$\int_{(0,1]^k} |f_k(\boldsymbol{\alpha}, X)|^{2s} \, \mathrm{d}\boldsymbol{\alpha} = J_{s,k}(X) \tag{1.1}$$

where

$$f_k(\boldsymbol{\alpha}, X) = \sum_{1 \leq x \leq X} e(\alpha_1 x + \alpha_2 x^2 + \ldots + \alpha_k x^k).$$

Identity (1.1) then follows in an identical fashion as when talking about g_k. Note then that we also have

$$\int_0^1 |g_k(\alpha, X)|^{2s} \, \mathrm{d}\alpha = \sum_{|h_1| \leq sX} \cdots \sum_{|h_k| \leq sX^{k-1}} \int_{(0,1]^k} |f_k(\boldsymbol{\alpha}, X)|^{2s} e(-h_1\alpha_1 - h_2\alpha_2 - \ldots - h_{k-1}\alpha_{k-1}) \, \mathrm{d}\boldsymbol{\alpha},$$

by a similar counting argument. By uniformly bounding the integral by $J_{k,s}$ and counting the number of terms in the above $(k-1)$-fold sum we see that

$$\int_0^1 |g_k(\alpha, X)|^{2s} \, \mathrm{d}\alpha \ll_{k,s} X^{\frac{k(k-1)}{2}} J_{s,k}(X),$$

where $\ll_{k,s}$ means that the inequality holds with a constant depending on k and s. We can now present the main conjecture and the recent results on the problem by Bourgain, Demeter and Guth.

Conjecture 1. For all $s, k \geq 1, X \geq 1$ and arbitrary ε, the function $J_{s,k}(X)$ has the following asymptotic estimate:

$$J_{s,k}(X) \ll_{s,k,\varepsilon} X^\varepsilon (X^s + X^{2s - \frac{k(k+1)}{2}})$$

Of particular interest is the critical case when $2s - \frac{k(k+1)}{2} = s$, i.e. when both terms in the asymptotic are of equal size. It turns out that proving the statment in this case, when $s_k = \frac{k(k+1)}{2}$ is enough to establish it for all other $s \geq 1$! Assume that we have proven the statement in the critical case, that

$$J_{s_k,k}(X) = \int_{(0,1]^k} |f(\boldsymbol{\alpha}, X)|^{k(k+1)} \,\mathrm{d}\boldsymbol{\alpha} \ll_{s,k,\varepsilon} X^{\varepsilon + \frac{k(k+1)}{2}}.$$

Then for $s > s_k$ we can compute that

$$\begin{aligned} J_{s,k}(X) = \int_{(0,1]^k} |f(\boldsymbol{\alpha}, X)|^{2s} \,\mathrm{d}\boldsymbol{\alpha} &\le \sup_{\boldsymbol{\alpha} \in (0,1]^k} |f(\boldsymbol{\alpha}, X)|^{2s-k(k+1)} \int_{(0,1]^k} |f(\boldsymbol{\alpha}, X)|^{k(k+1)} \,\mathrm{d}\boldsymbol{\alpha} \\ &\ll_{k,s,\varepsilon} X^{\varepsilon} \cdot X^{2s-k(k+1)} \cdot X^{k(k+1)} = X^{\varepsilon + 2s - k(k+1)} \end{aligned}$$

which is the dominant term when $s > s_k$. When $s < s_k$, let $r = s/s_k$ and r' its conjugate exponent. Now by Hölder we can compute that

$$\begin{aligned} \int_{(0,1]^k} |f_k(\boldsymbol{\alpha}, X)|^{2s} \,\mathrm{d}\boldsymbol{\alpha} &\le \left(\int_{(0,1]^k} 1^{r'} \right)^{\frac{1}{r'}} \left(\int_{(0,1]^k} |f_k(\boldsymbol{\alpha}, X)|^{2sr} \,\mathrm{d}\boldsymbol{\alpha} \right)^{\frac{1}{r}} \\ &\ll_{k,s,\varepsilon} \left(X^{\varepsilon + s_k} \right)^{\frac{1}{r}} \ll X^{\varepsilon + s}. \end{aligned}$$

Therefore it suffices to show the cases when $s = s_k$. The case $k = 1, s = 2$ is trivial, and the case when $k = 2, s_k = 3$ can be established after observing that the system of equations

$$\begin{cases} x_1 + x_2 + x_3 = y_1 + y_2 + y_3 \\ x_1^2 + x_2^2 + x_3^2 = y_1^2 + y_2^2 + y_3^2 \end{cases}$$

implies that

$$(x_1 - y_3)(x_2 - y_3) = (y_1 - x_3)(y_2 - x_3),$$

and then doing a bit of counting. The case when $k = 3$ proved out for reach until Wooley proved it in 2016, and later the conjecture was shown for all $k \ge 4$ by Bourgain, Demeter and Guth. Later proofs which have simplified their argument have also appeared, and the one by Guo, Li, Yung and Zorin-Kranich will be the focus of the rest of our essay.

2 Baby Decoupling

This chapter is mainly based on [2]. Let us first turn to a very familiar phenomenon, namely orthogonality in L^2. We know that for a collection of orthogonal functions $\{f_j\}$ we have the following identity:

$$\| \sum_j f_j \|_{L^2} = \left(\int \left(\sum_j f_j \right) \left(\overline{\sum_i f_i} \right) \right)^{1/2} = \left(\sum_j \|f_j\|_{L^2}^2 \right)^{1/2} \tag{2.1}$$

To make the condition of orthogonality a bit more concrete, we can specialize the above identity to the case where the f_j:s have disjoint Fourier support, i.e. the sets $\operatorname{supp} \hat{f}_j$ are disjoint. Without any additional conditions on the functions f_j, if we assumed that the collection J of indices j were finite, we could by Cauchy-Schwarz obtain the inequality in (2.1) with an additional factor of $|J|^{1/2}$ on the RHS.

