MULTILINEAR OPERATORS IN HARMONIC ANALYSIS: METHODS AND APPLICATIONS

BY

DONG DONG

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Mathematics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2018

Urbana, Illinois

Doctoral Committee:

Professor M. Burak Erdogan, Chair
Professor Xiaochun Li, Director of Research
Professor Kevin B. Ford
Professor Florin P. Boca

Abstract

We first give a survey on multilinear Hilbert transforms. Then we study several variants of bilinear Hilbert transform such as bilinear Hilbert transform along two polynomials, discrete (integer) bilinear Hilbert transform along polynomials, finite field version of bilinear Radon transform, a hybrid of bilinear Hilbert transform and the paraproduct, etc. Our aim is to find operator norms of these operators: showing that they are finite or have certain decay. Applications in Roth type theorems will also be given.

To my family.

Acknowledgments

I would like to express my deepest gratitude to my advisor, Prof. Xiaochun Li, for his constant support and enthusiastic guidance. Through numerous discussions, he taught me so much math and made me gain confidence in doing research.

Besides my advisor, I would also like to thank the rest of my committee members: Prof. M. Burak Erdogan, Prof. Kevin B. Ford, and Prof. Florin P. Boca, for offering valuable comments during my research.

Special thanks to Xiaochun Li, M. Burak Erdogan, Christoph Thiele, Francesco Di Plinio, Cheelan Bo-Linn, Robert Muncaster and Scott Ahlgren for their support in my job application.

Last but not least, I want to extend my greatest gratitude to my wife Xiumin, my daughter Sharon, my parents and parents-in-law, and many friends who are always offering support and love.

Table of Contents

Chapter 1 Multilinear operators in a 3 by 3 matrix **1**
- 1.1 First column: linear operators 1
- 1.2 Second column: bilinear operators 2
- 1.3 Third column: trilinear operators 3

Chapter 2 Polynomial Roth theorems in finite fields **5**
- 2.1 Roth theorems in finite fields 5
  - 2.1.1 Results 6
  - 2.1.2 Proof of Theorem 2.2 9
  - 2.1.3 Proof of Theorem 2.4 12
- 2.2 A Hörmander type theorem in finite fields 19
  - 2.2.1 Introduction 19
  - 2.2.2 σ-Uniformity and the proof of Theorem 2.15 21
  - 2.2.3 An application 24
- 2.3 A note on Saŕközy's theorem 26

Chapter 3 Discrete bilinear singular Radon transform **28**
- 3.1 Introduction 28
- 3.2 Proof of theorem 3.1 30
- 3.3 Extensions to some arithmetic functions 33
  - 3.3.1 Arithmetic functions with many common values 36
  - 3.3.2 Proof of Theorem 3.5 39

Chapter 4 Bilinear Hilbert transform along two polynomials **43**
- 4.1 Introduction 43
- 4.2 Decomposition and reduction 46
- 4.3 TT^* method 49
- 4.4 σ-uniformity method 52
- 4.5 L^r estimates and the maximal function 56
- 4.6 Necessary conditions 59

Chapter 5 A hybrid of bilinear Hilbert transform and paraproduct **63**
- 5.1 Introduction 63
  - 5.1.1 Background 63
  - 5.1.2 Main results and application 65
  - 5.1.3 Notations 66
- 5.2 Reduction to model form 67
- 5.3 Proof of Theorem 5.2 71
- 5.4 Size estimates 75
- 5.5 Organizing tiles 78
- 5.6 Telescoping 84

References **86**

Chapter 1

Multilinear operators in a 3 by 3 matrix

Boundedness of multilinear operators in harmonic analysis is a problem that demands ideas and techniques from various fields of math. It not only has many applications in PDE and ergodic theory, but also plays a more and more important role in number theory and combinatorics.

My research lies in the study of some variants of Hilbert transforms, with a focus on their applications. In the first Chapter, I will give a brief (and certainly incomplete) survey on Hilbert transforms, including my contributions.

I will explain the colorful diagram and present Hilbert transforms and its eight variants in an organized way. The fundamental problem about these operators is the L^p-boundedness.

1.1 First column: linear operators

Hilbert transform: $H(f)(x) = \int f(x-t)\frac{1}{t}\,dt$

This operator appears in every harmonic/Fourier analysis textbook. Besides its important role in engineering (such as signal processing), it is the prototype of singular integral operators which is a core object of harmonic analysis. The L^p-boundedness of H for $p \in (1,\infty)$ was proved by M. Riesz in 1928 [88].

Hilbert transform along curves: $H_C(f) = \int f(x-\gamma(t))\frac{1}{t}\,dt$

Here $\gamma : \mathbb{R} \to \mathbb{R}^n$ is a "well-behaved" curve. This operator is also called Radon transform. Since 1960s, the L^p-boundedness of H_C has been obtained for various curves and curvature plays an important role in this line of investigations ([97]). Christ-Nagel-Stein-Wainger [17] established the boundedness of H_C in the most general setting.

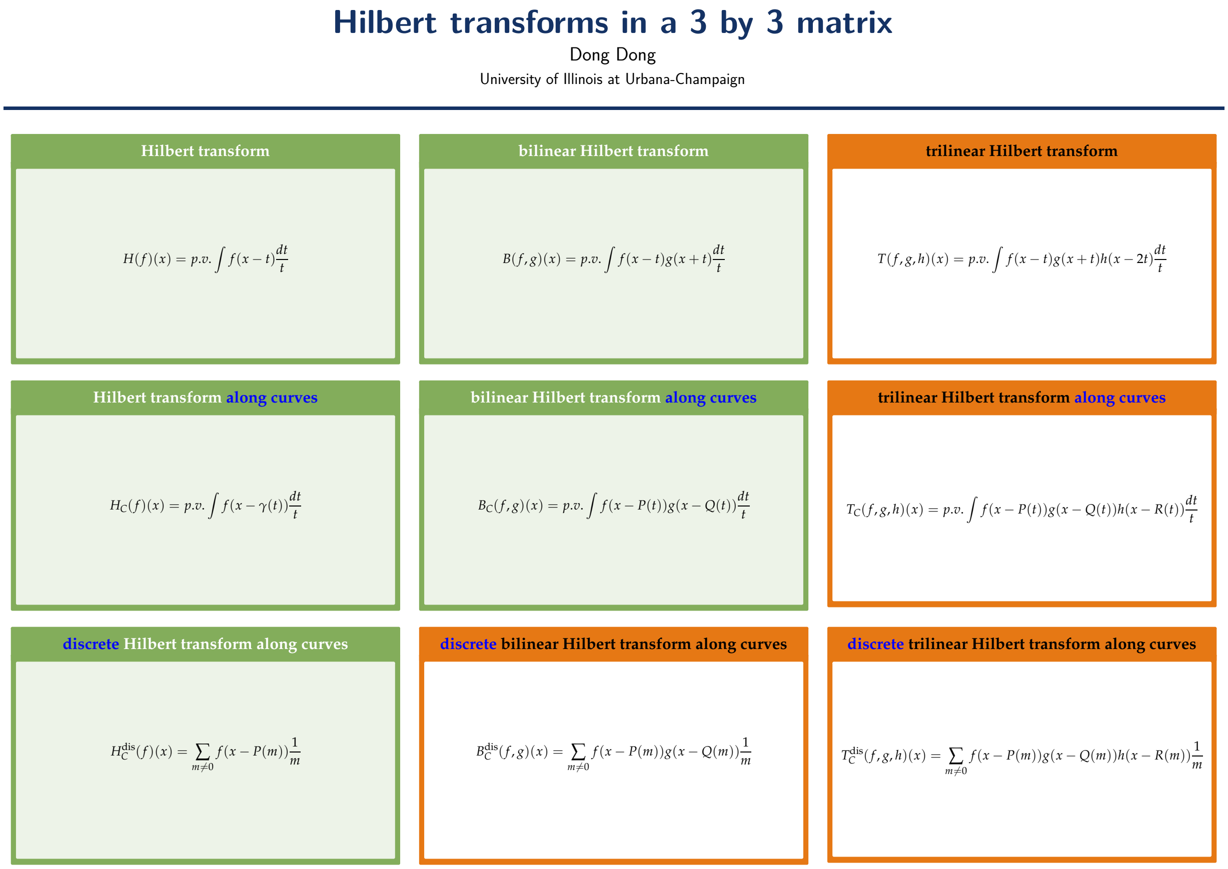

Figure 1.1: The whole picture

Discrete Hilbert transform along curves: $H_C^{\textbf{dis}}(f)(x) = \sum_{m\neq 0} f(x-P(m))\frac{1}{m}$

This discrete analogue of H_C is much more difficult to handle than its continuous counterpart H_C: the boundedness of H_C and some number theoretical tools such as circle method are needed to obtain the boundedness of H_C^{dis} (Ionescu-Wainger [53]). It has many applications in ergodic theory [54].

1.2 Second column: bilinear operators

Bilinear Hilbert transform: $B(f,g)(x) = \int f(x-t)g(x+t)\frac{dt}{t}$

Calderón encountered this bilinear operator in his study of Cauchy integral along Lipschitz curve. The boundedness of B is proved by Lacey and Thiele using time-frequency analysis [60, 61]. This technique originates from Fefferman's proof of Carleson Theorem [33], and has now become a standard tool to deal with operators that are modulation invariant. Open problems about this operator include $L^{\frac{2}{3}}$-boundedness: see [5, 6, 21] for some progress.

Bilinear Hilbert transform along curves: $B_C(f,g)(x) = \int f(x-P(t))g(x-Q(t))\frac{dt}{t}$

This operator is a natural extension of the bilinear Hilbert transform. X. Li [64] first proved the $L^2 \times L^2 \to L^1$-boundedness of B_C for monomial curves: $P(t) = t$ and $Q(t) = t^d$. Together with Xiao, Li later obtained the boundedness of B_C in full range for $P(t) = t$ and Q a general polynomial [65]. Lie [66, 67] and Guo-Xiao [45] considered the case when $P(t) = t$ and Q is a "non-flat" curve. By refining the arguments in [64], I obtained the boundedness of B_C in the case both P and Q are polynomials [23, 24]. An application of B_C in number theory can be found in [29, 27].

Discrete bilinear Hilbert transform along curves:

$B_C^{\mathbf{dis}}(f,g)(x) = \sum_{m\neq 0} f(x-P(m))g(x-Q(m))\frac{1}{m}$

Even the $l^2 \times l^2 \to l^1$ boundedness of this operator is extremely difficult and still open. Using exponential sum estimates, Hu-Li [52] first obtained the $l^2 \times l^2 \to l^{1+\epsilon}$-boundedness ($\epsilon > 0$) when $P(m) = m$ and $Q(m) = m^2$. I proved Hu-Li's result to all polynomials P and Q [22]. This boundedness property is later extended by Meng and I [28] to include the case when P or Q is an arithmetic function such as the Euler totient or the prime counting function (because of the discrete feature of B_C^{dis}, it is natural to consider these arithmetic functions). See [39, 51, 107] for studies on ergodic analogue of this discrete operator.

1.3 Third column: trilinear operators

Trilinear Hilbert transform: $T(f,g,h)(x) = \int f(x-t)g(x+t)h(x-2t)\frac{dt}{t}$

This problem has been open for a long time. A hidden quadratic modulation invariance makes the usual time-frequency analysis arguments ineffective. Muscalu-Tao-Thiele [79] proved the boundedness of T when one L^p space is replaced with a smaller space. Tao [102] used additive combinatorics in an elegant way to obtain a cancellation property of T which, however, does not lead to the boundedness. See [110] for a generalization of Tao's result.

Trilinear Hilbert transform along curves:

$T_C(f,g,h)(x) = \int f(x-P(t))g(x-Q(t))h(x-R(t))\frac{dt}{t}$

As proving boundedness of the trilinear Hilbert transform T seems to be hopeless at the moment, it is natural to study its curved version T_C first. T_C is expected to enjoy a better decay property than T which one could take advantage of. Depending on the efficiency of the oscillatory integral estimate, T_C can be decomposed as a sum of a good part and a bad part. I successfully established the desired bound for the good part [25], based on a joint work with Li [26]. We are actively investigating the bad part, where a generalization of the results in [16] seems to be needed.

Discrete trilinear Hilbert transform along curves:

$T_C^{\mathbf{dis}}(f,g,h)(x) = \sum_{m\neq 0} f(x-P(m))g(x-Q(m))h(x-R(m))\frac{1}{m}$

Based on the research progress in the linear and bilinear operators, we expect that a proof of the boundedness of this discrete trilinear operator requires tools from number theory and the boundedness of its continuous version T_C. Therefore, we should focus on T_C first at the moment.

Remarks The 3×3 matrix can also be read row by row: the first row contains classical operators, the second includes the curved versions and the last row covers the discrete analogues. The operators in green boxes already have satisfactory results, while the boundedness of those operators in orange boxes are still open.

Chapter 2

Polynomial Roth theorems in finite fields

2.1 Roth theorems in finite fields

Let N be a large integer and $A \subseteq \{1, 2, \ldots, N\}$ with density $\delta = \frac{|A|}{N}$. A famous theorem of Roth [89] says that if the density δ is not too small, then A must contain a 3-term arithmetic progression: $x, x+y, x+2y, y \neq 0$. Roth's theorem has been generalized in various directions (e.g. the work of Szemerédi [99, 100], Furstenberg [38], Gowers [40, 41], Bergelson-Leibman [4], Green-Tao [44], Tao-Ziegler [103], etc). It is very interesting that the techniques used in proving these theorems come from many fields such as number theory, ergodic theory, combinatorics and harmonic analysis.

Recently, Bourgain and Chang [11] established a polynomial Roth theorem in finite fields, with surprising quantitative estimates that are not available in the space of integers. More precisely, they proved that for any prime p and $A \subseteq \mathbb{F}_p = \{\bar{0}, \bar{1}, \ldots, \overline{p-1}\}$ with density $\delta = \frac{|A|}{p}$ greater than $p^{-1/15}$, A must contain a triplet of the form $x, x+y, x+y^2$, i.e. a quadratic polynomial progression. Bourgain and Chang then asked several natural questions. For instance, is the the lower bound $p^{-\frac{1}{15}}$ for the density sharp? Can one obtain the same result for polynomial progression $x, x+P(y), x+Q(y)$? Peluse [83] successfully established the polynomial progression case, but with a worse lower bound for the density.

Using harmonic analysis techniques, Li, Sawin and I [27] reduced this number theory problem to an algebraic geometry problem (which is solvable by theories of Deligne [19] and Katz [58]), and fully answered the second question of Bourgain-Chang by not only extending Roth theorem to the polynomial progression case, but also improving the lower bound for the density from $p^{-\frac{1}{15}}$ to $p^{-\frac{1}{12}}$. See the recent work of Peluse [84] for further generalizations.

2.1.1 Results

Fix a large prime p and denote $e_p(x) := e^{2\pi i \frac{x}{p}}$. For any $\varphi_1, \varphi_2 : \mathbb{F}_p \to \mathbb{F}_p$, we are interested in the bilinear average along the "curve" $\Gamma = (\varphi_1, \varphi_2)$: for any $x \in \mathbb{F}_p$,

$$\mathcal{A}_\Gamma(f_1, f_2)(x) := \frac{1}{p} \sum_{y \in \mathbb{F}_p} f_1(x + \varphi_1(y)) f_2(x + \varphi_2(y)) \,. \tag{2.1}$$

The behavior of the bilinear average is closely related to the following exponential sum associated to Γ,

$$K_\Gamma(x, y) := \begin{cases} \frac{1}{p} \sum_{z \in \mathbb{F}_p} e_p(x\varphi_1(z) + y\varphi_2(z)) & y \neq 0; \\ 0 & y = 0. \end{cases} \tag{2.2}$$

To state our main result, we first set up some notation. For $f : \mathbb{F}_p \to \mathbb{C}$, define

$$\mathbb{E}[f] = \mathbb{E}_x[f] = \frac{1}{p} \sum_{x=0}^{p-1} f(x)$$

$$\|f\|_r = \left(\frac{1}{p} \sum_x |f(x)|^r \right)^{\frac{1}{r}}$$

$$\|f\|_{l^r} = \left(\sum_x |f(x)|^r \right)^{\frac{1}{r}}$$

$$\hat{f}(z) = \frac{1}{p} \sum_x f(x) e_p(-xz)$$

With this notation, it is easy to verify that (using the fact that the sum of all p-th roots of unity is 0)

$$\|f\|_r \leq \|f\|_s \text{ if } s > r \quad \text{(a special case of Hölder inequality);}$$

$$\|f\|_2 = \|\hat{f}\|_{l^2} \quad \text{(Parseval)}$$

$$f(x) = \sum_z \hat{f}(z) e_p(xz) \quad \text{(Fourier inversion)}$$

We also need a notion of generalized diagonal sets.

Definition 2.1 *A set $D \subset \mathbb{F}_p \times \mathbb{F}_p$ is called **generalized diagonal** if for any $x \in \mathbb{F}_p$, there are $O(1)$ y's such that $(x, y) \in D$ and for any $y \in \mathbb{F}_p$ there are $O(1)$ x's such that $(x, y) \in D$. The implied constant must be independent of p.*

Our main theorem below provides a framework to obtain decay estimate for the bilinear operator $\mathcal{A}_\Gamma$ associated with various function pairs (φ_1, φ_2). Throughout the paper, $A \lesssim B$ denotes the statement that

$|A| \le C|B|$ for some positive constant C independent of the prime p and the coefficients of polynomials where relevant.

Theorem 2.2 *Let the kernel K_Γ be defined as in (2.2). We define for any $h, y, y' \in \mathbb{F}_p$,*

$$I_\Gamma := \sum_{x\in\mathbb{F}_p} K_\Gamma(x,y)\overline{K_\Gamma(x-h,y+h)K_\Gamma(x,y')}K_\Gamma(x-h,y'+h). \tag{2.3}$$

Suppose that the following three conditions hold:

1. *There exists $\theta \in (0,1]$ such that $\frac{1}{p}\sum_{y\in\mathbb{F}_p} e_p(s\varphi_1(y)) \lesssim p^{-\theta}$ for any $s \neq 0$;*
2. *There exists $\alpha \in \left(\frac{1}{4}, 1\right)$ such that $K_\Gamma(x,y) \lesssim p^{-\alpha}$ for any $x, y \in \mathbb{F}_p$;*
3. *There exists $\beta > 1$ such that for any $h \in \mathbb{F}_p^* := \mathbb{F}_p \backslash \{0\}$ we can find a generalized diagonal set $D_{\Gamma,h}$ so that $I_\Gamma \lesssim p^{-\beta}$ for any $(y,y') \notin D_{\Gamma,h}$.*

Then the bilinear average defined by (2.1) obeys

$$\|\mathcal{A}_\Gamma(f_1,f_2) - \mathbb{E}[f_1]\mathbb{E}[f_2]\|_2 \lesssim p^{-\gamma}\|f_1\|_2\|f_2\|_2 \tag{2.4}$$

with $\gamma = \min\{\theta, \alpha - \frac{1}{4}, \frac{\beta}{4} - \frac{1}{4}\}$.

Motivated by the non-conventional ergodic averages considered by Bergelson [3] and Frantzikinakis and Kra [36], Bourgain and Chang [11] were the first to consider quantitative estimate of the form (2.4). They established (2.4) with $\gamma = \frac{1}{10}$ for the quadratic monomial curve $\Gamma = (y, y^2)$, via an elegant way combining discrete Fourier analysis, explicit evaluation of quadratic Gauss sums and Bombieri's estimate for Weil sums of rational functions [7].

Peluse [83, Theorem 2.2] generalized Bourgain-Chang's result to the polynomial curve $(\varphi_1(y), \varphi_2(y))$ for any linearly independent polynomials φ_1, φ_2. Her result also applies over arbitrary finite fields of large characteristic (and not just $\mathbb{F}_p$). However, she must take $\gamma = 1/16$. Her method is based on careful analysis of the dimension of varieties created by multiple applications of Cauchy-Schwartz, and an exponential sum bound due to Kowalski.

Our result improves the decay rate from $\frac{1}{16}$ to $\frac{1}{8}$ in Peluse's bound. This also improves the decay rate from $\frac{1}{10}$ to $\frac{1}{8}$ in the cases handled by Bourgain and Chang. Moreover, in the special case $\Gamma = (y, y^2)$, our approach does not rely on Bombieri's estimate: When $\Gamma = (y, y^2)$, $K_\Gamma(x,y)$ is a quadratic Gauss sum which can be evaluated explicitly. Condition (3) can therefore be verified by

$$|I_\Gamma| \leq \frac{1}{p^2}\left|\sum_x e_p\left(-\frac{x^2}{4y}\right)e_p\left(\frac{(x-h)^2}{4(y+h)}\right)e_p\left(\frac{x^2}{4y'}\right)e_p\left(-\frac{(x-h)^2}{4(y'+h)}\right)\right| \leq p^{-\frac{3}{2}} \text{ for } y\neq y',$$

using only the quadratic Gauss sum estimate. Hence $\beta=\frac{3}{2}$. It is easy to check that $\theta=\alpha=\frac{1}{2}$, and thus $\gamma=\frac{1}{8}$.

To extend to the polynomial curve $\Gamma=(y,P(y))$, the condition (3) can be verified by Deligne's fundamental work on exponential sums over finite fields [19]. When extending to the bi-polynomial case, we need to use Katz's generalization [58] of Deligne's theorem on exponential sums over smooth affine varieties.

Theorem 2.2 immediately implies a quantitative Roth type theorem:

Corollary 2.3 *Let $\varphi_1,\varphi_2:\mathbb{F}_p\to\mathbb{F}_p$ be functions satisfying conditions* (1), (2) *and* (3) *(with parameters θ, α and β, resp.) in Theorem 2.2. Then for any $A\subset\mathbb{F}_p, |A|=\delta p$ with $\delta>cp^{-\frac{2}{3}\gamma}$, $\gamma=\min\{\theta,\alpha-\frac{1}{4},\frac{\beta}{4}-\frac{1}{4}\}$, there are $\gtrsim\delta^3p^2$ triplets $x,x+\varphi_1(y),x+\varphi_2(y)\in A$.*

We include its short proof (which is the same as that of Corollary 1.2 in [11]) here for the reader's convenience. Indeed, set both f_1 and f_2 to be the indicator function of the set A. By Cauchy-Schwarz inequality and (2.4),

$$\sum_{x,y} f(x)f(x+\varphi_1(y))f(x+\varphi_2(y)) \geq p^2\big(\mathbb{E}[f]^3-\|f\|_2\|\mathcal{A}_\Gamma(f,f)-\mathbb{E}[f]^2\|_2\big)\gtrsim p^2\delta^3, \tag{2.5}$$

from which the corollary follows.

One interesting case of Theorem 2.2 is the following theorem:

Theorem 2.4 *Let $\Gamma=(\varphi_1,\varphi_2)$ with $\varphi_1,\varphi_2\in\mathbb{F}_p[X]$, $\varphi_1(0)=\varphi_2(0)=0$. Suppose that φ_1,φ_2 are linearly independent. Then the average function $\mathcal{A}_\Gamma$ satisfies*

$$\|\mathcal{A}_\Gamma(f_1,f_2)-\mathbb{E}[f_1]\mathbb{E}[f_2]\|_2\lesssim p^{-1/8}\|f_1\|_2\|f_2\|_2, \tag{2.6}$$

with the implied constant depending only on the degrees of φ_1 and φ_2.

As before, we can obtain the corresponding Roth type theorem in which the lower bound $p^{-\frac{1}{12}}$ of δ is slightly better than the bound $p^{-\frac{1}{15}}$ obtained in Bourgain-Chang's paper [11].

Corollary 2.5 *Let $\varphi_1, \varphi_2 \in \mathbb{F}_p[X]$, $\varphi_1(0) = \varphi_2(0) = 0$, be linearly independent. Then for any $A \subset \mathbb{F}_p$, $|A| = \delta p$ with $\delta > cp^{-\frac{1}{12}}$, there are $\gtrsim \delta^3 p^2$ triplets $x, x + \varphi_1(y), x + \varphi_2(y) \in A$.*

Remark 2.6 *The results of this paper can be generalized to an arbitrary finite field $\mathbb{F}_q$ with $q = p^m$. In this general setting, one should be careful that the degree of the polynomial should be coprime to p in order to get the Weil's estimate [12] (and using Deligne-Katz theory). However, as we are usually only interested in the case that p is very large compared with the degrees of the relative polynomials, the coprime condition is automatically satisfied. More precisely, we just need to redefine $e_p(x) := e^{2\pi i \frac{x}{p}}$ to be $e_p(x) := e^{2\pi i \frac{Tr(x)}{p}}$, where Tr is the trace function from $\mathbb{F}_q$ to $\mathbb{F}_p$. All other arguments remain the same.*

Remark 2.7 *Our results can be compared with the recent solution to Cap set problem [30]. On $\mathbb{F}_{3^n}$, the size of a progression-free set is $o(2.756^n)$ by [30]. We can consider the size of a polynomial-progression-free set: Bourgain-Chang's result gives $O((3^{1-1/15})^n) = 2.788^n$, while our result gives $(3^{1-1/12})^n = 2.738^n$. One can also consider the size of progression-free sets in $\mathbb{F}_{p^n}$, which has an order of $(cp)^n$ for some constant $c < 1$ when p is large by [30]. So for p sufficiently large our upper bound for nonlinear progression-free sets beats the lower bound for true progression-free sets.*

Remark 2.8 *Some rational functions could be included in our results. For instance, when $\varphi_1(y) = y$, $\varphi_2(y) = \frac{1}{y}$ (this case is also considered in [11]), we can get the same conclusion as in Theorem 2.4, using Kloosterman sum estimates (Corollary 3.3. in [37]).*

Remark 2.9 *Theorem 2.4, in the case $\varphi_1(x) = x$, implies that the polynomial $x + \varphi_2(y - x)$ is an almost strong asymmetric expander in the sense of Tao's paper [101]. It is possible that this result could also be established using [101, Theorem 3], but we do not pursue this.*

We will prove Theorem 2.2 in the Section 2.1.2. In Section 2.1.3 we will verify the three conditions (1), (2), and (3) for certain polynomial pairs and henceforth prove Theorem 2.4.

2.1.2 Proof of Theorem 2.2

We prove the main theorem in this section. We follow the spirit in the second author's work on the bilinear Hilbert transform along curves in [64]. First, by using Fourier inversion for f_1 and f_2, it is clear that

$$\mathcal{A}_\Gamma(f_1, f_2)(x) = \sum_{n_1, n_2} \hat{f}_1(n_1)\hat{f}_2(n_2) e_p((n_1 + n_2)x) \mathbb{E}_y[e_p(n_1\varphi_1(y) + n_2\varphi_2(y))].$$

Changing variables $n_2 = n, n_1 = s - n$, we then split the bilinear average $\mathcal{A}_\Gamma(f_1, f_2)(x)$ into three terms:

$$\mathcal{A}_\Gamma(f_1, f_2)(x) = J_1 + J_2 + J_3,$$

where

$$\begin{aligned} J_1 &= \hat{f}_1(0)\hat{f}_2(0) = \mathbb{E}[f_1]\mathbb{E}[f_2], \\ J_2 &= \hat{f}_2(0)\sum_{s\neq 0}\left(\hat{f}_1(s)\mathbb{E}_y[e_p(s\varphi_1(y))]\right) e_p(sx), \\ J_3 &= \sum_s \left(\sum_{n\neq 0}\hat{f}_1(s-n)\hat{f}_2(n)\mathbb{E}_y[e_p((s-n)\varphi_1(y)+n\varphi_2(y))]\right) e_p(sx). \end{aligned}$$

By the assumption (1), when $s \neq 0$, we get

$$\mathbb{E}_y[e_p(s\varphi_1(y))] = \frac{1}{p}\sum_y e_p(s\varphi_1(y)) \lesssim \frac{1}{p^\theta}. \tag{2.1}$$

Therefore, using Parseval's identity, triangle inequality, and Hölder inequality, we see that

$$\begin{aligned} &\|\mathcal{A}_\Gamma(f_1, f_2) - \mathbb{E}[f_1]\mathbb{E}[f_2]\|_2 \leq \|\widehat{J_2}\|_{l^2} + \|\widehat{J_3}\|_{l^2} \\ &\lesssim \frac{1}{p^\theta}\|f_1\|_2\|f_2\|_2 + \left(\sum_s \left|\sum_n \hat{f}_1(s-n)\hat{f}_2(n)K_\Gamma(s-n,n)\right|^2\right)^{\frac{1}{2}}, \end{aligned}$$

where K_Γ is given by (2.2).

Set $\gamma_0 = \min\{\alpha - \frac{1}{4}, \frac{\beta}{4} - \frac{1}{4}\}$. Hence it remains to show

$$\sum_s \left|\sum_n \hat{f}_1(s-n)\hat{f}_2(n)K_\Gamma(s-n,n)\right|^2 \lesssim \frac{1}{p^{2\gamma_0}}\|f_1\|_2^2\|f_2\|_2^2. \tag{2.2}$$

Next we choose to employ a TT^* method (Our method and Bourgain-Chang's diverge from here). The left hand side of (2.2) equals

$$\sum_s \sum_{n_1,n_2} \hat{f}_1(s-n_1)\overline{\hat{f}_1(s-n_2)}\hat{f}_2(n_1)\overline{\hat{f}_2(n_2)}K_\Gamma(s-n_1,n_1)\overline{K_\Gamma(s-n_2,n_2)},$$

which, after changing variables $n_1 = v, n_2 = v + h, s = u + v$, can be rewritten as

$$\sum_h \left(\sum_{u,v} F_h(u) G_h(v) K_\Gamma(u,v) \overline{K_\Gamma(u-h, v+h)} \right), \tag{2.3}$$

where

$$F_h(x) = \hat{f}_1(x) \overline{\hat{f}_1(x-h)};$$
$$G_h(x) = \hat{f}_2(x) \overline{\hat{f}_2(x+h)}.$$

When $h = 0$, using condition (2), we see that the inner double sum in (2.3) is bounded by

$$p^{-2\alpha} \|F_0\|_{l^1} \|G_0\|_{l^1} = p^{-2\alpha} \|f_1\|_2^2 \|f_2\|_2^2,$$

which is better than $p^{-2\gamma_0} \|f_1\|_2^2 \|f_2\|_2^2$ as $\alpha > \gamma_0$. Therefore, it remains to handle the case when h is nonzero. The tool is the following bilinear form estimate, which may be interesting on its own right (see [59] for applications of some related bilinear forms).

Proposition 2.10 *Fix $h \neq 0$. Let $\varphi_1, \varphi_2 : \mathbb{F}_p \to \mathbb{F}_p$ satisfy* (2) *and* (3) *(with parameters α and β, resp.) in Theorem 2.2. Let $\gamma_0 = \min\{\alpha - \frac{1}{4}, \frac{\beta}{4} - \frac{1}{4}\}$. Then for any $F, G : \mathbb{F}_p \to \mathbb{C}$,*

$$\sum_{u,v} F(u) G(v) K_\Gamma(u,v) \overline{K_\Gamma(u-h, v+h)} \lesssim \frac{1}{p^{2\gamma_0}} \|F\|_{l^2} \|G\|_{l^2}. \tag{2.4}$$

Once this proposition is proved, one can use (2.4) and apply Cauchy-Schwarz inequality a few times to (2.3) to get the desired estimate (2.2).

By duality, it is easy to see that Proposition 2.10 can be reduced to the following finite field version of Hörmander principle (see Theorem 1.1 in [50] for its continuous counterpart):

Lemma 2.11 *Fix $h \neq 0$. Let $\varphi_1, \varphi_2 : \mathbb{F}_p \to \mathbb{F}_p$ satisfy* (2) *and* (3) *(with parameters α and β, resp.) in Theorem 2.2. Let $\gamma_0 = \min\{\alpha - \frac{1}{4}, \frac{\beta}{4} - \frac{1}{4}\}$. Define an operator*

$$T(g)(x) = \sum_y g(y) K_\Gamma(x,y) \overline{K_\Gamma(x-h, y+h)}.$$

Then

$$\|T(g)\|_{l^2} \lesssim \frac{1}{p^{2\gamma_0}} \|g\|_{l^2}.$$

Proof: *We will show that*

$$\|T(g)\|_{l^2}^2 \lesssim \frac{1}{p^{4\gamma_0}} \|g\|_{l^2}^2. \tag{2.5}$$

A straightforward calculation gives

$$\begin{aligned}\|T(g)\|_{l^2}^2 &= \sum_{x,y,y'} g(y)\overline{g(y')} K_\Gamma(x,y)\overline{K_\Gamma(x-h,y+h)K_\Gamma(x,y')}K_\Gamma(x-h,y'+h) \\ &\leq \sum_{(y,y')\in D_{\Gamma,h}} |g(y)||g(y')||I| + \sum_{(y,y')\notin D_{\Gamma,h}} |g(y)||g(y')||I|, \end{aligned} \tag{2.6}$$

where $D_{\Gamma,h}$ is the generalized diagonal set in condition (3) and

$$I_\Gamma = \sum_x K_\Gamma(x,y)\overline{K_\Gamma(x-h,y+h)K_\Gamma(x,y')}K_\Gamma(x-h,y'+h).$$

We estimate the two terms in (2.6) by different methods. Using the definition of generalized diagonal set and the trivial estimate $I_\Gamma \lesssim \frac{p}{p^{4\alpha}}$ from (2), the first term in (2.6) is estimated by

$$\sum_{(y,y')\in D_{\Gamma,h}} |g(y)||g(y')||I_\Gamma| \lesssim \sum_y |g(y)|^2 \frac{p}{p^{4\alpha}} = \frac{1}{p^{4\alpha-1}}\|g\|_{l^2}^2. \tag{2.7}$$

For the second term in (2.6), we use the assumption $I_\Gamma \lesssim \frac{1}{p^\beta}$ for $(y,y') \notin D_{\Gamma,h}$ and Cauchy-Schwarz inequality to get the estimate

$$\sum_{(y,y')\notin D_{\Gamma,h}} |g(y)||g(y')||I_{\Gamma,h}| \lesssim \frac{\sqrt{p}\sqrt{p}}{p^\beta}\|g\|_{l^2}^2 = \frac{1}{p^{\beta-1}}\|g\|_{l^2}^2. \tag{2.8}$$

Combining (2.7) and (2.8), we obtain

$$\|T(g)\|_{l^2}^2 \lesssim \max\left\{\frac{1}{p^{4\alpha-1}}, \frac{1}{p^{\beta-1}}\right\}\|g\|_{l^2}^2 = \frac{1}{p^{4\gamma_0}}\|g\|_{l^2}^2,$$

which is exactly what we aimed for: (2.5). ■

2.1.3 Proof of Theorem 2.4

To prove Theorem 2.4, first note that we can assume without loss of generality that the two polynomials φ_1 and φ_2 have distinct leading terms. This is because we can rewrite (2.6) in its dual form as

$$|\mathbb{E}_{x,y} f_1(x+\varphi_1(y))f_2(x+\varphi_2(y))f_3(x) - \mathbb{E}[f_1]\mathbb{E}[f_2]\mathbb{E}[f_3]| \lesssim p^{-1/8}\|f_1\|_2\|f_2\|_2\|f_3\|_2, \tag{2.1}$$

and do a change of variable $x \to x + \varphi_1(y)$ on the left-hand-side of (2.1) if necessary (We are indebted to Sarah Peluse for pointing this out).

We will verify that for linearly independent polynomials $\varphi_1, \varphi_2 \in \mathbb{F}_p[X]$ with distinct leading terms, the conditions (1), (2) and (3) in Theorem 2.2 are satisfied with parameters $\theta = \frac{1}{2}$, $\alpha = \frac{1}{2}$ and $\beta = \frac{3}{2}$, resp, and thus prove Theorem 2.4 using Theorem 2.2.

Let d_1 and d_2 denote the degrees of φ_1 and φ_2, resp. Without loss of generality, we assume that $d_1 \leq d_2$.

Conditions (1) and (2) can be verified in the same way, using the well-known square-root cancellation result of Weil [106] (see also [12]). Therefore, $\theta = \alpha = \frac{1}{2}$. Note that the linearly independence of the two polynomials is crucial to obtain (2).