The bound given by Cauchy-Schwarz does not require the space L^2 in any way, so every L^2-norm in (2.1) could be replaced by a L^p norm to get the estimate

$$\| \sum_j f_j \|_{L^p} \leq |J|^{1/2} \left(\sum_j \|f_j\|_{L^p}^2 \right)^2 . \tag{2.2}$$

The question then arises of which conditions we need to impose on functions in L^p to get a constant better than this $|J|^{1/2}$ in (2.2).

It turns out that the condition we will use to define decoupling inequalities regards the Fourier supports of the functions f_j, as we will use the following definition of a decoupling inequality:

Definition (Decoupling inequality)**.** Let Ω be a subset of $\mathbb{R}^d$, and $\{\theta\}$ a partition of Ω. Furthermore, let f_θ be Schwartz functions with $\operatorname{supp} \hat{f}_\theta \subset \theta$. Now the *decoupling constant* $\mathrm{Dec}_p(\Omega)$ is the least constant such that we have the inequality

$$\| \sum_\theta f_\theta \|_{L^p(\mathbb{R}^d)} \leq \mathrm{Dec}_p(\Omega) \left(\sum_\theta \|f_\theta\|_{L^p(\mathbb{R}^d)}^2 \right)^{1/2} .$$

Note that we allow this constant to depend on the partition, and to this end, some authors write the decoupling constant as

$$\mathrm{Dec}_p(\Omega = \bigcup \theta).$$

Decoupling inequalities have a natural context in the theory of restriction problems, where one works on curved submanifolds of $\mathbb{R}^d$. In most cases (including for Conjecture 1.1) what we want to show is that

$$\mathrm{Dec}_p(\Omega) \lesssim_{p,d,\varepsilon} |\{\theta\}|^\varepsilon$$

for every $\varepsilon > 0$. To show Conjecture 1.1 in the case $k = 2$, we would pick Ω to be the δ^2-neighbourhood of the two-dimensional moment curve $\Gamma = \{(t, t^2 + s) : t \in [0,1], |s| \leq \delta^2\}$ above $[0,1]$. The partition $\{\theta\}$ is defined using a partition P_δ of the interval $[0,1]$ into pieces of length $\approx \delta$. For $I \in P_\delta$, we define $\theta_I = \{(t, t^2 + s) : t \in I, |s| \leq \delta^2\}$, and so what is to be shown is that $\mathrm{Dec}_p(\Omega) \lesssim_{p,\varepsilon} \delta^{-\varepsilon}$.

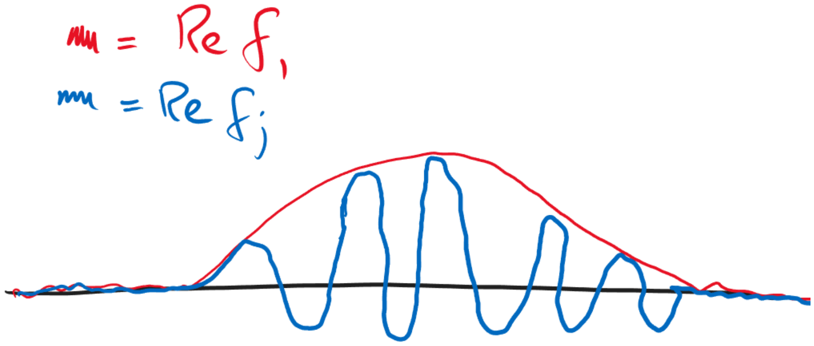

Figure 1: We have in red marked the real part of the waveform of f_1, and in blue the real part of the waveform of f_j. The main observation is that near some small neighbourhood of the origin dependent on N, both waves are essentially 1.

To end this chapter, let us give an example why the geometry of the submanifold is important. Let $\Omega = [0, N] \subset \mathbb{R}$, and let $\theta_j = [j-1, j]$ form a partition of Ω. Now

$$\mathrm{Dec}_p(\Omega) \gtrsim_p N^{\frac{1}{2}-\frac{1}{p}}$$

for $p > 2$ (in fact $\sim_p$ holds in the above but we will only motivate the lower bound). Let f_1 be a function concentrated on $[-1, 1]$ with $f_1(0) = 1$ and its Fourier transform is supported in the interval $[0, 1]$. Now define $f_j(x) = e^{-2\pi i(j-1)} f_1(x)$ and note that the Fourier support of f_j is in $[j, j-1]$. Now note that each f_j oscillates with frequency around $1/j$, and so we get the maximal constructive interference at least in a neighbourhood of size $\approx 1/N$ around the origin, so

$$|\sum_{j=1}^{N} f_j(x)| \sim N$$

in this neighbourhood of size $1/N$, and so

$$\|\sum_{j=1}^{N} f_j\|_{L^p(\mathbb{R})} \gtrsim N^1 N^{-1/p} = N^{1-\frac{1}{p}}.$$

We can also compute that $\|f_j\|_{L^p} \sim 1$, and so specializing our decoupling inequality for these choices of f_j we find that

$$N^{1-\frac{1}{p}} \leq \mathrm{Dec}_p(\Omega) N^{\frac{1}{2}},$$

and so we must have

$$\mathrm{Dec}_p(\Omega) \gtrsim N^{\frac{1}{2}-\frac{1}{p}}.$$

To end this chapter, note that the desired estimate in the critical case of Conjecture 1.1 looks at least cosmetically like a decoupling inequality, namely we would wish to show that

$$\int_{(0,1]^k} |f_k(\boldsymbol{\alpha}, X)|^{k(k+1)} \,\mathrm{d}\boldsymbol{\alpha} \ll X^{\varepsilon + k(k+1)}.$$