Now we focus on the verification of condition (3). We will from now on write for simplicity that $K = K_\Gamma$ and $I = I_\Gamma$. Recall that for $y \neq 0$,

$$K(x, y) = \frac{1}{p} \sum_{z \in \mathbb{F}_p} e_p(x\varphi_1(z) + y\varphi_2(z)).$$

Plug in the definition of K, put the sum over x innermost, and we see that

$$\begin{aligned}
I &= \sum_{x \in \mathbb{F}_p} K(x, y)\overline{K(x-h, y+h)K(x, y')}K(x-h, y'+h) \\
&= \frac{1}{p^4} \sum_{x} \sum_{z_1, z_2, z_3, z_4} e_p[x\varphi_1(z_1) + y\varphi_2(z_1) - (x-h)\varphi_1(z_2) - (y+h)\varphi_2(z_2) - x\varphi_1(z_3) \\
&\quad - y'\varphi_2(z_3) + (x-h)\varphi_1(z_4) + (y'+h)\varphi_2(z_4)] \\
&= \frac{1}{p^3} \sum_{\substack{z_1, z_2, z_3, z_4 \\ G(z_1, z_2, z_3, z_4) = 0}} e_p(F(z_1, z_2, z_3, z_4)),
\end{aligned}$$

where

$$\begin{aligned}
G(z_1, z_2, z_3, z_4) &= \varphi_1(z_1) - \varphi_1(z_2) - \varphi_1(z_3) + \varphi_1(z_4), \\
F(z_1, z_2, z_3, z_4) &= y\varphi_2(z_1) + h\varphi_1(z_2) - (y+h)\varphi_2(z_2) - y'\varphi_2(z_3) - h\varphi_1(z_4) + (y'+h)\varphi_2(z_4).
\end{aligned}$$

It remains to get the estimate

$$\sum_{\substack{z_1, z_2, z_3, z_4 \\ G(z_1, z_2, z_3, z_4) = 0}} e_p(F(z_1, z_2, z_3, z_4)) \lesssim p^{\frac{3}{2}}. \tag{2.2}$$

We need machinery of algebraic geometry to prove (2.2). To benefit readers who are not very familiar with algebraic geometry, we first prove (2.2) in a simpler case. We assume $\varphi_1(z) = z$, and consequently φ_2 has degree at least 2 by the linearly independence assumption. In this case, the restriction $G(z_1, z_2, z_3, z_4) = 0$ can be dropped once z_4 is replaced with $z_2 + z_3 - z_1$. Therefore, (2.2) is reduced to

$$\sum_{z_1, z_2, z_3} e_p(F(z_1, z_2, z_3, z_2 + z_3 - z_1)) \lesssim p^{\frac{3}{2}}. \tag{2.3}$$

Such character sum is studied by Deligne in his resolution of Weil conjectures:

Theorem 2.12 (Theorem 8.4, [19]) *Let $f \in \mathbb{F}_p[X_1, \ldots, X_n]$ be a polynomial of degree $d \geq 1$. Suppose that d is prime to p, and the projective hypersurface defined by the highest degree homogeneous term f_d is smooth, i.e., the gradient of f_d is non-zero at any point in $\{f_d = 0\} \setminus \{\mathbf{0}\}$. Then*

$$\sum_{z_1, \ldots, z_n} e_p(f(z_1, \ldots, z_n)) \lesssim p^{\frac{n}{2}}.$$

For notational convenience, we write $d = d_2$, the degree of φ_2. Let bz^d denote the leading term of $\varphi_2(z)$. Then the highest degree homogeneous term of $F(z_1, z_2, z_3, z_2 + z_3 - z_1)$ is

$$F_d(z_1, z_2, z_3) = byz_1^d - b(y+h)z_2^d - by'z_3^d + b(y'+h)(z_2 + z_3 - z_1)^d.$$

We need to verify the smoothness $\{F_d = 0\}$. By straightforward calculations, $\nabla F_d = \mathbf{0}$ implies

$$\begin{cases} z_1 = \left(\frac{y'+h}{y}\right)^{\frac{1}{d-1}} (z_2 + z_3 - z_1) \\ z_2 = \left(\frac{y'+h}{y+h}\right)^{\frac{1}{d-1}} (z_2 + z_3 - z_1) \\ z_3 = \left(\frac{y'+h}{y'}\right)^{\frac{1}{d-1}} (z_2 + z_3 - z_1) \end{cases}$$

The above system has nonzero solutions only when

$$\left(\frac{y'+h}{y+h}\right)^{\frac{1}{d-1}} + \left(\frac{y'+h}{y'}\right)^{\frac{1}{d-1}} - \left(\frac{y'+h}{y}\right)^{\frac{1}{d-1}} = 1 \tag{2.4}$$

Put those pairs (y, y') satisfying (2.4) as a set $D_{\Gamma,h}$, and it is not hard to check that $D_{\Gamma,h}$ is generalized diagonal. By Deligne's Theorem, (2.3) holds for any $(y, y') \notin D_{\Gamma,h}$. This finishes the verification of condition (3) with $\beta = \frac{3}{2}$, assuming $\varphi_1(z) = z$.

Now we turn to the general case. In [57], Katz generalizes Deligne's theorem to exponential sums over smooth affine varieties, and in [58], to singular algebraic varieties. We need the following special case of [58, Theorem 4] (The reader could skip its long proof and use it as a "black box" on an early reading of the paper):

Theorem 2.13 *Let $F, G \in \mathbb{F}_p[X_1, \ldots, X_4]$. Assume that the degree of F is indivisible by p, the homogeneous leading term of G defines a smooth projective hypersurface, and the homogeneous leading terms of G and that of F together define a smooth co-dimension-2 variety in the projective space. Then (2.2) holds, i.e.,*

$$\sum_{\substack{z_1,z_2,z_3,z_4 \\ G(z_1,z_2,z_3,z_4)=0}} e_p(F(z_1, z_2, z_3, z_4)) \lesssim p^{\frac{3}{2}}.$$

Proof: We explain in detail how to realize this theorem as a special case of Katz's theorem. We will try to explain this derivation for mathematicians who are not experts in algebraic geometry. (However, Katz's proof requires much more advanced algebraic geometry than we can go into here).

We first restate part of Katz's theorem. Then we will explain Katz's notation and how it applies to our case.

Theorem 2.14 (Katz, Theorem 4 [58]) *Let N and d be natural numbers, let k be a finite field in which d is invertible, let $\psi : k \to \mathbb{C}^{\times}$ be an additive character. Let X be a closed subscheme of $\mathbb{P}^N$ of dimension d. Let L be a section of $H^0(X, \mathcal{O}(1))$ and H a section of $H^0(X, \mathcal{O}(D))$. Let V, f, ϵ, δ be defined as in [58, pp. 878-879]. If assumptions (H1)' and (H2) of [58, pp. 878] hold, and $\epsilon \leq \delta$, then*

$$\left| \sum_{x \in V(k)} \psi(f(x)) \right| \leq C \times (\#k)^{(n+1+\delta)/2}$$

where C is a constant depending only on N, d, and the number and degree of the equations defining X.

We will choose our data so that $k = \mathbb{F}_p$, $V(k) = \{z_1, z_2, z_3, z_4 \in \mathbb{F}_p \mid G(z_1, z_2, z_3, z_4) = 0\}$, $\psi(f(x)) = e_p(F(z_1, z_2, z_3, z_4))$ for $x = (z_1, z_2, z_3, z_4) \in V(k)$, $n = 3$, and $\epsilon = \delta = -1$. Furthermore C will depend only on the degree of F and G.

Examining Katz's bound, and plugging in these statements, it is clear that if we can in fact choose our data in this way, while verifying Katz's conditions, we obtain exactly our stated bound.

In what remains, we will first explain all of Katz's notation that is needed to choose (X, L, H) so that

$$V(k) = \{z_1, z_2, z_3, z_4 \in \mathbb{F}_p \mid G(z_1, z_2, z_3, z_4) = 0\} \text{ and } \psi(f(x)) = e_p(F(z_1, z_2, z_3, z_4)),$$

and second we will verify (H1)' and (H2) and calculate ϵ, δ, explaining more of Katz's notation along the way.

For the first part, because we are interested in the $\mathbb{F}_p$-points $V(\mathbb{F}_p)$ of a scheme V, we will describe schemes mostly by their set of $\mathbb{F}_p$-points (though schemes in fact have more structure than this.) First, we take $N = 4$, so $\mathbb{P}^N = \mathbb{P}^4$ is the space whose $\mathbb{F}_p$-points $\mathbb{P}^4(\mathbb{F}_p)$ are the set of quintuples $(z_1, z_2, z_3, z_4, z_5) \in \mathbb{F}_p$, not all zero, up to multiplication by nonzero scalars. We let $\tilde{G}$ be the homogenization of G, where we add additional powers of z_5 to all the non-leading terms of G to make every term have equal degree. Let X be the vanishing set of $\tilde{G}$, so that $X(\mathbb{F}_p)$ is the subset of $\mathbb{P}^4(\mathbb{F}_p)$ consisting of tuples $(z_1, z_2, z_3, z_4, z_5)$ with $\tilde{G}(z_1, z_2, z_3, z_4, z_5) = 0$. We must choose L as an element of $H^0(X, \mathcal{O}(1))$, which is the space of linear functions in the variables z_1, z_2, z_3, z_4, z_5, and we choose $L = z_5$. Now Katz defines V to be the locus in X where L is nonzero. Hence $V(\mathbb{F}_p)$ is the set of tuples $(z_1, z_2, z_3, z_4, z_5)$, with z_5 nonzero, up to scalar multiplication, that solve the equation $\tilde{G}(z_1, z_2, z_3, z_4, z_5) = 0$. For each such tuple there exists a unique scalar multiplication that sends z_5 to 1, so we can express it equally as the set of tuples (z_1, z_2, z_3, z_4) with $\tilde{G}(z_1, z_2, z_3, z_4, 1) = 0$. By construction, $\tilde{G}(z_1, z_2, z_3, z_4, 1) = G(z_1, z_2, z_3, z_4)$, so $V(\mathbb{F}_p) = \{z_1, z_2, z_3, z_4 \in \mathbb{F}_p \mid G(z_1, z_2, z_3, z_4) = 0\}$, as desired.

Next, because $e_p : \mathbb{F}_p \to \mathbb{C}^\times$ is an additive character, we set $\psi = e_p$. We then need to choose H, a homogeneous form of degree d in the variables z_1, z_2, z_3, z_4, z_5, so that $f(x) = F(z_1, z_2, z_3, z_4)$. Katz defines f as H/L^d. We take d to be the degree of F and H to be the homogenization $\tilde{F}$ of F, just as we did with G. Because we are using the bijection between 4-tuples and 5-tuples that sends (z_1, z_2, z_3, z_4) to $(z_1, z_2, z_3, z_4, 1)$, we need to check that $f(z_1, z_2, z_3, z_4, 1) = F(z_1, z_2, z_3, z_4)$. This follows because

$$f(z_1, z_2, z_3, z_4, 1) = \frac{\tilde{F}(z_1, z_2, z_3, z_4, 1)}{L(z_1, z_2, z_3, z_4, 1)^d} = \frac{\tilde{F}(z_1, z_2, z_3, z_4, 1)}{1^d} = F(z_1, z_2, z_3, z_4).$$

We have therefore shown how to specialize the left side of Katz's bound to the left side of our own bound. It remains to check Katz's assumptions and also the assumptions we made in applying Katz's bound. These are as follows:

1. d is invertible in k.

2. Katz's assumption (H1)' holds.

3. Katz's assumption (H2) holds.

4. $\delta = -1$.

5. $\epsilon = -1$.

6. $n = 3$.

7. C depends only on the degree of F and G.

The first condition, that d is invertible in k, is easy to interpret, as we set $k = \mathbb{F}_p$ and set d to equal the degree of F, so this is equivalent to the degree of F being prime to p, which we have already assumed in the statement of the theorem.

Katz's assumption (H1)' is that X is Cohen-Macauley and equidimensional of dimension $n \geq 1$. Because H is the hypersurface defined by a single equation $\tilde{G} = 0$ in $\mathbb{P}^4$, a smooth variety of dimension 4, it is automatically Cohen-Macauley of dimension 3. This verifies assumptions (2) and (6).

Katz defines C as an explicit function of his numerical data, which consists of N, the number r of equations needed to define X, the degrees of those equations, and d. In our case $N = 4$, $r = 1$, the degree of the unique equation needed to define X is the degree of G, and d is the degree of F. Hence C is some explicit function of those degrees (assumption (7)).

Katz defines ϵ as the dimension of the singular locus of the scheme-theoretic intersection $X \cap L$. For us L is the closed subset of $\mathbb{P}^4$ where $z_5 = 0$. (Katz abuses notation slightly to use L also to refer to the vanishing locus of L.) So $X \cap L$ is the closed subset where $z_5 = 0$ and $\tilde{G} = 0$. Because $z_5 = 0$, we can ignore z_5 and work in $\mathbb{P}^3$ with coordinates z_1, z_2, z_3, z_4. When we do this, because all non-leading monomials of G were multiplied by a positive power of z_5 in $\tilde{G}$, all non-leading monomials become 0 and we are left with just the zero-locus. So $X \cap L$ is the vanishing locus of the leading term of G in $\mathbb{P}^3$, which we assumed in the statement of the theorem is a nonsingular hypersurface, so its singular locus is empty, which by convention Katz assigns dimension -1, verifying $\epsilon = -1$ (assumption (5)).

Katz defines δ as the dimension of the singular locus of the scheme-theoretic intersection $X \cap L \cap H$, and (H2) is his assumption that this has dimension $n-2$. This is the joint vanishing locus of $\tilde{G}, z_5$, and $\tilde{F}$ in $\mathbb{P}^4$, which for the same reason as before is the vanishing locus of the leading terms of F and G in $\mathbb{P}^3$. Because we assumed this is a smooth subscheme of codimension 2, it has dimension $3-2 = n-2$, verifying condition (H2), and its singular locus is empty and has dimension -1, verifying $\delta = -1$ (assumptions (3) and (4)). ■

Now we are ready to prove (2.2) using Theorem 2.13. The first two conditions in the theorem are easy to check. To check the third condition, we handle two cases separately: $d_1 < d_2$ and $d_1 = d_2$.

First assume $d_1 < d_2$. Let az^{d_1} and bz^{d_2} denote the leading term of φ_1 and φ_2, resp. The homogeneous leading term of G and F are

$$G_{d_1}(z_1, z_2, z_3, z_4) := az_1^{d_1} - az_2^{d_1} - az_3^{d_1} + az_4^{d_1},$$

and

$$F_{d_2}(z_1, z_2, z_3, z_4) := byz_1^{d_2} - b(y+h)z_2^{d_2} - by'z_3^{d_2} + b(y'+h)z_4^{d_2},$$

resp. We need to show that the Jacobian matrix

$$J = \begin{bmatrix} \nabla G_{d_1} \\ \nabla F_{d_2} \end{bmatrix} = \begin{bmatrix} d_1 a z_1^{d_1-1} & -d_1 a z_2^{d_1-1} & -d_1 a z_3^{d_1-1} & d_1 a z_4^{d_1-1} \\ d_2 b y z_1^{d_2-1} & -d_2 b(y+h) z_2^{d_2-1} & -d_2 b y' z_3^{d_2-1} & d_2 b(y'+h) z_4^{d_2-1} \end{bmatrix}$$

has full rank at any point in $\{G_{d_1} = F_{d_2} = 0\} \setminus \{\mathbf{0}\}$. When J has rank less than 2, assuming $z_1 z_2 z_3 z_4 \neq 0$, we can solve for each z_i and plug in $G_{d_1} = 0$ to get the equation

$$\left(\frac{1}{y}\right)^{\frac{d_1}{d_2-1}} - \left(\frac{1}{y+h}\right)^{\frac{d_1}{d_2-1}} - \left(\frac{1}{y'}\right)^{\frac{d_1}{d_2-1}} + \left(\frac{1}{y'+h}\right)^{\frac{d_1}{d_2-1}} = 0 \tag{2.5}$$

If one or two of the four variables z_1, z_2, z_3, z_4 are zero, then a new equation can be obtained by deleting the corresponding term(s) in the above equation. The solutions to (2.5) and its variants lie in a generalized diagonal set. So we can apply Theorem 2.13 for pairs (y, y') outside this set.

Secondly consider the case $d_1 = d_2 = d$. The homogeneous leading term of G and F are

$$G_d(z_1, z_2, z_3, z_4) := az_1^d - az_2^d - az_3^d + az_4^d,$$

and

$$F_d(z_1, z_2, z_3, z_4) := byz_1^d - (b(y+h) - ah)z_2^d - by'z_3^d + (b(y'+h) - ah)z_4^d,$$

resp. The Jacobian matrix becomes

$$J = \begin{bmatrix} \nabla G_d \\ \nabla F_d \end{bmatrix} = \begin{bmatrix} da z_1^{d-1} & -da z_2^{d-1} & -da z_3^{d-1} & da z_4^{d-1} \\ dby z_1^{d-1} & -d(b(y+h) - ah) z_2^{d-1} & -dby' z_3^{d-1} & d(b(y'+h) - ah) z_4^{d-1} \end{bmatrix}$$

When $z_1 z_2 z_3 z_4 \neq 0$, J has rank 1 only when

$$by = b(y+h) - ah = by' = b(y'+h) - ah. \tag{2.6}$$

One or two terms in the above equation can be dropped if the corresponding variable is zero. Since we assume that φ_1 and φ_2 have distinct leading terms, $a \neq b$. It is then easy to see that the solutions to (2.6) and its variants form a generalized diagonal set. So Theorem 2.13 applies in most cases, and we are done.

2.2 A Hörmander type theorem in finite fields

Let p be a prime. For any kernel $K : \mathbb{F}_p \times \mathbb{F}_p \to \mathbb{C}$, define a discrete integral operator

$$T(f)(x) = \sum_{y \in \mathbb{F}_p} f(y)K(x,y).$$

We proved that if K satisfy some natural size condition and cancellation condition, the $l^2 - l^2$-operator norm of T is bounded by $p^{-\gamma}$ for some positive γ. This result can be viewed as a discrete analogue of Hörmander theorem. As an application, we recovered a polynomial Roth type theorem in finite fields by Li, Sawin and the author.

2.2.1 Introduction

Notations and results

For suitable $\alpha(x,y)$ and $\phi(x)$, Hörmander [50] proved that the oscillatory integral operator

$$T_N(f)(x) = \int f(y)\alpha(x,y)e^{iN\phi(x,y)}\,dy \tag{2.7}$$

has decaying $L^2 \to L^2$ norm, i.e.,

$$\|T_N(f)\|_{L^2(\mathbb{R})} \ll \frac{1}{N^{\frac{1}{2}}}\|f\|_{L^2(\mathbb{R})}. \tag{2.8}$$

Here $A \ll B$ denotes the statement that $|A| \leq C|B|$ for some constant C independent of N (and p below). This result has many applications in harmonic analysis and differential equations. In this note we aim to establish a discrete (finite field) version of (2.8), which could be useful in some problems in number theory and combinatorics (See Section 2.2.3 for an example).

Let p be an odd prime and $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ denote the prime field with characteristic p. For any function $K : \mathbb{F}_p \times \mathbb{F}_p \to \mathbb{C}$, we define a discrete analogue of the operator T_N in (2.7) by

$$T(f)(x) = \sum_{y \in \mathbb{F}_p} f(y)K(x,y),\ x \in \mathbb{F}_p. \tag{2.9}$$

Our main theorem below provides sufficient conditions (on K) under which T has decaying $l^2 \to l^2$-norm.

Theorem 2.15 *Let $K : \mathbb{F}_p \times \mathbb{F}_p \to \mathbb{C}$ satisfy the size condition:*

$$\|K\|_{l^\infty} \ll \frac{1}{p^a} \tag{2.10}$$

for some $a > 0$. Then for the operator T defined by (2.9) we have

(i) (Cancellation condition of type I) If for each $y \in \mathbb{F}_p$,

$$\left| \sum_{x \in \mathbb{F}_p} K(x,y)\overline{K(x,y')} \right| \ll \frac{1}{p^b} \tag{2.11}$$

holds for all but at most C $y's$, then $\|T\|_{l^2 \to l^2} \ll p^{-\gamma_1}$ with $\gamma_1 = \min\{a, \frac{b}{2}\} - \frac{1}{2}$.

(ii) (Cancellation condition of type II) If for each $x \in \mathbb{F}_p$,

$$\left| \sum_{y \in \mathbb{F}_p} K(x,y)\overline{K(x',y)} \right| \ll \frac{1}{p^b} \tag{2.12}$$

holds for all but at most C $x's$, then $\|T\|_{l^2 \to l^2} \ll p^{-\gamma_2}$ with $\gamma_2 = \min\{a - \frac{1}{4}, \frac{b}{2}\} - \frac{1}{2}$.

Remarks 1. The decay rate γ_1 in part (i) of the theorem is larger than or equal to the decay rate γ_2 in part (ii). However, the two rates are often the same in applications. This is because the estimate for the size of K alone is usually much better (and easier to obtain) than that for the sum of a product in the cancellation condition and thus $\frac{b}{2}$ is usually "much" smaller than a.

2. Part (i) in Theorem 2.15 can be proved by a standard TT^* argument. Indeed,

$$\begin{aligned}
\|T(f)\|_{l^2}^2 &= \sum_x |\sum_y f(y)K(x,y)|^2 = \sum_{y,y'} f(y)\overline{f(y')} \sum_x K(x,y)\overline{K(x,y')} \\
&\ll \sum_y |f(y)|^2 \sum_x |K(x,y)\overline{K(x,y')}| + \sum_{y,y'} |f(y)||f(y')| \frac{1}{p^b} \\
&\ll \|f\|_{l^2}^2 \frac{p}{p^{2a}} + \|f\|_{l^2}^2 \frac{p}{p^b} \ll \frac{1}{p^{\min\{2a,b\}-1}} \|f\|_{l^2}^2.
\end{aligned}$$

Therefore, the main interesting of Theorem 2.15 is part (ii), which will be proved using a σ-uniformity argument.

3. There are some some related problems to consider. One direction is to improve the values of γ_1 and γ_2 or show that they are sharp. Another direction is to view T as a discrete "singular" integral operator and try to find the conditions on K so that T is bounded on l^2. We shall not pursuit these directions here.

An application in number theory

As an application of Theorem 2.15, we give a new proof of the polynomial Roth theorem in finite fields (this is also one of our motivations to introduce the operator T):

Theorem 2.16 (D.-Li-Sawin [27]) *Let $P, Q \in \mathbb{F}_p[X]$, $\varphi_1(0) = \varphi_2(0) = 0$, be linearly independent polynomials. Then for any $A \subseteq \mathbb{F}_p, |A| = \delta p$ with $\delta > c_1 p^{-\frac{1}{12}}$, there are at leasst $c_2\delta^3 p^2$ triplets $x, x+P(y), x+Q(y) \in A$. Moreover, the implied constants c_1 and c_2 depend only on the degrees of P and Q, not on the coefficients of P, Q.*

Motivated by the non-conventional ergodic averages considered by Bergelson [3] and Frantzikinakis and Kra [36], Bourgain and Chang [11] are the first to study non-linear Roth type theorem in finite fields. They proved Theorem 2.16 in the case $P(m) = m, Q(m) = m^2$ and with $\delta > cp^{-\frac{1}{15}}$ [11]. Then Peluse established the above theorem for general polynomial pairs, but with a bigger lower bound for δ [83] and non-uniform implied constants. It is expected that higher order Fourier analysis [41] should be used to establish the 4-term or longer term polynomial progression case of Theorem 2.16 (See [4] for the corresponding theorem in the setting of integers). One purpose of this paper is to demonstrate that another technique, σ-uniformity, is also powerful in studying Roth type theorem. It should be noted that σ-uniformity has been successful in a few Euclidean harmonic analysis problems [16, 45, 64]. To the best of our knowledge, this paper provides its first application in a discrete setting.

2.2.2 σ-Uniformity and the proof of Theorem 2.15

In this section, we first give an abstract version of the σ-uniformity technique, following the treatments by X. Li in [64], and then prove part (ii) of Theorem 2.15.

Let $\mathcal{H}$ be a Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\| \cdot \|$.

Definition 2.17 *Let $\mathcal{Q}$ be a non-empty subset of $\mathcal{H}$ such that the norms of the elements of $\mathcal{Q}$ are uniformly bounded by a constant. For any $\sigma \in (0, \infty]$, we say that an element $f \in \mathcal{H}$ is **σ-uniform in $\mathcal{Q}$** if*

$$| \langle f, q \rangle | \leq \sigma \|f\| \quad \text{for all } q \in \mathcal{Q}. \tag{2.13}$$

The following lemma, which is essentially the same as Theorem 7.1 in [64], will allow us to calculate operator norms more easily by adding additional conditions on testing functions.

Lemma 2.18 *Fix $\sigma > 0$, and a subset $\mathcal{Q}$ of a Hilbert space $\mathcal{H}$. Let $\mathfrak{L}$ be a bounded sublinear map from $\mathcal{H}$ to $\mathbb{C}$. Let S_σ denote the set of all elements in $\mathcal{H}$ that are σ-uniform in $\mathcal{Q}$. Then the operator norm of $\mathfrak{L}$ is* *bounded by*

$$\|\mathfrak{L}\| \leq \max\{U_\sigma, 2\sigma^{-1}Q\},$$

where

$$U_\sigma := \sup_{f \in S_\sigma} \frac{|\mathfrak{L}f|}{\|f\|} \text{ and } Q := \sup_{q \in \mathcal{Q}} |\mathfrak{L}(q)|\,.$$

Proof: We follow closely the proof of Theorem 7.1 in [64]. Denote

$$A_1 := \sup_{f \in S_\sigma^C} \frac{|\mathfrak{L}(f)|}{\|f\|}$$

where S_σ^C is the complement of S_σ in $\mathcal{H}$. It suffices to show that $A_1 \leq 2\sigma^{-1}Q$, provided that $U_\sigma < A_1$.

By the definition of A_1, for any $\epsilon > 0$, there exists $f \in S_\sigma^C$ such that

$$|\mathfrak{L}(f)| \geq (A_1 - \epsilon)\|f\|\,. \tag{2.14}$$

Since $f \in S_\sigma^C$, there is $q \in \mathcal{Q}$ with

$$|\langle f, q\rangle| > \sigma\|f\|\,. \tag{2.15}$$

Decompose f as

$$f = \langle f, g\rangle\, g + \frac{\langle f, q\rangle}{\|q\|^2} q\,, \tag{2.16}$$

for some $g \in \mathcal{H}$ with $g \perp q$ and $\|g\| = 1$. Since $\mathfrak{L}$ is sublinear,

$$|\mathfrak{L}(f)| \leq |\langle f, g\rangle|\,|\mathfrak{L}(g)| + \frac{|\langle f, q\rangle|}{\|q\|^2}|\mathfrak{L}(q)|. \tag{2.17}$$

Note that $\|\mathfrak{L}\| = A_1$ if $U_\sigma < A_1$, and by orthogonality we also have

$$\|f\|^2 = |\langle f, g\rangle|^2 + \frac{|\langle f, q\rangle|^2}{\|q\|^2}\,. \tag{2.18}$$

Combine (2.14) and(2.17), and we see that

$$(A_1 - \epsilon)\|f\| \leq A_1 \|g\| \|f\| \sqrt{1 - \frac{|\langle f, q \rangle|^2}{\|f\|^2 \|q\|^2}} + \frac{|\langle f, q \rangle|}{\|q\|^2} Q\,. \tag{2.19}$$

Applying the inequality $\sqrt{1-x} \leq 1 - x/2$ if $x \leq 1$, we then get

$$\begin{aligned} A_1 &\leq \frac{2\epsilon \|f\|^2 \|q\|^2}{|\langle f, q \rangle|^2} + \frac{2\|f\|}{|\langle f, q \rangle|} Q \\ &\leq 2\epsilon\sigma^{-2}\|q\|^2 + 2\sigma^{-1} Q\,, \end{aligned}$$

where the second inequality follows from (2.15). The above inequality holds for every $\epsilon > 0$ and some element $q = q(\epsilon) \in \mathcal{Q}$. Now let ϵ approach 0. Since the norms of the elements of $\mathcal{Q}$ are uniformly bounded by a constant, we get $A_1 \leq 2\sigma^{-1}Q$, as desired. ■

Now we turn to the proof of part (ii) in Theorem 2.15, using Lemma 2.18. Let $\mathcal{H}$ be the space of complex-valued functions on $\mathbb{F}_p$ equipped with l^2-norm and standard inner product

$$\langle f, g \rangle = \sum_{x \in \mathbb{F}_p} f(x)\overline{g(x)}.$$

Define $\mathfrak{L} : \mathcal{H} \to \mathbb{R}$ by $\mathfrak{L}(f) = \|T(f)\|_{l^2}$ where $T(f)(x) = \sum_{y \in \mathbb{F}_p} f(y)K(x,y)$. Let σ be a positive number to be determined later. Let

$$\mathcal{Q} = \{\overline{K(x,\cdot)} : x \in \mathbb{F}_p\}.$$

We first calculate U_σ. Note that

$$\begin{aligned} S_\sigma &= \{f : f \text{ is } \sigma\text{-uniform in } \mathcal{Q}\} \\ &= \{f : |\langle f, \overline{K(x,\cdot)} \rangle| \leq \sigma \|f\|_{l^2},\ \text{for any } x \in \mathbb{F}_p\} \\ &= \{f : |T(f)(x)| \leq \sigma \|f\|_{l^2} \text{ for any } x \in \mathbb{F}_p\} \end{aligned}$$

Therefore, for any $f \in S_\sigma$,

$$\mathfrak{L}(f) = (\sum_x |T(f)(x)|^2)^{\frac{1}{2}} \leq \sigma\sqrt{p}\|f\|_{l^2},$$

and thus

$$U_\sigma = \sup_{f \in S_\sigma} \frac{|\mathfrak{L} f|}{\|f\|} \leq \sigma p^{\frac{1}{2}}. \tag{2.20}$$

Next we estimate $Q = \sup_{q \in \mathcal{Q}} |\mathfrak{L}(q)| = \sup_{x \in \mathbb{F}_p} \|T(\overline{K(x,\cdot)})\|_{l^2}$. For any $x \in \mathbb{F}_p$,

$$\begin{aligned}
\|T(\overline{K(x,\cdot)})\|_{l^2}^2 &= \sum_{x' \in \mathbb{F}_p} |\sum_y \overline{K(x,y)} K(x',y)|^2 \\
&= \sum_{\text{at most } C\ x'} |\sum_y \overline{K(x,y)} K(x',y)|^2 + \sum_{\text{all but at most } C\ x'} \frac{1}{p^{2b}} \\
&\ll \left(\frac{p}{p^{2a}}\right)^2 + \frac{p}{p^{2b}} \ll \frac{1}{p^{\min\{4a-2, 2b-1\}}}.
\end{aligned}$$

So we have

$$Q \ll \frac{1}{p^{\min\{4a-2, 2b-1\}}}. \tag{2.21}$$

Applying Lemma 2.18 and using (2.20) and (2.21), we obtain

$$\|T\|_{l^2 \to l^2} = \|\mathfrak{L}\| \ll \max\{U_\sigma, \frac{Q}{\sigma}\} \ll \max\{\sigma p^{\frac{1}{2}}, \frac{1}{\sigma p^{\min\{4a-2, 2b-1\}}}\}$$

Choose $\sigma = \frac{1}{p^{\min\{a-\frac{1}{4}, \frac{b}{2}\}}}$ and we get

$$\|T\|_{l^2 \to l^2} \ll \frac{1}{p^{\min\{a-\frac{1}{4}, \frac{b}{2}\} - \frac{1}{2}}}.$$

This finishes the proof of Theorem 2.15 part (ii).

2.2.3 An application

We give a new proof of a main result in [27] using Theorem 2.15 (ii). In [27], with the help of Fourier analysis and exponential sum estimates obtained by Weil [106], Deligne [19, 20] and Katz [57, 58], Theorem 2.16 (Corollary 1.4 in [27]) has been reduced to the following proposition

Proposition 2.19 (Proposition 2.1 [27]) *Let* $K : \mathbb{F}_p \times \mathbb{F}_p \to \mathbb{C}$ *satisfy*

$$|K(x,y)| \ll \frac{1}{p} \text{ for any } x,y; \tag{2.22}$$

$$\text{For any } y, \text{there are at most } C\ y' \text{such that } |\sum_x K(x,y)\overline{K(x,y')}| \ll \frac{1}{p^{\frac{3}{2}}}. \tag{2.23}$$

Then

$$\left|\sum_{x,y} f(x)g(y)K(x,y)\right| \ll \frac{1}{p^{\frac{1}{4}}}\|f\|_{l^2}\|g\|_{l^2} \tag{2.24}$$

for any functions f and g defined on $\mathbb{F}_p$.

The bilinear form in (2.24) is very important in analytic number theory (for example, see [59] for some applications), and thus it is useful to have different methods to estimate this bilinear form. Proposition 2.19 was proved in [27] via a TT^* argument (which is essentially part (i) of Theorem 2.15). We will show that (2.24) also follows from Theorem 2.15 (ii).

To establish (2.24), we apply Cauchy-Schwarz to the left-hand-side:

$$\left|\sum_{x,y} f(x)g(y)K(x,y)\right| = \left|\sum_y g(y)\sum_x f(x)K(x,y)\right| \le \|g\|_{l^2}\left\|\sum_x f(x)K(x,y)\right\|_{l^2}.$$

Therefore, it suffices to show that the linear operator

$$\tilde{T}(f)(x) = \sum_y f(y)K(y,x)$$

satisfies the estimate

$$\|\tilde{T}(f)\|_{l^2} \ll \frac{1}{p^{\frac{1}{4}}}\|f\|_{l^2}. \tag{2.25}$$

To apply Theorem 2.15 (ii), rewrite $\tilde{T}$ as

$$\tilde{T}(f)(x) = \sum_y f(y)\tilde{K}(x,y),$$

where $\tilde{K}(x,y) = K(y,x)$. From (2.22), the size condition in the theorem is satisfied with $a = 1$. Cancellation condition of type II also holds with $b = \frac{3}{2}$ by (2.23) and the fact that

$$|\sum_y \tilde{K}(x,y)\overline{\tilde{K}(x',y)}| = |\sum_y K(y,x)\overline{K(y,x')}|.$$

Apply Theorem 2.15 (ii) to $\tilde{T}$, and we see that $\|\tilde{T}\|_{l^2 \to l^2} \ll \frac{1}{p^\gamma}$, with $\gamma = \min\{a - \frac{1}{4}, \frac{b}{2}\} - \frac{1}{2} = \frac{1}{4}$, which is exactly (2.25).

2.3 A note on Saŕközy's theorem

We can also consider the existence of 2-term polynomial progressions in subsets of a finite field. In the integer setting, this kind of result is usually called Saŕközy's Theorem, which was first proved independently by Saŕközy [90] and Furstenberg [38] (Saŕközy's result is quantitative). See [2, 47, 68, 69, 70, 86, 92] for more references about ongoing investigation on this topic.

Let $P \in \mathbb{F}_p[x]$ be a polynomial without constant term. Define

$$F(x) = \frac{1}{p} \sum_{y \in \mathbb{F}_p} f(x - P(y)).$$

Then we have a sharp power-saving estimate

Theorem 2.20

$$\|F - Ef\|_2 \le p^{-\frac{1}{2}} \|f\|_2 \tag{2.26}$$

holds for any $f : \mathbb{F}_p \to \mathbb{C}$. Moreover, this is an equality when $P(y) = y^2$ and $Ef = 0$.

Proof: As in the bilinear version of F, we can use Fourier transform to rewrite F as

$$F(x) = \hat{f}(0) + \sum_{\xi} \hat{f}(\xi) K(\xi) e_p(\xi x),$$

where the kernel K satisfies $K(0) = 0$ and

$$K(\xi) = \frac{1}{p} \sum_{y} e_p(-\xi P(y)), \ \xi \neq 0.$$

By Parseval and the square-root cancellation estimate for K,

$$\|F - Ef\|_2 = \|\hat{f} K\|_{l^2} \lesssim \frac{1}{\sqrt{p}} \|f\|_2.$$

When $P(y) = y^2$, the l^∞ norm of K is exactly $\frac{1}{\sqrt{p}}$. Therefore, the above inequality becomes equality when $Ef = 0$. ∎

Corollary 2.21 *Let $P \in \mathbb{F}_p[X]$ be a polynomial without constant term. Then for any $A \subset \mathbb{F}_p, |A| = \delta p$ with $\delta > cp^{-\frac{1}{2}}$, there are $\gtrsim \delta^2 p^2$ pairs $x, x + P(y) \in A$ with $y \neq 0$.*

Proof: Given $A \subset \mathbb{F}_p$ with density $\delta \gtrsim p^{-\frac{1}{2}}$, let $f = \chi_A$. Then $Ef = \delta$ and $\|f\|_2 = \delta^{\frac{1}{2}}$.

$$\frac{1}{p^2} \sum_x \sum_y f(x) f(x + P(y)) = E(fF) = E(f(F - Ef) + fEF)$$
$$\geq (EF)^2 - \|f(F - EF)\|_1 \geq (EF)^2 - c\|f\|_2 p^{-\frac{1}{2}} \|f\|_2 = \delta^2 - cp^{-\frac{1}{2}}\delta \gtrsim \delta^2.$$

Note that this lower bound is larger than the trivial lower bound $\frac{\delta}{p}$ (simply sum over x when $y = 0$). ■

We are interested in how the result in the corollary can be transformed to the integer setting $\{1, 2, \ldots, p\}$ with a possible worse lower bound for the density. This will be a future direction of study.

Chapter 3

Discrete bilinear singular Radon transform

This chapter presents my work on the boundedness of a discrete bilinear Hilbert transform [22] and joint work with Meng [28]. We prove that for a large class of functions P and Q, the discrete bilinear operator $T_{P,Q}(f,g)(n) = \sum_{m\in\mathbb{Z}\setminus\{0\}} f(n-P(m))g(n-Q(m))\frac{1}{m}$ is bounded from $l^2 \times l^2$ into $l^{1+\epsilon,\infty}$ for any $\epsilon \in (0,1]$.