Estimating

$$\int_{(0,1]^k} |e(\alpha_1 x + \ldots + \alpha_k x^k)|^{k(k+1)} \,\mathrm{d}\boldsymbol{\alpha} \leq \int_{(0,1]^k} 1 \,\mathrm{d}\boldsymbol{\alpha} = 1,$$

we note that it suffices to show that

$$\int_{(0,1]^k} \left| \sum_{1 \leq x \leq X} e(\alpha_1 x + \alpha_2 x^2 + \ldots + \alpha_k x^k) \right|^{k(k+1)} \mathrm{d}\boldsymbol{\alpha}$$

$$\ll X^{\varepsilon} \left(\sum_{1 \leq x \leq X} \left(\int_{(0,1]^k} |e(\alpha_1 x + \alpha_2 x^2 + \ldots + \alpha_k x^k)|^{k(k+1)} \,\mathrm{d}\boldsymbol{\alpha} \right)^{\frac{1}{k(k+1)}} \right)^{k(k+1)}.$$

Finally taking p:th roots and fiddling around we see that it suffices to show that

$$\| \sum_{1 \leq x \leq X} e(\alpha_1 x + \ldots + \alpha_k x^k) \|_{L^{k(k+1)}((0,1]^k)} \ll X^{\varepsilon} \left(\sum_{1 \leq x \leq X} \| e(\alpha_1 x + \ldots + \alpha_k x^k) \|^2_{L^{k(k+1)}((0,1]^k)} \right)^{1/2},$$

which is reminiscent of a decoupling inequality with "decoupling constant" of order X^{ε}. This is in fact the optimal case for decoupling (and what we wish to show for the moment curve, which will be defined later), and next we will show a case where this optimality is not achieved.

3 The proof of Guo-Li-Yang-Zorin-Kranich

In this section, we will go over the proof laid out in [3] for decoupling for the moment curve. Let us first introduce some notation. Let

$$\Gamma : [0,1] \to \hat{\mathbb{R}}^k, \Gamma(t) = (t, t^2, \ldots, t^k)$$

be the moment curve in $\mathbb{R}^k$, where $\hat{\mathbb{R}}^k$ is the Pontryagin dual of $\mathbb{R}^k$ (as $\mathbb{R}^k$ is self-dual, we can just identify it with $\mathbb{R}^k$ and then use the parametrization). For a dyadic interval I and $\delta > 0$, let $\mathcal{P}(I, \delta)$ be the partition of I into dyadic intervals of length $2^{\lceil \log_2 \delta \rceil}$, and as a special case, let $\mathcal{P}(\delta) = \mathcal{P}([0,1], \delta)$.

Define also for a sufficiently nice smooth curve γ the parallelepipeds $\mathcal{U}_{J,\gamma}$ to be the parallelepiped of dimensions $|J| \times |J|^2 \times \ldots \times |J|^k$ with sides parallel to $\partial^1\gamma(c_J), \partial^2\gamma(c_J), \ldots \partial^k\gamma(c_J)$ where c_J is the midpoint of the interval J. We also write $\mathcal{U}_J$ as shorthand for $\mathcal{U}_{J,\Gamma}$.

We are now ready to state the decoupling theorem for the moment curve.

Theorem 1 (ℓ^2 decoupling for the moment curve)**.** Let $\mathrm{Dec}_{k,p}(\delta)$, and $\mathrm{Dec}_k(\delta) = \mathrm{Dec}_{k,p_k}(\delta)$, where $p_k = k(k+1)$ be the smallest constant for which we have the inequality

$$\| \sum_{J \in \mathcal{P}(\delta)} f_J \|_{L^p(\mathbb{R}^k)} \leq \mathrm{Dec}_k(\delta) \left(\sum_{J \in \mathcal{P}(\delta)} \|f_J\|^2_{L^p(\mathbb{R}^k)} \right)^{1/2}.$$

Then for all $\delta > 0$ we have

$$\mathrm{Dec}_p(\delta) \lesssim_{k,\varepsilon} \delta^{-\varepsilon}.$$

Note that this is a decoupling inequality as described in the previous chapter, but with the role of Ω omitted and some notational quirks. The reason why we omit p and instead mark the dimension is that our argument will use induction on the dimension, while the exponent p is going to be held fixed at the critical exponent $p = p_k = k(k+1)$. As an aside, one could work as we described in the two-dimensional case, but it is not necessary.

The first idea that is required before we perform a series of reductions is one of rescaling, which uses the fact that one piece of the moment curve (some $\mathcal{U}_J$) is equivalent to some other piece (some $\mathcal{U}_{J'}$) by an affine transformation A_J depending on J. This means that we can take decompositions to decompositions affinely, that is for $\delta < \delta_0$

$$\{A_I \mathcal{U}_J\}_{J \in \mathcal{P}(I,\delta)} = \{\mathcal{U}_{J'}\}_{J' \in \mathcal{P}(\delta/\delta_0)}$$

and so we can take decompositions of some interval I into decompositions of $[0,1]$ of scale δ/δ_0. This one of the main parts where the geometry of the moment curve is used; for manifolds for which we are unable to prove a rescaling lemma, one must work a lot harder.

Lemma 1 (Rescaling for the moment curve)**.** Let $I \in \mathcal{P}(2^{-n})$, $\delta \in (0, 2^{-n})$, and $\{f_J\}_{J \in \mathcal{P}(I,\delta)}$ functions with Fourier support in J. Then

$$\| \sum_{J \in \mathcal{P}(I,\delta)} f_J \|_{L^p(\mathbb{R}^k)} \leq \mathrm{Dec}_k(2^n\delta) \left(\sum_{J \in \mathcal{P}(I,\delta)} \|f_J\|_{L^p(\mathbb{R}^k)} \right)^{1/2}.$$

Note that using this rescaling lemma we are able to pass from one scale to a smaller scale. Throughout the paper of Guo-Li-Yang-Zorin-Kranich, dyadic intervals are used, so we will conform to them as well. In this case though, it is maybe a bit easier to see the multiplicative consequence of rescaling when we give the rescaling lemma in a slightly different form.

Lemma 2 (Rescaling for the moment curve, non-dyadic version)**.** Let $\mathcal{P}'(I,\delta)$ be the partition of I into intervals of length δ, and everything else defined identically as for dyadic intervals. Now for $0 < \delta \leq \delta_0 \leq 1$ we have

$$\| \sum_{J \in \mathcal{P}'(I,\delta_0)} f_J \|_{L^p(\mathbb{R}^k)} \leq \mathrm{Dec}_k(\delta/\delta_0) \left(\sum_{J \in \mathcal{P}'(I,\delta)} \|f_J\|_{L^p(\mathbb{R}^k)} \right)^{1/2} .$$

From this version of the lemma, we can see that if we had $0 < \delta_2 \leq \delta_1 \leq 1$, then

$$\begin{aligned} \| \sum_{J \in \mathcal{P}'(I,\delta_1)} f_J \|_{L^p(\mathbb{R}^k)} &\leq \mathrm{Dec}_k(\delta_1) \left(\sum_{J \in \mathcal{P}'(I,\delta)} \|f_J\|_{L^p(\mathbb{R}^k)} \right)^{1/2} \\ &\leq \mathrm{Dec}_k(\delta_1)\, \mathrm{Dec}_k(\delta_2/\delta_1) \left(\sum_{J \in \mathcal{P}'(I,\delta_1)} \sum_{J' \in \mathcal{P}'(J,\delta_2)}) \|f_{J'}\|_{L^p(\mathbb{R}^k)} \right)^{1/2} , \end{aligned}$$

which implies that $\mathrm{Dec}_k(\delta_2) \leq \mathrm{Dec}_k(\delta_1)\, \mathrm{Dec}_k(\delta_2/\delta_1)$.

The next main idea is that of replacing our current "linear" decoupling inequality by a bilinear one, that is, we try to prove an inequality of the form

$$\int_{\mathbb{R}^k} | \sum_{J \in \mathcal{P}(I,\delta)} f_J |^{p_k/2} | \sum_{J' \in \mathcal{P}(I',\delta)} f_{J'} |^{p_k/2} \leq C^{p_k} \left(\sum_{J \in \mathcal{P}(I,\delta)} \|f_J\|_{p_k}^2 \right)^{p_k/4} \left(\sum_{J' \in \mathcal{P}(I',\delta)} \|f_J'\|_{p_k}^2 \right)^{p_k/4} , \tag{3.1}$$

where the f_J satisfy the same conditions as previously postulated. We also want to make sure that this inequality does not directly devolve into our linear one, so adding a condition to ensure the separatedness of the intervals I and I' is needed. We will choose to mandate that $I, I' \in \mathcal{P}(1/4)$ and $\mathrm{dist}(I, I') \geq 1/4$, i.e. that cases where $I = I'$ are not considered.

Going from a linear to a bilinear estimate does make things a bit more cumbersome to work with, but in exchange it allows us to get cancellations between the terms which are not possible to get in the linear case. Of course, this is not of much use unless we can prove that these bilinear estimates somehow control the "linear" decoupling constant Dec_k. To this end, let the bilinear decoupling constant $\mathrm{B_k}(\delta)$ be the least constant such that the equation (3.1) holds. It is important to note that the rescaling lemma holds also for this bilinear

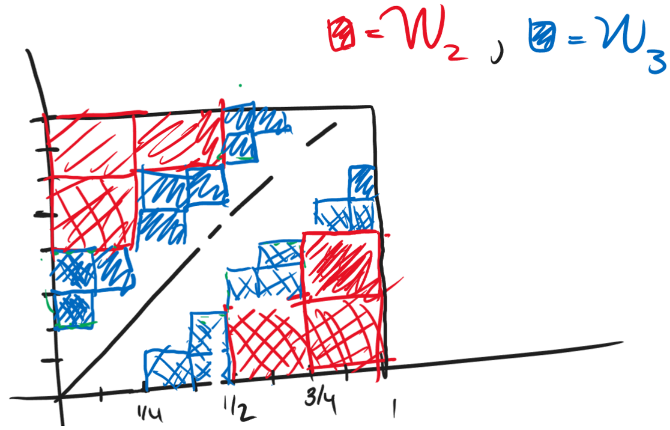

Figure 2: In red are the squares in $\mathcal{W}_2$ and in blue the squares in $\mathcal{W}_3$.

decoupling inequality. It requires that the intervals I and I' be sufficiently separated, and instead of $\mathrm{Dec}_k(2^n\delta)$ gives us $\mathrm{B_k}(2^{n-2}\delta)$. Now we claim that

$$\mathrm{Dec}_k(2^{-N}) \lesssim \left(1 + \sum_{n=2}^{N} \mathrm{B_k}(2^{-N+n-2})^2\right)^{1/2}. \tag{3.2}$$

By the above, we can then shift our focus to studying these bilinear decoupling constants.

Proof of (3.2). We decompose the unit square into pieces according to the following procedure: Let $\mathcal{W}_1 =$, and for $n \geq 2$ define

$$\mathcal{W}_n = \left\{ (I_1, I_2) \in \mathcal{P}(2^{-n}) : 2^n \operatorname{dist}(I_1, I_2) \in \{1, 2\} \quad \text{and} \quad I_1 \times I_2 \not\subset \bigcup_{(I_1', I_2') \in \mathcal{W}_{n-1}} I_1' \times I_2' \right\}.$$

Define also the collections

$$\tilde{\mathcal{W}}_n = \{(I_1, I_2) \in \mathcal{P}(2^{-n}) : \operatorname{dist}(I_1, I_2) = 0\}.$$