3.1 Introduction

Recall that the Hilbert transform (HT for short) is defined by

$$H(f)(x) = \int f(x-t)\,\frac{dt}{t},\ f \in \mathcal{S}(\mathbb{R}),$$

where $\mathcal{S}(\mathbb{R}^n)$, $n \in \mathbb{N}$, is the Schwartz space on $\mathbb{R}^n$. It was proved in 1928 ([88]) that HT is bounded on L^p for $p \in (1,\infty)$. An interesting generalization of HT is the so-called HT along curves:

$$H_C(f)(x) = \int f(x-\gamma(t))\frac{dt}{t}, f \in \mathcal{S}(\mathbb{R}^n).$$

Here $\gamma : \mathbb{R} \to \mathbb{R}^n$ is a well-behaved curve. The L^p boundedness of H_C has been obtained for various curves. See [97] for a comprehensive survey and [17] for a generalization of H_C to the non-translation-invariant setting. When γ is a polynomial with integer coefficients, there is a discrete version of H_C defined by

$$H_C^{\text{dis}}(f)(n) = \sum_{m\in\mathbb{Z}\setminus\{0\}} f(n-\gamma(m))\frac{1}{m},\ f \in D(\mathbb{Z}^n),$$

where $D(\mathbb{Z}^n)$ is the space of compactly supported complex-valued functions defined on $\mathbb{Z}^n$. On the one hand H_C^{dis} has many applications in ergodic theory ([54, 73, 74, 75, 77]), but one the other hand this discrete operator is more subtle to handle than its continuous counterpart H_C, as many number theoretical tools are involved. H_C^{dis} was at first proved to be bounded on l^p for $p \in (\frac{3}{2}, 3)$ ([98]). This restricted range was extended to the full range $(1,\infty)$ a long time later ([53, 72]).

Another direction of generalizing HT is to consider its bilinear analogue, which is significantly more difficult to analyze since Plancherel Theorem is unavailable in the bilinear setting. The bilinear Hilbert transform (BHT for short) can be defined as

$$B(f,g)(x) = \int f(x-t)g(x+t)\,\frac{dt}{t},\ \ f,g \in \mathcal{S}(\mathbb{R}).$$

It was about 70 years after the first proof of the boundedness of HT that Lacey and Thiele ([60, 61]) obtained the L^p estimates for BHT. Very recently, L^p estimates for BHT along curves

$$B_C(f,g)(x) = \int f(x-t)g(x-\gamma(t))\,\frac{dt}{t},\ \ f,g \in \mathcal{S}(\mathbb{R}),$$

were also established when γ is a polynomial ([65]). Note that B_C is a natural bilinear version of H_C.

Following the development of the linear case, in this paper we consider the discrete version of B_C, that is,

$$B_C^{\rm dis}(f,g)(n) = \sum_{m\in\mathbb{Z}\setminus\{0\}} f(n-m)g(n-P(m))\frac{1}{m},\ \ f,g \in D(\mathbb{Z}),$$

where P is a polynomial with integer coefficients. This operator can also be viewed as a bilinear analogue of $H_C^{\rm dis}$. As $H_C^{\rm dis}$ is harder to handle than H_C, it is reasonable to expect that proving boundedness of $B_C^{\rm dis}$ should be more difficult than that of B_C. As a starting point of the long journey of investigation on $B_C^{\rm dis}$, in this article we show the $l^2\times l^2 \to l^{1+\epsilon,\infty}$ boundedness of $B_C^{\rm dis}$ (Theorem 4.2).

We will study an operator that is more general than $B_C^{\rm dis}$ (see (3.1)). Given two functions P and Q that map $\mathbb{Z}$ into $\mathbb{Z}$, define

$$A^{P,Q} := \left\{(m_1,m_2)\in(\mathbb{Z}\setminus\{0\})^2 : P(m_1)-Q(m_1) = P(m_2)-Q(m_2)\right\}.$$

We say that the pair of functions (P,Q) satisfies **condition** $(\star)$ if there are constants D_1 and D_2 such that $\frac{|m_1|}{|m_2|} \le D_1$ for all $(m_1,m_2)\in A^{P,Q}$ and for each $m_1\in\mathbb{Z}$, there are at most D_2 pairs (m_1,m_2) in the set $A^{P,Q}$.

Theorem 3.1 *Given two functions P and Q that map $\mathbb{Z}$ into $\mathbb{Z}$, let*

$$T_{P,Q}(f,g)(n) := \sum_{m\in\mathbb{Z}\setminus\{0\}} f(n-P(m))g(n-Q(m))\frac{1}{m},\ \ f,g\in D(Z). \tag{3.1}$$

Assume that (P,Q) satisfies condition $(\star)$. Then for any $\epsilon\in(0,1]$, there is a constant C_ϵ depending only on

ϵ, D_1 and D_2 such that

$$\|T_{P,Q}(f,g)\|_{l^{1+\epsilon,\infty}} \leq C_\epsilon \|f\|_{l^2}\|g\|_{l^2}. \tag{3.2}$$

Remarks. (1). Condition ($\star$) is mild. A pair of polynomials with integer coefficients (P,Q) satisfies condition ($\star$) as long as $P-Q$ is not constant. Note that D_1 depends on the coefficients of P and Q, so does C_ϵ in the theorem. It is natural to expect that this dependence can be removed, as uniform estimates exist for related operators ([42, 43, 62, 63, 65, 98, 104]). We shall not pursuit this here.

(2). We conjecture that at least for some special pairs of P and Q (for example, $P(t)=t$ and $Q(t)=t^2$), $T_{P,Q}$ is bounded from $l^p \times l^q$ into l^r, where $p,q \in (1,\infty)$, $\frac{1}{p}+\frac{1}{q}=\frac{1}{r}$. This problem is very difficult and currently out of reach.

(3). A useful operator related with $T_{P,Q}$ is the corresponding maximal operator

$$T^*_{P,Q}(f,g)(n) = \sup_{M\in[1,\infty)} |\frac{1}{M}\sum_{m=1}^{M} f(n-P(m))g(n-Q(m))|,$$

which is at first proved to be bounded from $l^2 \times l^2$ to l^r when $r>1$ ([52]). By using Hölder inequality and boundedness of the corresponding discrete linear maximal function $f \to \sup_{M\in[1,\infty)]} \frac{1}{M}\sum_{m=1}^{M} |f(n-P(m))|$ (see, for example, [10, 56, 76, 109]), we can prove that $T^*_{P,Q}$ is bounded from $l^p \times l^q$ into l^r, whenever $p,q \in (1,\infty)$, $\frac{1}{p}+\frac{1}{q}=\frac{1}{r}$, and $r>1$ (see p.75 in [105] for a similar trick). Whether the restriction $r>1$ can be dropped is still unknown.

(4). See [64] for a discussion about an ergodic analogue of $T_{P,Q}$.

The rest of paper is devoted to the proof of Theorem 3.1.

3.2 Proof of theorem 3.1

We will use $A \lesssim B$ to denote the statement that $A \leq CB$ for some positive constant C. When the implied constant C depends on r, we write $A \lesssim_r B$. All the constants may depend on D_1 and D_2 (appeared in the definition of condition ($\star$)), but this dependence will be suppressed since D_1 and D_2 are often fixed in applications. $A \simeq B$ is short for $A \lesssim B$ and $B \lesssim A$. For any set of integers E, $|E|$ and χ_E will be used to denote the counting measure and the indicator function of E, receptively.

Let P and Q be a pair of functions satisfying condition ($\star$). For notational convenience, we will simply write T for $T_{P,Q}$ and $r := 1+\epsilon$. For any $\lambda > 0$ and $f, g \in D(\mathbb{Z})$, define the level set $E_\lambda = \{n \in \mathbb{Z} :$

$|T(f,g)(n)| > \lambda\}$. Our goal is to prove the following the level set estimate

$$|E_\lambda| \lesssim_r \frac{1}{\lambda^r}, \text{ whenever } \|f\|_{l^2} = \|g\|_{l^2} = 1. \tag{3.3}$$

We first write $T = \sum_{m\in\mathbb{Z}\setminus\{0\}} f(n-P(m))g(n-Q(m))\frac{1}{m}$ as a bilinear multiplier operator. Recall the Fourier transform for any $f \in D(\mathbb{Z})$ is defined by $\hat{f}(\xi) := \sum_{m\in\mathbb{Z}} f(m)e^{-2\pi i\xi m}$. Hence

$$T(f,g)(n) = \int_{\mathbf{T}}\int_{\mathbf{T}} \hat{f}(\xi)\hat{g}(\eta)e^{2\pi i(\xi+\eta)n}\sigma(\xi,\eta)\,d\xi d\eta,$$

where $\mathbf{T}$ is the unit circle and σ is the periodic multiplier (a.k.a symbol) given by

$$\sigma(\xi,\eta) = \sum_{m\in\mathbb{Z}\setminus\{0\}} \frac{1}{m}e^{-2\pi i(P(m)\xi+Q(m)\eta)}.$$

Then we decompose dyadically the symbol σ as follows. Pick an odd function $\rho \in \mathcal{S}(\mathbb{R})$ supported in the set $\{x : |x| \in (\frac{1}{2}, 2)\}$ with the property that

$$\frac{1}{x} = \sum_{j=0}^{\infty} \frac{1}{2^j}\rho\left(\frac{x}{2^j}\right) \text{ for any } x \in \mathbb{R} \text{ with } |x| \geq 1.$$

So the symbol σ can be written as $\sigma(\xi,\eta) = \sum_{j=0}^{\infty}\sigma_j(\xi,\eta)$, where

$$\sigma_j(\xi,\eta) := \frac{1}{2^j}\sum_{m\in\mathbb{Z}}\rho\left(\frac{m}{2^j}\right)e^{-2\pi i(P(m)\xi+Q(m)\eta)}.$$

Correspondingly $T = \sum_{j=0}^{\infty} T_j$, where

$$\begin{aligned} T_j(f,g)(n) &= \int_{\mathbf{T}}\int_{\mathbf{T}} \hat{f}(\xi)\hat{g}(\eta)e^{2\pi i(\xi+\eta)n}\sigma_j(\xi,\eta)\,d\xi d\eta \\ &= \frac{1}{2^j}\sum_{m\in\mathbb{Z}}\rho\left(\frac{m}{2^j}\right)f(n-P(m))g(n-Q(m)). \end{aligned}$$

By the support of ρ and Hölder inequality, it is easy to see $\|T_j(f,g)\|_{l^1} \lesssim \|f\|_{l^2}\|g\|_{l^2}$. So we have the following level set estimate for each T_j.

Lemma 3.2 *For any $f, g \in D(\mathbb{Z})$ with l^2-norm 1, $j \in \mathbb{N}$, and $\lambda > 0$, we have*

$$|\{n \in \mathbb{Z} : |T_j(f,g)(n)| > \lambda\}| \lesssim \frac{1}{\lambda}$$

This lemma says that each single T_j is under good (and uniform) control. The difficulty is how to get desired estimates for the sum of T_j's. In the following we will apply the idea of TT^* method.

Define an auxiliary function $h(n) = \frac{\overline{T(f,g)(n)}}{|T(f,g)(n)|}\chi_{E_\lambda}(n)$. It is easy to verify that

$$\lambda^2 |E_\lambda|^2 \leq \left(\sum_{n \in \mathbb{Z}} T(f,g)(n) h(n) \right)^2. \tag{3.4}$$

By Fubini theorem and the definition of Fourier transform,

$$\begin{aligned}\sum_{n \in \mathbb{Z}} T(f,g)(n)h(n) &= \int_{\mathbf{T}} \int_{\mathbf{T}} \hat{f}(\xi)\hat{g}(\eta)\sigma(\xi,\eta)\hat{h}(-(\xi+\eta))\, d\xi d\eta \\ &= \int_{\mathbf{T}} \int_{\mathbf{T}} \hat{f}(\xi-\eta)\hat{g}(\eta)\sigma(\xi-\eta,\eta)\hat{h}(-\xi)\, d\xi d\eta\end{aligned}$$

Apply 2-dim Cauchy-Schwarz inequality to separate $\hat{f}(\xi-\eta)\hat{g}(\eta)$ and $\sigma(\xi-\eta,\eta)\hat{h}(-\xi)$ and then invoke change of variable and Plancherel Theorem. Then we get

$$\left(\sum_{n \in \mathbb{Z}} T(f,g)(n) h(n) \right)^2 \leq B|E_\lambda|, \tag{3.5}$$

where

$$B := \sup_{\xi \in \mathbf{T}} \int_{\mathbf{T}} |\sigma(\xi-\eta,\eta)|^2 d\eta$$

Alternatively, we could apply 1-dim Cauchy-Schwarz inequality twice w.r.t. $d\xi$ and $d\eta$ respectively, and then get the same upper bound as in (3.5) with B being replaced with $\sup_{\xi\in\mathbf{T}} \int_{\mathbf{T}} |\sigma(\xi,\eta)|^2 d\eta$. This will change the condition on $P-Q$ to the same one on P.

Combining (3.4) and (3.5), we see that $|E_\lambda| \leq \frac{B}{\lambda^2}$. Hence to prove (3.3), it suffices to obtain the estimate

$$B \lesssim_r \lambda^{2-r}. \tag{3.6}$$

To control B we make use of the dyadic decomposition of σ, aiming for some cancellations. For any $\xi \in \mathbf{T}$,

$$\begin{aligned}\int_{\mathbf{T}} |\sigma(\xi-\eta,\eta)|^2\, d\eta &= \int_{\mathbf{T}} \left| \sum_{j=0}^{\infty} \sigma_j(\xi-\eta,\eta) \right|^2 \, d\eta \\ &\leq \sum_{j_1,j_2=0}^{\infty} \frac{1}{2^{j_1}} \frac{1}{2^{j_2}} \sum_{m_1,m_2\in\mathbb{Z}} \left| \rho\left(\frac{m_1}{2^{j_1}}\right) \rho\left(\frac{m_2}{2^{j_2}}\right) \right| \chi_{A^{P,Q}}(m_1,m_2)\end{aligned} \tag{3.7}$$

By condition $(\star)$, $\frac{|m_1|}{|m_2|} \leq D_1$ for all $(m_1, m_2) \in A^{P,Q}$. The support of ρ forces $|m_1| \simeq 2^{j_1}$ and $|m_2| \simeq 2^{j_2}$. These facts show that $|j_1 - j_2| \lesssim 1$. Also note that for each m_1, there are only bounded number of m_2's such that $(m_1, m_2) \in A^{P,Q}$. Thus (3.7) implies

$$B = \sup_{\xi \in \mathbf{T}} \int_{\mathbf{T}} |\sigma(\xi - \eta, \eta)|^2 \, d\eta \lesssim \sum_{j=0}^{\infty} \frac{1}{2^j}.$$

When $\lambda \geq 1$, as $r \in (1, 2]$, trivially $B \lesssim \lambda^{2-r}$ and we are done. Let $M = [(2-r)\log_2 \frac{1}{\lambda}] + 1$, where $[x]$ denotes the integer part of x. In the case $\lambda < 1$, since $\sum_{j=M+1}^{\infty} \frac{1}{2^j} \lesssim_r \lambda^{2-r}$, the above method still gives the desired estimate for $\sum_{j=M+1}^{\infty} T_j$, the operator associated with the symbol $\sum_{j=M+1}^{\infty} \sigma_j$. It remains to control the level set of the operator $\sum_{j=0}^{M} T_j$ for $\lambda < 1$. Applying Lemma 3.2, we have

$$\left| \left\{ n \in \mathbb{Z} : \left| \sum_{j=0}^{M} T_j(f, g)(n) \right| > \lambda \right\} \right| \lesssim \frac{M^2}{\lambda} \lesssim_r \frac{1}{\lambda^r},$$

where we used the facts $r > 1$ and $\lambda < 1$ in the last inequality. This finishes the proof of Theorem 3.1.

3.3 Extensions to some arithmetic functions

Recall what we proved in the previous section:

Theorem 3.3 *Given two functions P and Q that map $\mathbb{Z}$ into $\mathbb{Z}$, assume that P or Q satisfies Condition $(\star)$. Then for any $\epsilon \in (0, 1]$, there is a constant C_ϵ depending only on ϵ, D_1 and D_2 such that the operator $B^{dis}_{P,Q}$ defined by*

$$B^{dis}_{P,Q}(f, g)(n) := \sum_{m \in \mathbb{Z} \setminus \{0\}} f(n - P(m)) g(n - Q(m)) \frac{1}{m}, \; f, g \in D(Z). \tag{3.8}$$

satisfies

$$\|B^{dis}_{P,Q}(f, g)\|_{l^{1+\epsilon,\infty}} \leq C_\epsilon \|f\|_{l^2} \|g\|_{l^2}, \; \textit{for any } f, g \in l^2. \tag{3.9}$$

Monotonic functions and non-constant polynomials satisfy Condition $(\star)$. However, Condition $(\star)$ requires that the function can attain each value for only bounded number of times, which excludes numerous arithmetic functions. For example, Ford [34] proves that for any $k \geq 2$ there exists n_k such that the Euler's totient function $\phi(n)$ equals n_k for at least k times. Thus ϕ does not satisfy Condition $(\star)$. By the fact that gaps between primes can be arbitrarily large, the prime counting function $\pi(n)$ does not satisfy Condition $(\star)$ either. Due to the discrete nature of the operator $B^{\text{dis}}_{P,Q}$, it is interesting to seek for a weaker condition for P and Q that includes some important arithmetic functions having many common values. The definition and the main theorem of our paper below serve as this purpose.

Definition 3.4 *For any function R that maps $\mathbb{Z}$ into $\mathbb{Z}$, let*

$$S_{M,N}^{R} := \left\{ (m,n) : R(m) = R(n), \frac{1}{2}N \le |n| \le 2N, \frac{1}{2}M \le |m| \le 2M \right\}.$$

*We say that R satisfies **Condition** $(\star\star)$ if there exist constants $\delta > 0$ and $\delta' > 0$ such that*

$$|S_{M,N}^{R}| \le \frac{\delta' MN}{(\log M \log N)^{1+\delta}}. \tag{3.10}$$

Roughly speaking, Condition $(\star\star)$ says that the solutions of $R(m) = R(n)$ in each dyadic strip have density slightly less than that of prime pairs. It is easy to see that Condition $(\star)$ implies Condition $(\star\star)$ for any $\delta > 0$ and thus the following main theorem of this paper extends Theorem 3.3.

Theorem 3.5 *Given two functions P and Q that map $\mathbb{Z}$ into $\mathbb{Z}$, assume that P or Q satisfies Condition $(\star\star)$. Then for any $\epsilon \in (\frac{1}{2\delta+1}, 1)$, there exists a constant $C_{\delta,\delta'}$ depending only on δ and δ' appeared in Condition $(\star\star)$ such that the operator $B_{P,Q}^{dis}$ defined by*

$$B_{P,Q}^{dis}(f,g)(n) := \sum_{m \in \mathbb{Z}\setminus\{0\}} f(n - P(m))g(n - Q(m))\frac{1}{m}, \ f, g \in D(Z). \tag{3.11}$$

satisfies

$$\|B_{P,Q}^{dis}(f,g)\|_{l^{1+\epsilon}} \le C_{\delta,\delta'}\|f\|_{l^2}\|g\|_{l^2}, \ \textit{for any } f, g \in l^2. \tag{3.12}$$

Remarks. (1). When $\epsilon = \frac{1}{2\delta+1}$ or $\epsilon = 1$, we have weak-$l^{1+\epsilon}$ estimate. See the proof of Theorem 3.5 in Section 2.1.2.

(2) When $P(m) = m$ and Q is a polynomial, the operator norm of $B_{P,Q}^{\text{dis}}$ we obtained is independent of Q. Such uniform estimates also appear in the continuous setting ([42, 62, 65, 104]).

(3) Note that the lower bound for ϵ goes to 0 as δ tends to ∞. This gives an evidence that $l^2 \times l^2 \to l^1$-boundedness of $B_{P,Q}^{\text{dis}}$ may be true for at least some special P and Q.

Very interestingly, Condition $(\star\star)$ covers some important arithmetic functions from number theory.

Corollary 3.6 *If P or Q equals the Euler's totient function $\phi(|m|)$ or the prime counting function $\pi(|m|)$, then for any $\epsilon \in (0, 1)$, we have*

$$\|B_{P,Q}^{dis}(f,g)\|_{l^{1+\epsilon}} \le C_{\epsilon}\|f\|_{l^2}\|g\|_{l^2}, \ \textit{for any } f, g \in l^2. \tag{3.13}$$

To better demonstrate the behaviors of $B^{\rm dis}_{P,Q}(f,g)(n)$ when $P(m)=\phi(|m|)$ or $P(m)=\pi(|m|)$, we exhibit the graphs of the operator for $f(x)=g(x)=\frac{1}{x^2+1}$. Fix $Q(m)=\operatorname{sgn}(m)d(|m|)$, where $d(m):=\sum_{a|m}1$ $(m>0)$ is the divisor function. We have

$$|S^d_{M,N}|\gtrsim\frac{MN}{\log M\log N},$$

as $d(p)=2$ for any prime p. Therefore Q does not satisfies Condition $(\star\star)$. We truncate the sum $B^{\rm dis}_{P,Q}(f,g)(x)=\sum_{m\in\mathbb{Z}\setminus\{0\}}f(x-P(m))g(x-Q(m))\frac{1}{m}$ up to $|m|\le T_0=1000$. For any $|x|\le 15$, the error will be bounded by $\frac{1}{(T_0-15)^2}$ which is good enough for us to plot Figure 3.1 and 3.2.

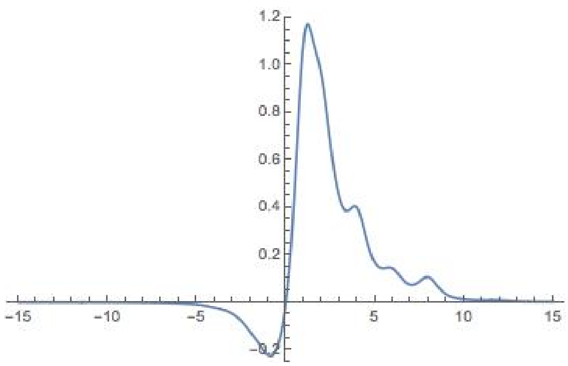

Figure 3.1: $B^{\rm dis}_{P,Q}(f,g)(x)$: $P(m)=\phi(|m|)$, $Q(m)=\operatorname{sgn}(m)d(|m|)$

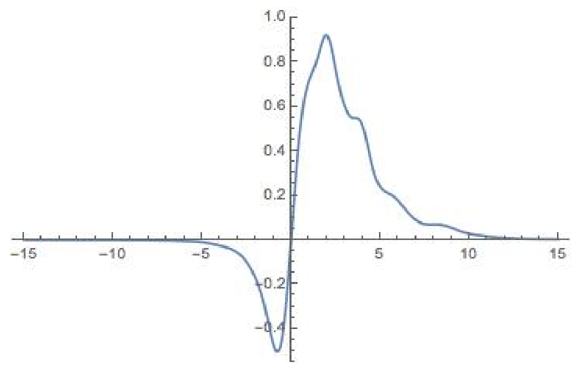

Figure 3.2: $B^{\rm dis}_{P,Q}(f,g)(x)$: $P(m)=\pi(|m|)$, $Q(m)=\operatorname{sgn}(m)d(|m|)$

Open Problems

There are a few related open problems to consider.

(1). When P and Q are polynomials, we believe that the operator norm of $B_{P,Q}$ may be chosen to be independent of the coefficients of both P and Q. We do not know how to achieve this.

(2). A useful operator related with $B^{\rm dis}_{P,Q}$ is the corresponding maximal operator

$$B^*_{P,Q}(f,g)(n)=\sup_{M\in[1,\infty)}\left|\frac{1}{M}\sum_{m=1}^{M}f(n-P(m))g(n-Q(m))\right|.$$

It is conjectured that this operator is bounded from $l^2\times l^2$ into $l^{1,\infty}$. See [22, 52] for some positive results about this operator.

(3). In our proof of Corollary 3.6, we use results about the Carmichael conjecture and gaps between primes. In converse, we wonder if the boundedness of this kind of operators could imply some information of the Carmichael conjecture and prime gaps.

(4). Note that if P is a constant function, then $B^{\mathrm{dis}}_{P,Q}$ is bounded using the theory of discrete linear Radon transform [53, 72]. In this extreme case, $|S^P_{M,N}| \simeq MN$. It remains to understand what happens if $|S^P_{M,N}|$ lies in between:

$$\frac{MN}{\log M \log N} \lesssim |S^P_{M,N}| \lesssim MN. \tag{3.14}$$

For example, besides the divisor function d, Möbius function μ and Ω function (the number of prime divisors) also satisfy (3.14). Therefore, Theorem 3.5 does not cover the cases when both P and Q are among these functions. Nevertheless, let us examine the graphs of the operator as before:

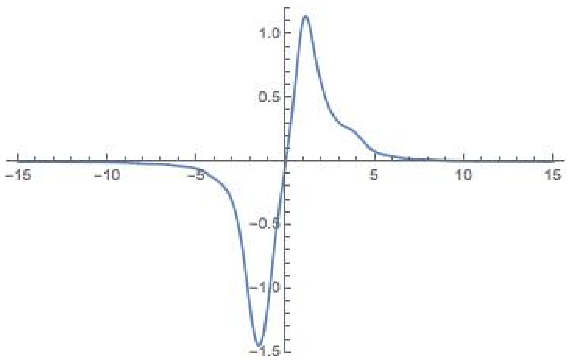

Figure 3.3: $B^{\mathrm{dis}}_{P,Q}(f,g)(x)$: $P(m) = \mu(|m|)$, $Q(m) = \mathrm{sgn}(m)d(|m|)$

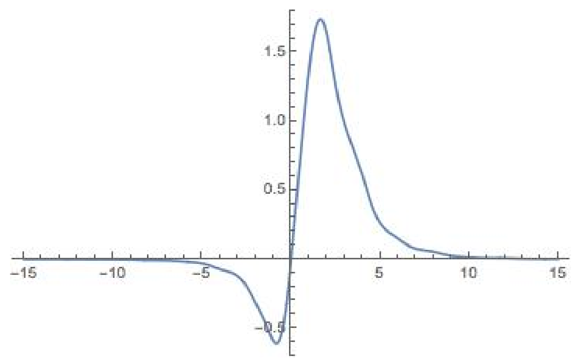

Figure 3.4: $B^{\mathrm{dis}}_{P,Q}(f,g)(x)$: $P(m) = \Omega(|m|)$, $Q(m) = \mathrm{sgn}(m)d(|m|)$

These pictures have similar shapes as those in Figures 3.1-3.2. It is reasonable to conjecture that our main theorem still holds in these cases.

Throughout this paper, we use $A \lesssim B$ to denote the statement that $A \le CB$ for some positive constant C. When the implied constant C depends on some parameter, say δ, we may write $A \lesssim_\delta B$. $A \simeq B$ is short for $A \lesssim B$ and $B \lesssim A$. For any set of integers E, $|E|$ and χ_E will be used to denote the counting measure and the indicator function of E, receptively.

3.3.1 Arithmetic functions with many common values

In this section, we show that Euler's totient function $\phi(n)$ and the prime counting function $\pi(n)$ satisfy Condition ($\star\star$), and thus prove Corollary 3.6 assuming Theorem 3.5. We will introduce some backgrounds in a friendly way, as to make our paper more readable to both number theorists and analysts.

Euler's totient function

Euler introduced the function $\phi(n)$ which counts the number of positive integers $\le n$ that are coprime to n. Euler's totient function not only has deep connections with prime numbers, but also appears in many classical theorems in number theory. We know that Euler's totient function $\phi(n)$ is multiplicative, i.e. if $\gcd(m,n)=1$, then $\phi(mn)=\phi(m)\phi(n)$. We have Euler's product formula

$$\phi(n) = n\prod_{p|n}\left(1-\frac{1}{p}\right),$$

and the following discrete Fourier representation [91],

$$\phi(n) = \sum_{k=1}^{n} \gcd(k,n) e^{-\frac{2\pi i k}{n}}.$$

In order to estimate the size of the set $S^{\phi}_{M,N}$, we need to consider the number of solutions of the equation $\phi(n)=\phi(m)$. Given m, Carmichael ([13], [14]) conjectured that there is at least one other integer $n\neq m$ such that $\phi(n)=\phi(m)$, which is the so called Carmichael's totient function conjecture. For each natural number m, let $A(m)$ be the number of n such that $\phi(n)=m$. An alternative way of stating Carmichael's conjecture is that $A(m)$ can never be 1.

We will use the bounds of $A(m)$ to verify Condition ($\star\star$) for ϕ. Ford [34] showed that, for any $k\geq 2$, there exist infinitely many m such that $A(m)=k$. For the upper bound, Pomerance [87] showed that

$$A(m) \leq m\exp\left(-(1+o(1))\log m \log\log\log m/\log\log m\right) =: U(m).$$

Therefore,

$$|S^{\phi}_{M,N}| \lesssim MU(\phi(2M)) \lesssim MU(M) \lesssim M\cdot\frac{M}{(\log M)^C} \text{ for any } C>0, \tag{3.1}$$

where the implied constant is absolute. Note that if $\phi(n)=m$, then $n\lesssim m\log\log m$ ([46], Thereom 328). Hence $S^{\phi}_{M,N}$ is not empty only when $M\lesssim N\log\log N$ and $N\lesssim M\log\log M$. Combining this fact with (3.1), we get

$$|S^{\phi}_{M,N}| \lesssim \frac{MN}{(\log M\log N)^{1+\delta}} \text{ for any } \delta>0,$$

as desired.

Gaps between primes

Let $\pi(x)$ be the number of primes no more than x. The famous Prime Number Theorem (PNT) states that, as $x\to\infty$,

$$\pi(x) \sim \frac{x}{\log x}. \tag{3.2}$$

Here $f(x) \sim g(x)$ means that $f(x)/g(x)$ goes to 1 as x goes to ∞. PNT was first proved independently in 1896 by Hadamard and de la Vallée Poussin. They both used the properties of the Riemann zeta-function $\zeta(s) = \sum_{n\geq 1} n^{-s}$ introduced by Riemann in his celebrated memoir.

Since $\pi(n) = \pi(m)$ can only occur when n and m are between two consecutive prime numbers, we need information about the gaps between primes, which have been extensively studied and many conjectures still remain open. PNT implies that the average gap between a prime p and the next prime is about $\log p$. By (3.2), one can derive that, for any $\epsilon > 0$, there exists a prime in the interval $(p, p + \epsilon p]$ for sufficiently large prime p. However, this is not enough for our application, as we need results for primes in shorter intervals.

Let $I(\theta, x)$ be the interval $[x, x + x^\theta]$. Hoheisel [49] showed that $I(\theta, x)$ contains primes for any $\theta > \frac{32999}{33000}$ as $x \to \infty$. Later, several authors made contributions to get smaller values of θ for which $I(\theta, x)$ contains primes for sufficiently large x. Iwaniec and Jutila ([55], $\theta > \frac{5}{9}$) introduced to this problem a sieve method, which was later refined by Heath-Brown and Iwaniec ([48], $\theta > \frac{11}{20}$). The best result to date is due to Baker, Harman, and Pintz [1], who showed that we can take $\theta \geq 0.525$. Riemann Hypothesis implies that $I(\theta, x)$ contains primes for any $\theta > \frac{1}{2}$ as $x \to \infty$ [15].

Using the result of Baker-Harman-Pintz, the size of the set $S^\pi_{M,N}$

$$|S^\pi_{M,N}| \lesssim M^{\theta_0} M,$$

where $\theta_0 = 0.525$. By the above results about gaps between primes, we have $M \simeq N$ if $S^\pi_{M,N} \neq \emptyset$. Since $\theta_0 < 1$, we deduce that

$$|S^\pi_{M,N}| \lesssim \frac{MN}{(\log M \log N)^{1+\delta}} \text{ for any } \delta > 0,$$

and hence we verify Condition $(\star\star)$ for π.

It is worth mentioning here some recent breakthroughs concerning gaps between primes. Let p_n be the n-th prime. Zhang [108] and Maynard [71] showed that there exists some absolute constant C such that $p_{n+1} - p_n < C$ happens infinitely often. For large gaps, Ford, Green, Konyagin, Maynard, and Tao [35] proved that there are infinitely many n's such that

$$p_{n+1} - p_n \gtrsim \frac{\log n \log\log n \log\log\log\log n}{\log\log\log n}.$$

It is possible that the boundedness of $B^{\text{dis}}_{P,Q}$ could provide a new approach to study gaps between primes. We shall not pursue this interesting idea here.

3.3.2 Proof of Theorem 3.5

By symmetry, we only consider the case that P satisfies Condition ($\star\star$) with parameter δ and δ' (and Q is arbitrary). For notational convenience, we will simply write T for $B^{\mathrm{dis}}_{P,Q}$. For any $\lambda > 0$ and $f, g \in D(\mathbb{Z})$, define the level set

$$E_\lambda := \{n \in \mathbb{Z} : |T(f,g)(n)| > \lambda\}.$$

Fix $\epsilon \in [\frac{1}{2\delta+1}, 1]$. Our goal is thus to prove the following the level set estimate

$$|E_\lambda| \lesssim_{\delta,\delta'} \frac{1}{\lambda^{1+\epsilon}}, \text{ whenever } \|f\|_{l^2} = \|g\|_{l^2} = 1. \tag{3.1}$$

The (strong) $l^{1+\epsilon}$-bound of $T(f,g)$ will follow immediately from interpolation.

We will only consider the case $\lambda < 1$. The other case can be proved similarly (In fact, the case $\lambda \geq 1$ is simpler: just set $M = 0$ in the proof below).

Define the Fourier transform for any $f \in D(\mathbb{Z})$ by

$$\hat{f}(\xi) := \sum_{m \in \mathbb{Z}} f(m) e^{-2\pi i \xi m}.$$

Then our operator can be rewritten as

$$\begin{aligned} T(f,g)(n) &= \sum_{m \in \mathbb{Z}\setminus\{0\}} f(n - P(m)) g(n - Q(m)) \frac{1}{m} \\ &= \int_{\mathbf{T}^2} \hat{f}(\xi)\hat{g}(\eta) e^{2\pi i(\xi+\eta)n} \sigma(\xi,\eta)\, d\xi d\eta, \end{aligned}$$

where $\mathbf{T}$ is the unit circle and σ is the periodic bilinear multiplier given by

$$\sigma(\xi,\eta) = \sum_{m \in \mathbb{Z}\setminus\{0\}} \frac{1}{m} e^{-2\pi i(P(m)\xi + Q(m)\eta)}.$$

Using a standard technique, we proceed to decompose the multiplier σ dyadically. Choose an odd function $\rho \in \mathcal{S}(\mathbb{R})$ supported in the set $\{x : |x| \in (\frac{1}{2}, 2)\}$ with the property that

$$\frac{1}{x} = \sum_{j=0}^{\infty} \frac{1}{2^j} \rho\left(\frac{x}{2^j}\right) \text{ for any } x \in \mathbb{R} \text{ with } |x| \geq 1.$$

Let

$$\sigma_j(\xi,\eta) := \frac{1}{2^j} \sum_{m \in \mathbb{Z}} \rho\left(\frac{m}{2^j}\right) e^{-2\pi i(P(m)\xi + Q(m)\eta)},$$

and consequently

$$\sigma(\xi,\eta) = \sum_{j=0}^{\infty} \sigma_j(\xi,\eta).$$

Correspondingly T can be written as the sum $\sum_{j=0}^{\infty} T_j$, where

$$\begin{aligned} T_j(f,g)(n) &= \int_{\mathbf{T}}\int_{\mathbf{T}} \hat{f}(\xi)\hat{g}(\eta)e^{2\pi i(\xi+\eta)n}\sigma_j(\xi,\eta)\,d\xi d\eta \\ &= \frac{1}{2^j}\sum_{m\in\mathbb{Z}} \rho\left(\frac{m}{2^j}\right) f(n-P(m))g(n-Q(m)). \end{aligned}$$

Let M be a non-negative integer to be determined later. Decompose T into two parts: $\sum_{j=0}^{M-1} T_j$ and $\sum_{j=M}^{\infty} T_j$. It remains to control the level sets

$$E_\lambda^{(1)} := \left\{ n\in\mathbb{Z} : \left| \sum_{j=0}^{M-1} T_j(f,g)(n) \right| > \lambda \right\} \tag{3.2}$$

and

$$E_\lambda^{(2)} := \left\{ n\in\mathbb{Z} : \left| \sum_{j=M}^{\infty} T_j(f,g)(n) \right| > \lambda \right\} \tag{3.3}$$

$E_\lambda^{(1)}$ can be estimated by the following simple lemma, whose proof is based on Hölder inequality and is omitted.

Lemma 3.7 *There is an absolute positive constant C such that for any $j \geq 0$,*

$$\|T_j(f,g)\|_{l^1} \leq C\|f\|_{l^2}\|g\|_{l^2}.$$

By Lemma 3.7 and triangle inequality, we see that

$$|E_\lambda^{(1)}| \leq \frac{M}{\lambda}. \tag{3.4}$$

Note that (3.4) is useful, i.e. better than the upper bound $\frac{1}{\lambda^{1+\epsilon}}$, only when $\lambda < 1$. We do not need this estimate in the case $\lambda \geq 1$.