One can note that the set $\tilde{\mathcal{W}}_n$ corresponds to the empty space left after the n:th stage of the decomposition. Therefore

$$\mathcal{W}^N = \{\mathcal{W}_1, \ldots, \mathcal{W}_N, \tilde{\mathcal{W}}_N\}$$

forms a essentially disjoint covering of the unit square. Note then that we have

$$\|\sum_{I \in \mathcal{P}(2^{-N})} f_I\|_{p_k} = \|\sum_{(I,I') \in \mathcal{W}^N} f_I \overline{f_{I'}}\|_{p_k/2}^{1/2} \leq \left(\sum_{(I,I') \in \mathcal{W}^N} \|f_I \overline{f_{I'}}\|_{p_k/2}\right)^{1/2},$$

and this final sum we can break up into the colored parts $\mathcal{W}_n, n = 1, \dots, N$ and the white parts $\tilde{\mathcal{W}}_N$. In the white parts, note that the product $\|f_I\|\|\overline{f_{I'}}\|$ is bounded by the sum

$$\|f_I\|_{p_k}^2 + \|\overline{f_{I'}}\|_{p_k}^2$$

by Hölders inequality. From our diagram we can also see that each I appears at most six times inside $\tilde{\mathcal{W}}_N$, so we get

$$\sum_{(I,I')\in\tilde{\mathcal{W}}_N} \|f_I\overline{f_{I'}}\|_{p_k} \leq 6 \sum_{I\in\mathcal{P}(2^{-N})} \|f_I\|_{p_k}^2.$$

To then estimate

$$\sum_{(I,I')\in\mathcal{W}_n} \|f_I\overline{f_{I'}}\|_{p_k/2}$$

we can apply affine rescaling directly (trust me, the conditions are met), to get

$$\|f_I\overline{f_{I'}}\|_{p_k/2} \leq \mathrm{B_k}(2^{-N+n-2})^2 \left(\sum_{J\in\mathcal{P}(I,2^{-N})} \|f_J\|_{p_k}^2\right)^{1/2} \left(\sum_{J'\in\mathcal{P}(I',2^{-N})} \|f_{J'}\|_{p_k}^2\right)^{1/2}.$$

Again bounding this product of square roots by a sum of squares we find that

$$\|f_I\overline{f_{I'}}\|_{p_k/2} \leq \mathrm{B_k}(2^{-N+n-2})^2 \left(\sum_{J\in\mathcal{P}(I,2^{-N})} \|f_J\|_{p_k}^2 + \sum_{J\in\mathcal{P}(I,2^{-N})} \|f_J\|_{p_k}^2\right).$$

To bound this, note that every interval $I \in \mathcal{P}(2^{-n})$ occurs at most 8 times in each $\mathcal{W}_n$, and so we get

$$\sum_{(I,I')\in\mathcal{W}_n} \|f_I\overline{f_{I'}}\|_{p_k/2} \leq C \sum_{I\in\mathcal{P}(2^{-n})} \sum_{J\in\mathcal{P}(I,2^{-N})} \|f_J\|_{p_k} = C \sum_{J\in\mathcal{P}(2^{-N})} \|f_J\|_{p_k},$$

where the last equality follows from the fact that the intervals are dyadic. Adding up the above estimate for all $n = 1, \dots, N$ we arrive at (3.2). $\square$

Before reducing our bilinear decoupling constants to asymmetric bilinear decoupling constants, we will have to take a look at uncertainty principle heuristics. To motivate the uncertainty principle, recall that the Fourier transform of a dilated function $f(x/a)$ in $\mathbb{R}^k$ is $a^k\hat{f}(ax)$. If now f were for example a function with support in the unit ball, the function $f(x/a)$ would have ever smaller support as a went to 0, but the region where the Fourier transform is large would increase.

The uncertainty principle may be thought of as a general class of heuristics that tell us about the behaviour of a function in some scale given that we know the behaviour of its Fourier transform in the dual scale. In our case, for a parallelipiped $\mathcal{U}_J$, we define its dual parallelipiped $\mathcal{U}_J^\circ$ as a parallelipiped centered at the origin, with dimensions $\sim |J|^{-1} \times |J|^{-2} \times \ldots \times |J|^{-k}$. The direction in which the parallelipiped points is not as important for now. The uncertainty principle then says that if we know the behaviour of the Fourier transform of f_J in $\mathcal{U}_J$, then we know how f_J behaves in $\mathcal{U}_J^\circ$. This takes on a even more pronounced form for our uses; if the Fourier support of f_J is in some $\mathcal{U}_J$, then f_J is controlled globally by its behaviour in $\mathcal{U}_J^\circ$. To translate this into actual mathematical language, we can write it as follows:

Lemma 3. (Uncertainty principle) Let $p \in [1, \infty)$, $J \subset [0,1]$, and φ_J a normalized bump function adapted to $\mathcal{U}_J^\circ$ (a normalized bump function in this case means that its largest value is $\|\mathcal{U}_J^\circ\|^{-1}$). Now for every g_J with Fourier support in some dilation $C\mathcal{U}_J$ of $\mathcal{U}_J$ we have

$$|g_J|^p \lesssim_p |g_J|^p * \varphi_J.$$

The lemma tells us that g_J is globally controlled by its behaviour in $\mathcal{U}_J^\circ$, so calling it a uncertainity principle is not a misnomer. To prove the lemma one defines ψ such that its Fourier transform is identically 1 on $C\mathcal{U}_J$ and that its L^1-norm is 1, and notes that $g_J = g_J * \psi$.

We are now ready for our final reduction and to set the stage for the induction stage of the argument.

Definition (Asymmetric bilinear decoupling constant)**.** Let $l \in \{0, \ldots, k-1\}, a, b \in [0,1]$ and $\delta \in (0,1)$. The asymmetric bilinear decoupling constant $B_{\ell,a,b}(\delta)$ is the smallest constant such that for all intervals $I \in \mathcal{P}(\delta^a), I' \in \mathcal{P}(\delta^b)$ with $\operatorname{dist}(I, I') \geq 1/4$ such that we have

$$\int_{\mathbb{R}^k} (|f_I|^{p_l} * \varphi_I)(|f_{I'}|^{p_k - p_l} * \varphi_{I'}) \leq B_{\ell,a,b}(\delta)^{p_k} \left(\sum_{J \in \mathcal{P}(I,\delta)} \|f_J\|_{p_k}^2 \right)^{p_l/2} \left(\sum_{J' \in \mathcal{P}(I',\delta)} \|f_{J'}\|_{p_k}^2 \right)^{(p_k - p_l)/2},$$

where we have written

$$f_I = \sum_{J \in \mathcal{P}(I,\delta)} f_J$$

and similarly for $f_{I'}$.

One can see from the definition of these asymmetric constants that they allow us to have a lot more control in exchange for a slightly more complicated object. Also, we know the LHS to be nicely behaved objects by our uncertainty principle. Again though, to get anything out of these asymmetric constants we need to show that they control their symmetric counterparts. We will do this through the following crude inequality:

Lemma 4. With the assumptions of the definition of the asymmetric bilinear decoupling constants, we have that

$$B_k(\delta) \lesssim \delta^{-ap_l/p_k} \delta^{-b(p_k-p_l)/p_k} B_{\ell,a,b}.$$

The proof of the lemma is a slightly tedious computation using Lemma 3.

Another observation is that when $l = 0, B_{l,a,b}(\delta)$ does not depend on a, but we allow this ambiguity to occur to avoid case distinctions. In this case we also have

$$\int_{\mathbb{R}^k} |f_{I'}|^{p_k} * \varphi_{I'} \leq B_{0,a,b}(\delta)^{p_k} \left(\sum_{J' \in \mathcal{P}(I',\delta)} \|f_J\|_{p_k} \right)^{p_k/2},$$

and as the Fourier support of $f_{I'}$ is in $\mathcal{U}_{I'}$ by the uncertainity principle the LHS is essentially just $\|f_{I'}\|_{p_k}^{p_k}$. On the other hand, by the (non-dyadic) rescaling lemma

$$\|f_{I'}\|_{p_k} \leq \mathrm{Dec}_k(\delta/\delta^b) \left(\sum_{J' \in \mathcal{P}(I',\delta)} \|f_{J'}\|_{p_k}^2 \right)^{1/2},$$

and so by the minimality of the asymmetric bilinear decoupling constant we must have

$$B_{0,a,b}(\delta) \sim_p \mathrm{Dec}_k(\delta^{1-b}). \tag{3.3}$$

The next step of our argument, lower dimensional decoupling, is quite intricate, so we will first set the stage. The idea is that any small segment of the moment curve is contained in a thickened arc of a lower-dimensional moment curve. For example, in three dimensions, one can think of the segment $\{(t,t^2,t^3) : t \in [0,\delta]\}$ for some small δ to be contained in a thickening of the arc $\{(t,t^2,0) : t \in [0,\delta]\}$, with the amount that we have to thicken the arc being $O(\delta)$. This structure of the moment curve makes induction on dimension a very natural idea to consider.

Lemma 5 (Lower-dimensional decoupling)**.** Let us assume that Theorem 1 holds for some $l \in \{1,\dots,k-1\}$. Then for $\delta \in (0,1), K \in \mathcal{P}(\delta), f_K$ with Fourier support in $\mathcal{U}_K$ and for $0 \leq a \leq \frac{k-l+1}{l} b, I \in \mathcal{P}(\delta^a), I' \in \mathcal{P}(\delta^b)$ with $\mathrm{dist}(I,I') \geq 1/4$ we have

$$\int_{\mathbb{R}^k} (|f_I|^{p_l} * \varphi_I)(|f_{I'}|^{p_k-p_l} * \varphi_{I'}) \lesssim_\varepsilon \delta^{-b\varepsilon} \left(\sum_{J \in \mathcal{P}(I,\delta^{(k-l+1)b/l})} \left(\int_{\mathbb{R}^k} (|f_J|^{p_l} * \varphi_J)(|f_{I'}|^{p_k-p_l} * \varphi_{I'}) \right)^{2/p_l} \right)^{p_l/2} \tag{3.4}$$

where f_I and $f_{I'}$ are defined as in the definition of the asymmetric bilinear decoupling constants.

Let us illustrate the idea of lower-dimensional decoupling in slightly more detail. For this, recall the projection-slice theorem.

Lemma 6 (Projection-slice theorem)**.** Let $f(\cdot,\cdot) : \mathbb{R}^d \times \mathbb{R}^{k-d} \to \mathbb{C}$ and $f_y : \mathbb{R}^d \to \mathbb{C}, f_y(x) = f(x)$. Let $\mathrm{proj}_{\mathbb{R}^d}$ denote the projection to the first d components. Then

$$\mathrm{supp}\,\widehat{(f_y)} \subset \mathrm{proj}_{\mathbb{R}^d}\,\mathrm{supp}\,\hat{f}.$$

From this we see that if we have some function f_J in $\mathbb{R}^k$ with Fourier support in $\mathcal{U}_J$, then the restriction $f_J|_H$ of f_J to some d-dimensional manifold H has its Fourier support contained the projection of $\mathcal{U}_J$ onto H.

Our plan then to estimate the asymmetric bilinear decoupling constants is as follows: We assume that Theorem 1 holds for some $l \in \{1,\ldots,k-1\}$. Then we will consider the projection of the k-dimensional moment curve to a suitable l-dimensional subspace H and apply our l-dimensional version of Theorem 1 on this subspace. Finally we will slice up $\mathbb{R}^k$ into these subspaces and then combine the l-dimensional decoupling estimate on each to yield the final estimate on the entirety of $\mathbb{R}^k$.

A bit of concreteness is almost always good, so let us write out the details in the above siege plan. The symbols we use, if they are in the statement of Lemma 5, are defined as in the statment of Lemma 5. We have the same inductive hypothesis, that Theorem 1 holds for some l. Define the partial tangent spaces of the moment curve at $\xi' \in I'$ to be $V^j_{\xi'}$ to be the span of the vectors

$$\partial^1\Gamma(\xi'),\ldots,\partial^j\Gamma(\xi'),$$

and let $\hat{H} = \mathbb{R}^k \setminus V^j_{\xi'}$ be the quotient space, so that its Pontryagin dual H is the orthogonal complement of $V^{k-l}(\xi')$ in $\mathbb{R}^k$. Write also P for the projection onto $\hat{H}$. Note then that by the discussion after the projection-slice theorem, for almost all z the restriction $f_J|_{H+z}$ to the translate $H+z$ has Fourier support in the projection of $\mathcal{U}_J$ onto H.