To control $|E_\lambda^{(2)}|$, we employ a TT^* method. Define an auxiliary function

$$h(n) = \frac{\overline{II(f,g)(n)}}{|II(f,g)(n)|}\chi_{E_\lambda^{(2)}}(n),$$

where

$$II(f,g)(n) := \sum_{j=M}^{\infty} T_j(f,g)(n).$$

It is easy to verify that

$$\lambda^2 |E_\lambda^{(2)}|^2 \leq \left(\sum_{n\in\mathbb{Z}} II(f,g)(n)h(n) \right)^2. \tag{3.5}$$

By Fubini's theorem and the definition of Fourier transform we obtain

$$\begin{aligned}
\sum_{n\in\mathbb{Z}} II(f,g)(n)h(n) &= \int_{\mathbf{T}^2} \hat{f}(\xi)\hat{g}(\eta) \sum_{j=M}^{\infty} \sigma_j(\xi,\eta)\hat{h}(-(\xi+\eta))\, d\xi d\eta \\
&= \int_{\mathbf{T}^2} \hat{f}(\xi)\hat{g}(\eta-\xi) \sum_{j=M}^{\infty} \sigma_j(\xi,\eta-\xi)\hat{h}(-\eta)\, d\eta d\xi
\end{aligned}$$

Invoking Cauchy-Schwarz inequality and Plancherel's Theorem we get

$$\begin{aligned}
&\left(\sum_{n\in\mathbb{Z}} II(f,g)(n)h(n) \right)^2 \\
&\leq \left(\int_{\mathbf{T}} |\hat{f}(\xi)||E_\lambda^{(2)}|^{\frac{1}{2}} \left(\int_{\mathbf{T}} |\hat{g}(\eta-\xi)|^2 \left| \sum_{j=M}^{\infty} \sigma_j(\xi,\eta-\xi) \right|^2 d\eta \right)^{\frac{1}{2}} d\xi \right)^2 \\
&\leq |E_\lambda^{(2)}| \int_{\mathbf{T}} \int_{\mathbf{T}} |\hat{g}(\eta)|^2 \left| \sum_{j=M}^{\infty} \sigma_j(\xi,\eta) \right|^2 d\eta d\xi \\
&\leq V|E_\lambda^{(2)}|
\end{aligned} \tag{3.6}$$

where

$$V := \sup_{\eta\in\mathbf{T}} \int_{\mathbf{T}} \left| \sum_{j=M}^{\infty} \sigma_j(\xi,\eta) \right|^2 d\xi.$$

Using (3.5) and (3.6), we see that

$$|E_\lambda^{(2)}| \leq \frac{V}{\lambda^2}. \tag{3.7}$$

To control V, we recall

$$S_{M,N}^P = \left\{ (m,n) : P(m) = P(n), \frac{1}{2}N \leq |n| \leq 2N, \frac{1}{2}M \leq |m| \leq 2M \right\},$$

and note that for any $\eta \in \mathbf{T}$,

$$\int_{\mathbf{T}} \left| \sum_{j=M}^{\infty} \sigma_j(\xi, \eta) \right|^2 d\xi$$
$$\leq \sum_{j_1, j_2 = M}^{\infty} \frac{1}{2^{j_1}} \frac{1}{2^{j_2}} \sum_{m_1, m_2 \in \mathbb{Z}} \left| \rho\left(\frac{m_1}{2^{j_1}}\right) \rho\left(\frac{m_2}{2^{j_2}}\right) \right| \chi_{S^P_{2^{j_1}, 2^{j_2}}}(m_1, m_2). \tag{3.8}$$

By the support of ρ and Condition $(\star\star)$, (3.8) implies that

$$V \lesssim \sum_{j_1, j_2 = M}^{\infty} \frac{1}{2^{j_1}} \frac{1}{2^{j_2}} |S^P_{2^{j_1}, 2^{j_2}}| \lesssim_{\delta'} \sum_{j_1, j_2 = M}^{\infty} \frac{1}{(j_1 j_2)^{1+\delta}} \lesssim_{\delta, \delta'} \frac{1}{M^{2\delta}}. \tag{3.9}$$

Combing (3.7) and (3.9), we get the estimate for $|E^{(2)}_\lambda|$

$$|E^{(2)}_\lambda| \lesssim_{\delta, \delta'} \frac{1}{\lambda^2 M^{2\delta}}. \tag{3.10}$$

Apply (3.4) and (3.10), and one gets

$$|E_\lambda| \leq |E^{(1)}_{\lambda/2}| + |E^{(2)}_{\lambda/2}| \lesssim_{\delta, \delta'} \frac{M}{\lambda} + \frac{1}{\lambda^2 M^{2\delta}}. \tag{3.11}$$

Optimize the above upper bound by choosing M to be an integer near $(\frac{1}{\lambda})^{\frac{1}{1+2\delta}}$, and we obtain

$$|E_\lambda| \lesssim_{\delta, \delta'} \frac{M}{\lambda} \lesssim \left(\frac{1}{\lambda}\right)^{\frac{1}{1+2\delta}} \frac{1}{\lambda} \lesssim \frac{1}{\lambda^{1+\epsilon}},$$

as $\epsilon \geq \frac{1}{1+2\delta}$ and $\lambda < 1$. This is our desired estimate (3.1), and the proof of Theorem 3.5 is thus complete.

Chapter 4

Bilinear Hilbert transform along two polynomials

This chapter include my research on bilinear Hilbert transform along two polynomials [23]. We prove that the bilinear Hilbert transform along two polynomials $B_{P,Q}(f,g)(x) = \int_{\mathbb{R}} f(x-P(t))g(x-Q(t))\frac{dt}{t}$ is bounded from $L^p \times L^q$ to L^r for a large range of (p,q,r), as long as the polynomials P and Q have distinct leading and trailing degrees. The same boundedness property holds for the corresponding bilinear maximal function $\mathcal{M}_{P,Q}(f,g)(x) = \sup_{\epsilon>0} \frac{1}{2\epsilon}\int_{-\epsilon}^{\epsilon} |f(x-P(t))g(x-Q(t))|dt$.

4.1 Introduction

The Hilbert transform along a curve $\gamma : \mathbb{R} \to \mathbb{R}^n$ is defined by

$$H_\gamma(f)(x) := \int_{\mathbb{R}} f(x-\gamma(t))\frac{dt}{t}, \ f \in \mathcal{S}(\mathbb{R}^n). \tag{4.1}$$

Here $\mathcal{S}(\mathbb{R}^n)$, $n \in \mathbb{N}$, denotes the space of Schwartz functions on $\mathbb{R}^n$. Stein ([93]) raised the question that under what condition on γ is H_γ bounded from $L^p(\mathbb{R}^n)$ to itself for some p. Among many curves, a simple but important two dimensional example is the curve $\gamma_{a,b}(t) = (t^a, t^b)$, where a, b are distinct natural numbers. For this particular type of curve, (4.1) becomes

$$H_{\gamma_{a,b}}(f)(x_1, x_2) = \int_{\mathbb{R}} f(x_1 - t^a, x_2 - t^b)\frac{dt}{t}, \ f \in \mathcal{S}(\mathbb{R}^2). \tag{4.2}$$

The L^2-boundedness of $H_{\gamma_{a,b}}$ was first proved by Fabes [32] and Stein and Wainger [95], using different methods. Nagel et.al. [80, 81] obtained the L^p-boundedness for $p \in (1,\infty)$. It turns out $\gamma_{a,b}$ is the model curve for the very general "well-curved" curves ([97]).

The purpose of this article is to investigate a bilinear analogue of $H_{\gamma_{a,b}}$. Given two polynomials P and Q on $\mathbb{R}$, define the bilinear Hilbert transform along P, Q by

$$B_{P,Q}(f,g)(x) := \int_{\mathbb{R}} f(x-P(t))g(x-Q(t))\frac{dt}{t}, \ f, g \in \mathcal{S}(\mathbb{R}). \tag{4.3}$$

In the above definition, instead of just t^a and t^b, two arbitrary polynomials are involved, which provides a more general framework. A natural question is that under what condition on P and Q does $B_{P,Q}$ satisfy any L^p estimates. For this problem, we can assume without loss of generality that both P and Q contain no constant term. There are already some positive results in the literature. For example, when P and Q are distinct linear polynomials, $B_{P,Q}$ is in fact the famous bilinear Hilbert transform, whose boundedness was proved by Lacey and Thiele in a pair of breakthrough papers ([60, 61]). Xiaochun Li first studied the case $P(t) = t$, $Q(t) = t^d$, $d \in \mathbb{N}$, and showed that $B_{P,Q}$ is bounded from $L^2 \times L^2$ to L^1 ([64]). Together with Lechao Xiao, Li later ([65]) obtained the L^p estimates in full range when $P(t) = t$ and Q is any polynomial without linear term. Following the approach in [64, 65], we obtain the theorems below which can be viewed as an extension of Li-Xiao's result to a larger range of pairs of polynomials.

Definition 4.1 *The* ***correlation degree*** *of any two polynomials P and Q is defined as the smallest natural number d such that any non-zero real root of $P'(x) - Q'(x)$ has multiplicity at most d.*

Theorem 4.2 *Given two polynomials P and Q without constant terms, we can always write them as*

$$P(t) = a_{d_1} t^{d_1} + a_{d_1-1} t^{d_1-1} + \cdots + a_{e_1} t^{e_1}, 1 \le e_1 \le d_1, a_{d_1} a_{e_1} \ne 0 \tag{4.4}$$

$$Q(t) = b_{d_2} t^{d_2} + b_{d_2-1} t^{d_2-1} + \cdots + b_{e_2} t^{e_2}, 1 \le e_2 \le d_2, b_{d_2} b_{e_2} \ne 0. \tag{4.5}$$

Assume $d_1 \ne d_2$ and $e_1 \ne e_2$. Then there is a constant $C_{P,Q}$ depending on P and Q (and of course p, q, r) such that $B_{P,Q}$ defined in (4.3) *satisfies $\|B_{P,Q}(f, g)\|_r \le C_{P,Q} \|f\|_p \|g\|_q$ for any $f, g \in \mathcal{S}(\mathbb{R})$, whenever $p, q \in (1, \infty)$, $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$, $r > \frac{d}{d+1}$. Here d is the correlation degree of P and Q.*

Remarks. 1. In the expressions (4.4) and (4.5), we can call d_1 and d_2 the **leading degrees**, as they are the degrees of the leading terms. Similarly, e_1 and e_2 may be called **trailing degrees** if we name $a_{e_1} t^{e_1}$ and $b_{e_2} t^{e_2}$ as **trailing terms**. So the condition imposed on P and Q in the theorem can be phrased in words as "P and Q have distinct leading and trailing degrees".

2. We conjecture that the constant $C_{P,Q}$ in the theorem may be chosen to be independent of the coefficients of the polynomials. This seems to be a hard and technical problem, whose solution may involve the ideas in the proof of uniform estimate for the bilinear Hilbert transform ([42, 62, 104]).

3. For any fixed natural number d, there exist polynomials P and Q with correlation degree d such that $B_{P,Q}$ is unbounded whenever $r < \frac{d}{d+1}$ (see Section 3.2 in [65] for an example). In this sense the lower bound for r given in Theorem 4.2 is sharp up to the endpoint. However, if we fix the polynomials P and Q, the lower bound of r in Theorem 4.2 may not be the best. For instance, let $P(t) = t^6$ and $Q(t) = 3t^4 - 3t^2$.

Then $B_{P,Q}$ is the zero operator, which is trivially bounded for $r > \frac{1}{2}$. But the correlation degree of P and Q is 2. It is interesting to find a way to determine the lowest r for any given P and Q. This task requires improvement on Lemma 4.4 (see Section 4.2).

As a byproduct of the proof of Theorem 4.2, we obtain the same estimate for the bilinear maximal function $\mathcal{M}_{P,Q}$ defined by

$$\mathcal{M}_{P,Q}(f,g)(x) := \sup_{\epsilon>0} \frac{1}{2\epsilon} \int_{-\epsilon}^{\epsilon} |f(x-P(t))g(x-Q(t))|\, dt. \tag{4.6}$$

Theorem 4.3 *Let P, Q and p, q, r satisfy the conditions stated in Theorem 4.2. Then $\mathcal{M}_{P,Q}$ is bounded from $L^p \times L^q$ to L^r.*

Just like the relationship between $B_{P,Q}$ and $H_{\gamma_{a,b}}$, $\mathcal{M}_{P,Q}$ can be viewed as a bilinear analogue of the the maximal function associated with $H_{\gamma_{a,b}}$,

$$M_{\gamma_{a,b}}(f)(x_1,x_2) := \sup_{h>0} \frac{1}{2h} \int_{-h}^{h} |f(x_1 - t^a, x_2 - t^b)|\, dt,\ f \in \mathcal{S}(\mathbb{R}^2).$$

The L^p-boundedness of $M_{\gamma_{a,b}}$ was proved in [82] (see [94, 96, 97] for further developments on more general curves), and Theorem 4.3 is the parallel result in the bilinear setting.

The rest of the paper is organized as follows. In section 4.2, we make careful decompositions on our operator, and after throwing away the paraproduct part, reduce Theorem 4.2 to two estimates (Proposition 4.7 and Proposition 4.8): a scale-type decay estimate when $p = q = 2$, and a moderate blow-up estimate for general p and q. The decay estimate will be proved in section 4.3 and 4.4, using TT* method and σ-uniformity method. In the last section, we show how to obtain the moderate blow-up estimate by adapting methods from [65], and prove Theorem 4.3.

Throughout the paper we use C to denote a positive constant (which may depend on P and Q) whose value is allowed to change from line to line. $A \lesssim B$ means $A \le CB$. $A \simeq B$ is short for $A \lesssim B$ and $B \lesssim A$. We use $A \sim B$ to denote the statement that B is the leading term (principal contribution) of A after using Taylor expansion or stationary phase method. χ_E will be used to denote the indicator function of a set E.

4.2 Decomposition and reduction

Pick an odd function $\rho \in \mathcal{S}(\mathbb{R})$ supported in the set $\{x : |x| \in (\frac{1}{2}, 2)\}$ with the property that $t^{-1} = \sum_{j\in\mathbb{Z}} 2^j \rho(2^j t)$ for any $t \neq 0$. Then we can write $B_{P,Q}(f,g)(x) = \sum_{j\in\mathbb{Z}} T_j(f,g)(x)$, where

$$\begin{aligned} T_j(f,g)(x) &:= \int f(x-P(t))g(x-Q(t))2^j\rho(2^j t)\, dt \\ &= \iint \hat{f}(\xi)\hat{g}(\eta)e^{2\pi i(\xi+\eta)x} m_j(\xi,\eta)\, d\xi d\eta, \end{aligned} \tag{4.1}$$

and

$$m_j(\xi,\eta) := \int 2^j \rho(2^j t) e^{-2\pi i(\xi P(t)+\eta Q(t))}\, dt. \tag{4.2}$$

We first prove that each T_j is bounded.

Lemma 4.4 *Let P and Q be two arbitrary polynomials. Then each T_j is bounded from $L^p \times L^q$ to L^r, whenever $p,q \in (1,\infty)$, $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$, $r > \frac{d}{d+1}$, where d is the correlation degree of P and Q.*

Proof: We only consider the operator T_0, as the other cases are similar. The idea of the proof is based on Lemma 9.1 in [64]. Note that when $r \geq 1$ the boundedness of T_0 follows from Minkowski inequality. So we assume now $r < 1$. Since $|t| \simeq 1$, we can restrict x and the support of f and g to a bounded interval $I_{P,Q}$. When the Jacobian $Q'(t) - P'(t) \neq 0$ for all t in the support of ρ, T_0 is bounded from $L^1 \times L^1$ to L^1 by changing variables $u = x - P(t)$ and $v = x - Q(t)$. Thus T_0 is bounded from $L^1 \times L^1$ to $L^{\frac{1}{2}}$ by Cauchy-Schwarz inequality.

Now we focus on the case that there is a root of $Q'(t) - P'(t)$ lying in the support of ρ. Let t_0 be such a root and $I(t_0)$ be a small neighborhood of t_0. It suffices to prove that

$$\int_{I_{P,Q}} \left| \int_{I(t_0)} f(x-P(t))g(x-Q(t))\rho(t)\, dt \right|^r dx \lesssim \|f\|_p^r \|g\|_q^r, \tag{4.3}$$

for $p,q \in (1,\infty)$, $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$, $r > \frac{d}{d+1}$. Because of the restriction on $I(t_0)$, the function ρ in (4.3) can be dropped. Let ρ_0 be a bump function supported in $\{t : |t| \in (\frac{1}{2}, 2)\}$ and satisfies $\sum_{j\in\mathbb{Z}} \rho_0(2^j t) = 1$ for all $t \in \mathbb{R}$. Then (4.3) will be proved once we can show that there is some $\epsilon > 0$ such that

$$\int_{I_{P,Q}} \left| \int f(x-P(t))g(x-Q(t))\rho_0(2^j(t-t_0))\, dt \right|^r dx \lesssim 2^{-\epsilon j} \|f\|_p^r \|g\|_q^r \tag{4.4}$$

holds for all large positive j. Changing variable $t - t_0 \to t$ and translating f and g by $P(t_0)$ and $Q(t_0)$

respectively, (4.4) becomes

$$\int_{I_{P,Q}} \left| \int f(x - P_1(t))g(x - Q_1(t))\rho_0(2^j t)\, dt \right|^r dx \lesssim 2^{-\epsilon j} \|f\|_p^r \|g\|_q^r, \tag{4.5}$$

where $P_1(t) := P(t + t_0) - P(t_0)$ and $Q_1(t) := Q(t + t_0) - Q(t_0)$. By the support of ρ_0, $|t| \simeq 2^{-j}$. This implies that $P_1(t) \lesssim 2^{-j}$ and $Q_1(t) \lesssim 2^{-j}$ by mean value theorem. So we can for free restrict x to an interval of length $\simeq 2^{-j}$. Let I_N be such an interval and define

$$T_N(f, g)(x) = \chi_{I_N}(x) \int f(x - P_1(t))g(x - Q_1(t))\rho_0(2^j t)\, dt.$$

It remains to show

$$\|T_N(f, g)\|_r \lesssim 2^{-\epsilon j} \|f\|_p \|g\|_q. \tag{4.6}$$

By Fubini theorem, T_N is bounded with norm $\lesssim 2^{-j}$ when $r = 1$. Next we aim to get a slow increasing $L^1 \times L^1 \to L^{\frac{1}{2}}$ norm. By Cauchy-Schwarz inequality,

$$\int |T_N(f, g)(x)|^{\frac{1}{2}}\, dx \lesssim 2^{-j/2} \|T_N(f, g)\|_1^{\frac{1}{2}}. \tag{4.7}$$

$\|T_N(f, g)\|_1$ can be calculated by changing variables $u = x - P_1(t)$ and $v = x - Q_1(t)$. Using Taylor expansion and the fact that t_0 has multiplicity at most d, the Jacobian $Q_1'(t) - P_1'(t)$ is bounded below by 2^{-dj}. Therefore

$$\|T_N(f, g)\|_1 \lesssim 2^{dj} \|f\|_1 \|g\|_1. \tag{4.8}$$

Combining (4.7) and (4.8), we get

$$\|T_N(f, g)\|_{\frac{1}{2}} \lesssim 2^{(d-1)j} \|f\|_1 \|g\|_1. \tag{4.9}$$

Interpolating (4.9) with the L^1-norm, we obtain (4.6). ■

By lemma 4.4, to prove Theorem 4.2 it suffices to prove the following theorem.