Here we arrive at our first geometric hiccup. To apply a lower-dimensional decoupling inequality, we need to ensure that $\mathcal{U}_{J,P\circ\Gamma}$ is non-degenerate. Consider that if $\mathcal{U}_{J,P\circ\Gamma}$ were degenerate, then its volume must be zero. But the volume of a parallelipiped can be computed as the determinant of the matrix with sides as in the columns of the matrix, and so it would suffice to show that

$$|\partial^1(P\circ\Gamma)(\xi) \wedge \partial^2(P\circ\Gamma)(\xi) \wedge \ldots \wedge \partial^l(P\circ\Gamma)(\xi)| \gtrsim 1$$

for all $\xi \in I$. In fact, GLYZK prove a stronger statement, namely that for the moment curve in k dimensions we have

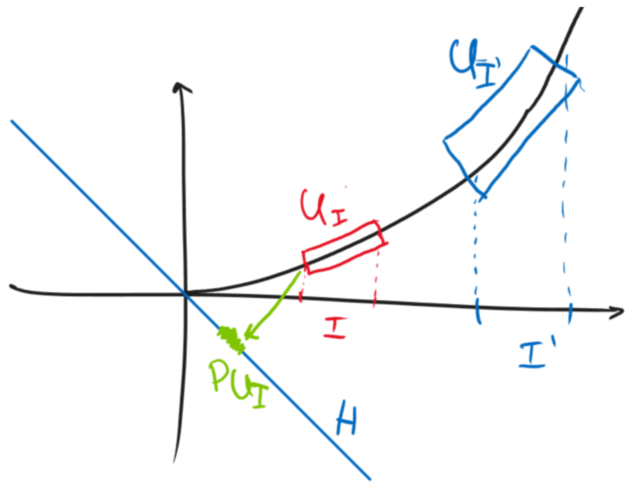

Figure 3: The case when $k = 2$. In blue we have the interval I', the paralellogram $\mathcal{U}_{I'}$ and the subspace H. In red we have the interval I and the parallelogram $\mathcal{U}_I$ respectively, with the green being the projection of $\mathcal{U}_I$ onto H.

Lemma 7. For $0 \leq l \leq k$ and $\xi_1, \xi_2 \in \mathbb{R}$ we have

$$|\partial^1\Gamma(\xi_1) \wedge \partial^2\Gamma(\xi_1) \wedge \ldots \wedge \partial^l\Gamma(\xi_1) \wedge \partial^1\Gamma(\xi_2) \wedge \ldots \wedge \partial^{k-l}\Gamma(\xi_2)| \geq |\xi_1 - \xi_2|^{l(k-l)}.$$

The proof of the lemma follows from writing out the Taylor expansion of $\Gamma(\xi_2)$ at ξ_1 and computing. We leave the details to the paper of GLYZK.

Finally, let us give a numerological explanation on why we decouple in the scale $\delta^{(k-l+1)b/l}$. When we are projecting onto H, we basically forgo the ℓ largest sides of the parallelipiped $\mathcal{U}_{I'}$, so the largest projected side will be of length at most $\delta^{(k-l+1)b}$. This also implies that $\mathcal{U}_{I'}$ is contained inside a ball of radius $\lesssim \delta^{(k-l+1)b}$, which implies that $C\mathcal{U}_{I'}^{\circ}$ (where C is some constant) contains the ball centered at the origin with radius $\delta^{-(k-l+1)b}$.

We can now give a short overview of the proof of Lemma 5. The idea is to in each translate of H, $H + z$ decouple f inside a ball $B_H(z, \delta^{-b'l})$, where $b' = (k - l + 1)b/l$, where by B_H we mean the ball inside the affine subspace $H + z$. Let us state this local decoupling lemma in greater detail.

Lemma 8 (Local decoupling)**.** Given we know Theorem 1 in the case $k = l$, we have the following inequality

$$\fint_B \left| \sum_{J \in \mathcal{P}(\delta)} f_J \right|^{p_l} \lesssim_\varepsilon \delta^{-\varepsilon} \left(\sum_{J \in \mathcal{P}(\delta)} \|f_J\|_{L^{p_l}(\varphi_B)} \right),$$

where φ_B is a normalized bump function on B, e.g.

$$\varphi_B(x) = \frac{1}{|B|}(1 + \delta^l \operatorname{dist}(x, B))^{-10k}.$$

To prove the lemma, we apply decoupling in $\mathbb{R}^l$ to the functions $f_J \psi_B$, where $|\psi| \sim 1$ on B, and the Fourier support of ψ lies in $B(0, \delta^l)$.

Returining to the proof of lower-dimensional decoupling, by Fubini we write

$$\int_{\mathbb{R}^k} (|f_I|^{p_l} * \varphi_I)(|f_{I'}|^{p_k - p_l} * \varphi_{I'}) = \int_{z \in \mathbb{R}^k} \left(\fint_{B_H(z, \delta^{b'l})} (|f_I|^{p_l} * \varphi_I)(|f_{I'}|^{p_k - p_l} * \varphi_{I'}) \right)(z). \tag{3.5}$$

Note though that as $C\mathcal{U}^{\circ}_{I'}$ contains the ball $B_H(0, \delta^{-b'l})$ we can use the uncertainty principle on the function $|f_{I'}|^{p_k - p_l} * \varphi_{I'}$ to find that

$$\sup_{x \in B_H(z, \delta^{-b'l})} (|f_{I'}|^{p_k - p_l} * \varphi_{I'})(x) \lesssim (|f_{I'}|^{p_k - p_l} * \varphi_{I'})(z),$$

and so the RHS of (3.5) is bounded up to a constant factor by

$$\int_{z \in \mathbb{R}^k} (|f_{I'}|^{p_k - p_l} * \varphi_{I'})(z) \left(\fint_{B_H(z, \delta^{b'l})} (|f_I|^{p_l} * \varphi_I) \right)(z).$$

We can think of the integral average as a convolution along H, and then by the commutativity of convolutions we have the equality

$$\left(\fint_{B_H(z, \delta^{b'l})} (|f_I|^{p_l} * \varphi_I \right)(z) = \int_{z' \in \mathbb{R}^k} \varphi_I(z - z') \fint_{B_H(z', \delta^{-b'l})} |f_I|^{p_l}.$$

Now we can apply local decoupling on the average integral term, and finishing from this point requires Minkowski's inequality followed by a few applications of the uncertainty principle.

Note that the lower-dimensional decoupling lemma gives us the following corollary.

$$B_{l,a,b} \lesssim_\varepsilon \delta^{-b\varepsilon} B_{l, \frac{k-l+1}{l} b, b}(\delta). \tag{3.6}$$

This will be the basis of the final stage of the proof, the iteration stage. Note that the corollary basically tells us that it suffices for us to control

$$B_{l, \frac{k-l+1}{l} b, b}(\delta).$$

We want two more estimates for our asymmetric bilinear decoupling constants, first the estimate

$$B_{l,a,b}(\delta) \le B_{k-l,b,a}(\delta)^{\frac{1}{k-l+1}} B_{l-1,a,b}(\delta)^{\frac{k-l}{k-l+1}}, \tag{3.7}$$

which is a straightforward consequence of Hölders inequality and for sufficiently small b,

$$B_{l,\frac{k-l+1}{l}b,b}(\delta) \lesssim_{\varepsilon} \delta^{-b\varepsilon} B_{k-l,\frac{l+1}{l}\frac{k-l+1}{k-l}b,\frac{k-l+1}{l}b}(\delta)^{\frac{1}{k-l+1}} B_{l-1,\frac{k-l+2}{l-1}b,b}(\delta)^{\frac{k-l}{k-l+1}} \tag{3.8}$$

which follows from applying (3.6) on $B_{l,(k-l+1)b/l,b}$ and then using lower-dimensional decoupling. We now have almost all the pieces to show Theorem 1.

As stated previously, we will induct on the dimension k. The case $k=1$ will be our basis case, and a proof of it can be found in [1]. Assume then that Theorem 1 holds for all $l \in \{1,\dots,k-1\}$. Let now η be the infimum of the exponents for which Theorem 1 holds for k and let $A_l(b)$ be the best exponent A such that we have the estimate

$$B_{l,\frac{k-l+1}{l}b,b}(\delta) \lesssim_{\varepsilon} \delta^{-A}.$$

Now by (3.3) we know that $A_0(b) = (1-b)\eta$. We note that using (3.8) we get a recursive estimate for the constants $A_l(b)$ for sufficiently small b as

$$A_l(b) \le \frac{1}{k-l+1} A_{k-l}\left(\frac{k-l+1}{l}b\right) + \frac{k-l}{k-l+1} A_{l-1}(b).$$

We are interested in what happens to the constants $A_l(b)$ as $b \to 0$, i.e. as we go to a larger and a larger scale. Let us therefore define

$$\mathcal{A}_l = \liminf_{b\to 0} \frac{\eta - A_l(b)}{b}.$$

Clearly we have $\mathcal{A}_0 = \eta$, and our recursive estimate takes the form

$$\mathcal{A}_l \ge \frac{1}{k-l+1}\mathcal{A}_{k-l} + \frac{k-l}{k-l+1}\mathcal{A}_{l-1}.$$

To solve this system of linear inequalities, one would have to show that the quantities $\mathcal{A}_l$ are finite, which we leave to the paper of GLYZK. Once this has been established, simply add up this modified recursive estimate over all $l = 1,\dots,k-1$ to get

$$\mathcal{A}_1+\mathcal{A}_2+\ldots+\mathcal{A}_{k-1} \ge \frac{1}{k-1+1}\mathcal{A}_{k-1}+\frac{k-1}{k-1+1}\mathcal{A}_0+\frac{1}{k-1+1}\mathcal{A}_{k-2}+\frac{k-2}{k-2+1}\mathcal{A}_1+\ldots.$$

The RHS is just equal to

$$\frac{k-1}{k}\mathcal{A}_0 + \mathcal{A}_1 + \mathcal{A}_2 + \ldots + \mathcal{A}_{k-1},$$

so we must have $\mathcal{A}_0 \le 0$, which immediately implies $\eta \le 0$ and so η must be equal to 0.

4 Some final remarks

We will leave how decoupling proves Vinogradov's Mean Value Theorem to be found in the text of Pierce [1]. As a final note, I would like to name a few other sources which I used during the writing of this essay, but which I did not use inline citations for. These are the excellent lecture notes of Guth [4] and Hickman [5], and the lecture recordings of Li [6].

References

[1] Lillian B. Pierce. *The Vinogradov Mean Value Theorem [after Wooley, and Bourgain, Demeter and Guth]*. 2020. arXiv: 1707.00119 [math.NT].

[2] Ciprian Demeter. *Fourier Restriction, Decoupling, and Applications*. 2020.

[3] Shaoming Guo et al. "A short proof of ℓ^2 decoupling for the moment curve". In: *American Journal of Mathematics* 143.6 (2021), pp. 1983–1998. DOI: 10.1353/ajm.2021.0048. URL: https://doi.org/10.1353%2Fajm.2021.0048.

[4] *Topics in Analysis: Decoupling*. https://math.mit.edu/~lguth/Math118.html. Accessed: 2023-26-06.

[5] *Modern Developments in Fourier Analysis*. https://www.maths.ed.ac.uk/~jhickman/MDFA.html. Accessed: 2023-25-06.

[6] *Introduction to Decoupling Inequalities*. https://www.youtube.com/watch?v=_ML8jAu5cqo&list=PLQZfZKhc0kiAXeVcTAnYT9_YwpQ13SMB5. Accessed: 2023-07-06.