Theorem 4.5 *Let P and Q be two polynomials with distinct leading and trailing degrees. Then there is a large N depending on P and Q such that $\sum_{|j|>N} T_j(f, g)(x)$ is bounded from $L^p \times L^q$ to L^r for all $p, q \in (1, \infty)$, $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$.*

From the definition (4.1), we see that $j > N$ corresponds to small $|t|$, in which case the trailing term dominates each polynomial; $j < -N$ corresponds to large $|t|$, in which case P and Q behave almost the same as their leading term. We will only deal with $\sum_{j>N} T_j(f,g)(x)$ since the other case is similar.

Let P, Q be polynomials written as (4.4) and (4.5). in When j is large (i.e. $|t|$ is close to 0), the trailing terms $a_{e_1}t^{e_1}$ and $b_{e_2}t^{e_2}$ dominate $P(t)$ and $Q(t)$, respectively. Since all the constants in our proof are allowed to depend on the coefficients of P and Q, we may assume without loss of generality that $a_{e_1} = b_{e_2} = 1$. For notation simplicity, from now on we denote $a := e_1$ and $b := e_2$. Recall that $e_1 \neq e_2$ and thus we may assume $a < b$. With these new notations, we can write $P(t) = t^a + P_\epsilon(t)$ and $Q(t) = t^b + Q_\epsilon(t)$, where $P_\epsilon(t)$ (resp. $Q_\epsilon(t)$) consists of terms whose degree is higher than a (resp. b). As $P_\epsilon(t)$ and $Q_\epsilon(t)$ are small when $j > N$ and can be viewed as error terms. We urge the reader to ignore them in the first reading of this paper.

The overall idea of the proof is to look at the size of the symbol $m_j(\xi,\eta)$ defined in (4.2), which can be estimated by stationary phrase method after proper cut-off and rescaling. By a change of variable,

$$m_j(\xi,\eta) = \int \rho(t) e^{-2\pi i\left(\frac{\xi}{2^{aj}}(t^a+\epsilon_P(t))+\frac{\eta}{2^{bj}}(t^b+\epsilon_Q(t))\right)}\, dt, \tag{4.10}$$

where

$$\epsilon_P(t) := 2^{aj} P_\epsilon(2^{-aj}t); \tag{4.11}$$

$$\epsilon_Q(t) := 2^{bj} Q_\epsilon(2^{-bj}t). \tag{4.12}$$

Clearly $|\epsilon_P(t)| \leq 2^{-N}|t^a|$ and $|\epsilon_Q(t)| \leq 2^{-N}|t^b|$ as $j > N$. The expression (4.10) suggests that we need to consider the sizes of $\frac{\xi}{2^{aj}}$ and $\frac{\eta}{2^{bj}}$. Therefore we choose $\Phi \in \mathcal{S}(\mathbb{R})$ such that $\hat{\Phi}$ is supported on $\{\xi : |\xi| \in (\frac{1}{2}, 2)\}$ and

$$\sum_{m\in\mathbb{Z}} \hat{\Phi}\left(\frac{\xi}{2^m}\right) = 1,\ \xi \neq 0.$$

Then decompose T_j as $T_j = \sum_{(m,n)\in\mathbb{Z}^2} T_{j,m,n}$ where

$$T_{j,m,n}(f,g)(x) := \iint \hat{f}(\xi)\hat{g}(\eta)e^{2\pi i(\xi+\eta)x} m_j(\xi,\eta)\hat{\Phi}\left(\frac{\xi}{2^{aj+m}}\right)\hat{\Phi}\left(\frac{\eta}{2^{bj+n}}\right) d\xi d\eta, \tag{4.13}$$

is the bilinear operator with symbol

$$\begin{aligned} M_{j,m,n}(\xi,\eta) :=& m_j(\xi,\eta)\hat{\Phi}\left(\frac{\xi}{2^{aj+m}}\right)\hat{\Phi}\left(\frac{\eta}{2^{bj+n}}\right) \\ =& \int \rho(t) e^{-2\pi i\left(\frac{\xi}{2^{aj}}(t^a+\epsilon_P(t))+\frac{\eta}{2^{bj}}(t^b+\epsilon_Q(t))\right)}\, dt \hat{\Phi}\left(\frac{\xi}{2^{aj+m}}\right)\hat{\Phi}\left(\frac{\eta}{2^{bj+n}}\right) \end{aligned} \tag{4.14}$$

In estimating its size, the symbol $M_{j,m,n}$ can be viewed roughly as

$$\int \rho(t) e^{-2\pi i (2^m t^a + 2^n t^b)}\, dt, \tag{4.15}$$

which decays rapidly if $|m-n|$ is large. In fact, $\sum_{j>N} \sum_{|m-n| \gtrsim 1} T_{j,m,n}(f,g)(x)$ can be reduced to the paraproduct studied in [63] (see Section 7.2 in [65] for details). To deal with the remaining $|m-n| \lesssim 1$ case, we can assume without loss of generality that $m = n$. For notation simplicity, denote $M_{j,m} := M_{j,m,m}$ and $T_{j,m} := T_{j,m,m}$. Using oddness of ρ and Taylor expansion, $\sum_{j>N} \sum_{m \leq 0} T_{j,m}$ can also be reduced to the paraproduct in [63]. Thus we will only focus on the most difficult case in proving Theorem 4.5: handing the operator $\sum_{j>N} \sum_{m>0} T_{j,m}$. Our goal is to prove

Theorem 4.6 *For all* $p, q \in (1, \infty)$, $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$,

$$\left\| \sum_{j>N} \sum_{m>0} T_{j,m}(f,g) \right\|_r \lesssim \|f\|_p \|g\|_q. \tag{4.16}$$

By interpolation, the above theorem follows from two propositions below.

Proposition 4.7

$$\left\| \sum_{j>N} T_{j,m}(f,g) \right\|_1 \lesssim 2^{-\epsilon m} \|f\|_2 \|g\|_2 \text{ for some } \epsilon > 0. \tag{4.17}$$

Proposition 4.8 *For* $p, q \in (1, \infty)$, $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$,

$$\left\| \sum_{j>N} T_{j,m}(f,g) \right\|_{r,\infty} \lesssim m \|f\|_p \|g\|_q, \tag{4.18}$$

4.3 TT^* method

We prove Proposition 4.7 in this section and the next.

Since we can for free insert cut-offs on $\hat{f}$ and $\hat{g}$ according to the support of $M_{j,m}$, in proving Proposition 4.7 we only need to consider the estimate for a single scale, i.e.

Proposition 4.9 $\|T_{j,m}(f,g)\|_1 \lesssim 2^{-\epsilon m} \|f\|_2 \|g\|_2$ *for any* $j > N$ *and* $m > 0$.

By rescaling, Proposition 4.9 is a consequence of

Proposition 4.10 *For any $j > N$ and $m > 0$, $\|B_{j,m}(f,g)\|_1 \lesssim 2^{-\epsilon m}\|f\|_2\|g\|_2$, where*

$$B_{j,m}(f,g)(x) := 2^{-\frac{(b-a)j}{2}} \int \rho(t) f * \Phi\left(\frac{x}{2^{(b-a)j}} - 2^m(t^a + \epsilon_P(t))\right) \\ g * \Phi(x - 2^m(t^b + \epsilon_Q(t)))\, dt \tag{4.1}$$

This proposition follows from the two estimates below.

Proposition 4.11 $\|B_{j,m}(f,g)\|_1 \lesssim 2^{\frac{(b-a)j-m}{8}}\|f\|_2\|g\|_2$ *for any $j > N$ and $m > 0$.*

Proposition 4.12 *There exists a positive δ such that $\|B_{j,m}(f,g)\|_1 \lesssim 2^{-\epsilon m}\|f\|_2\|g\|_2$ whenever $(b-a)j > (1-\delta)m$.*

Proposition 4.11 is efficient when m is large and Proposition 4.12 is useful for small m. The proofs for the above two propositions require different methods.

We prove Proposition 4.11 in this section, using a TT^* method. More precisely, we aim to obtain a $L^2 \times L^2 \to L^2$ bound with good decay. By making suitable partitions in time spaces, we see that x can be assumed to be supported in an interval of length $\simeq 2^{(b-a)j+m}$. This observation indicates that it suffices to prove

$$\|B_{j,m}(f,g)\|_2 \lesssim 2^{\frac{(b-a)j-m}{6}} 2^{-\frac{(b-a)j+m}{2}} \|f\|_2\|g\|_2. \tag{4.2}$$

Rewrite $B_{j,m}$ as

$$B_{j,m}(f,g)(x) = 2^{-\frac{(b-a)j}{2}} \iint \hat{f}(\xi)\hat{g}(\eta) e^{2\pi i\left(\frac{\xi}{2^{(b-a)j}}+\eta\right)x} I_{\rho,m}\hat{\Phi}(\xi)\hat{\Phi}(\eta)\, d\xi\, d\eta, \tag{4.3}$$

where

$$I_{\rho,m} := \int \rho(t) e^{-2\pi i 2^m(\xi(t^a+\epsilon_P(t))+\eta(t^b+\epsilon_Q(t)))}\, dt \tag{4.4}$$

Let $\varphi(t) := \xi(t^a + \epsilon_P(t)) + \eta(t^b + \epsilon_Q(t))$ and t_0 be a solution of $\varphi'(t) = 0$. Let $\phi(\xi,\eta) := \varphi(t_0)$. By stationary phase method,

$$I_{\rho,m}(\xi,\eta)\hat{\Phi}(\xi)\hat{\Phi}(\eta) \sim 2^{-\frac{m}{2}} e^{i2^m\phi(\xi,\eta)}. \tag{4.5}$$

Thus we can regard $B_{j,m}$ as

$$B_{j,m}(f,g)(x) \sim 2^{-\frac{(b-a)j}{2}} 2^{-\frac{m}{2}} \iint \hat{f}(\xi)\hat{g}(\eta) e^{2\pi i(\frac{\xi}{2^{(b-a)j}})x} e^{i2^m\phi(\xi,\eta)}\, d\xi d\eta. \tag{4.6}$$

Then

$$\begin{aligned}\|B_{j,m}\|_2^2 = &\int B_{j,m}(x)\overline{B_{j,m}(x)}\,dx\\
=&2^{-(b-a)j-m}\iiiint\limits_{\frac{\xi}{2^{(b-a)j}}+\eta=\frac{\xi_1}{2^{(b-a)j}}+\eta_1} \hat{f}(\xi)\hat{\Phi}(\xi)\overline{\hat{f}(\xi_1)\hat{\Phi}(\xi_1)}\\
&\hat{g}(\eta)\hat{\Phi}(\eta)\overline{\hat{g}(\eta_1)\hat{\Phi}(\eta_1)}e^{i2^m[\phi(\xi,\eta)-\phi(\xi_1,\eta_1)]}d\xi d\eta d\xi_1 d\eta_1\\
=&2^{-(b-a)j-m}\iiint F_\tau(\xi)G_\tau(\eta)e^{i2^m Q_\tau(\xi,\eta)}\,d\xi d\eta d\tau,\end{aligned}$$

where

$$\begin{aligned}F_\tau(\xi) &:= \hat{f}(\xi)\hat{\Phi}(\xi)\overline{\hat{f}(\xi-\tau)\hat{\Phi}(\xi-\tau)}\\
G_\tau(\eta) &:= \hat{g}(\eta)\hat{\Phi}(\eta)\overline{\hat{g}\left(\eta+\frac{\tau}{2^{(b-a)j}}\right)\hat{\Phi}\left(\eta+\frac{\tau}{2^{(b-a)j}}\right)}\\
Q_\tau(\xi,\eta) &:= \phi(\xi,\eta)-\phi\left(\xi-\tau,\eta+\frac{\tau}{2^{(b-a)j}}\right).\end{aligned}$$

We claim that whenever $\xi,\eta,\xi-\tau,\eta+\frac{\tau}{2^{(b-a)j}}\in \mathrm{supp}\hat{\Phi}$, we have

$$|\partial_\xi\partial_\eta Q_\tau(\xi,\eta)| \gtrsim |\tau| \tag{4.7}$$

Let's briefly justify (4.7). By the definition of Q_τ and mean value theorem, we need to show that $|\partial_\xi^2\partial_\eta\phi(\xi,\eta)|$ and $|\partial_\xi\partial_\eta^2\phi(\xi,\eta)|$ are bounded below by some positive C. Let t_0 be a root of $F_0(t) := \varphi'(t)=D_t(\xi(t^a+\epsilon_P(t))+\eta(t^b+\epsilon_Q(t)))$, and t_1 be a root of $F_1(t) := D_t(\xi t^a+\eta t^b)$. Let $\phi^*(\xi,\eta) := \xi t_1^a+\eta t_1^b = C\left(\frac{\xi^b}{\eta^a}\right)^{\frac{1}{b-a}}$. Then

$$\phi(\xi,\eta) = \varphi(t_0) = \phi^*(\xi,\eta) + Err(\xi,\eta),$$

where $Err(\xi,\eta) := \xi(t_0^a-t_1^a)+\eta(t_0^b-t_1^b)+\xi\epsilon_P(t_0)+\eta\epsilon_Q(t_0)$. Clearly the mixed derivatives of $\phi^*(\xi,\eta)$ are bounded below by some positive C. It remains to show that $|Err(\xi,\eta)| \le C^{-1}$ for some large C. Since F_0 and F_1 are "close", the difference of their inverses t_0-t_1 (and its derivatives) is also very small (see Definition A.1 and Lemma A.2 in [65] for details). By this observation and the facts that $|\epsilon_P(t_0)|$ and $|\epsilon_Q(t_0)|$ are tiny when N is large enough, we conclude that $|Err(\xi,\eta)|$ is very small compared with 1. This finishes the justification of (4.7).

By (4.7) and Hömander principle (Theorem 1.1 in [50]),

$$\iint F_\tau(\xi)G_\tau(\eta)e^{i2^m Q_\tau}\,d\xi d\eta \lesssim \min\{\|f\|_2^2\|g\|_2^2, 2^{-\frac{m}{2}}|\tau|^{-\frac{1}{2}}\|F_\tau\|_2\|G_\tau\|_2\}. \tag{4.8}$$

Therefore,

$$\begin{aligned}
\|B_{j,m}\|_2^2 &\lesssim 2^{-(b-a)j-m}\int \min\{\|f\|_2^2\|g\|_2^2, 2^{-\frac{m}{2}}|\tau|^{-\frac{1}{2}}\|F_\tau\|_2\|G_\tau\|_2\}\,d\tau \\
&\lesssim 2^{-(b-a)j-m}\left[\int_{|\tau|<\tau_0}\|f\|_2^2\|g\|_2^2\,d\tau + \int_{\tau_0\le|\tau|\lesssim 1} 2^{-\frac{m}{2}}|\tau|^{-\frac{1}{2}}\|F_\tau\|_2\|G_\tau\|_2\right] \\
&\lesssim 2^{-(b-a)j-m}\left[\tau_0\|f\|_2^2\|g\|_2^2 + 2^{-\frac{m}{2}}|\tau_0|^{-\frac{1}{2}}\|F_\tau\|_2\|G_\tau\|_2\right] \\
&\lesssim 2^{-(b-a)j-m}2^{\frac{(b-a)j-m}{3}}\|f\|_2^2\|g\|_2^2,
\end{aligned}$$

from which (4.2) follows.

4.4 σ-uniformity method

We prove Proposition 4.12 and hence finish the proof of Proposition 4.7 in this section. Let $I \subseteq \mathbb{R}$ be a fixed interval. Let $\sigma \in (0,1]$ and $\mathcal{Q}$ be a collection of real-valued functions.

Definition 4.13 *A function $f \in L^2(I)$ is called σ-uniform in $\mathcal{Q}$ if*

$$\left|\int_I f(\xi)e^{-iq(\xi)}\,d\xi\right| \le \sigma\|f\|_{L^2(I)}$$

for all $q \in \mathcal{Q}$.

The main tool of proving Proposition 4.12 is the following theorem, whose proof can be found in Theorem 6.2 in [64].

Theorem 4.14 *Let L be a bounded sub-linear functional from $L^2(I)$ to $\mathbb{C}$. Let S_σ be the set of all L^2 functions that are σ-uniform in $\mathcal{Q}$ and $U_\sigma := \sup\limits_{f\in S_\sigma}\frac{|L(f)|}{\|f\|_{L^2(I)}}$. Then for all functions $f \in L^2(I)$,*

$$|L(f)| \lesssim \max\left\{U_\sigma, \frac{Q_0}{\sigma}\right\}\|f\|_{L^2(I)}, \tag{4.1}$$

where $Q_0 := \sup\limits_{q\in\mathcal{Q}} L(e^{iq})$.

Now we start to estimate U_σ. Recall that we can assume x is restricted in an interval of length $\simeq 2^{(b-a)j+m}$. We fix such an interval and partition it into 2^m intervals of length $\simeq 2^{(b-a)j}$, which are denoted by

$$I_k = [\alpha_k - 2^{(b-a)j}, \alpha_k + 2^{(b-a)j}], k = 1, 2, \ldots, 2^m.$$

To each I_k we assign an enlarged interval

$$I'_k = [\alpha_k - C(2^{(b-a)j} + 2^m), \alpha_k + C(2^{(b-a)j} + 2^m)]$$

such that $x - 2^m(t^b + \epsilon_Q(t)) \in I'_k$ whenever $x \in I_k$ and $t \in \text{supp}\rho$. So $B_{j,m}$ can be partitioned accordingly as

$$\begin{aligned} B_{j,m}(f,g)(x) &= 2^{-\frac{(b-a)j}{2}} \sum_{k=1}^{2^m} \chi_{I_k}(x) \int f * \Phi\left(\frac{x}{2^{(b-a)j}} - 2^m(t^a + \epsilon_P(t))\right) \\ &\qquad\qquad \chi_{I'_k} g * \Phi(x - 2^m(t^b + \epsilon_Q(t)))\rho(t)\, dt \\ &= 2^{-\frac{(b-a)j}{2}} \sum_{k=1}^{2^m} \chi_{I_k}(x) \iint \hat{f}(\xi)\hat{\Phi}(\xi) e^{2\pi i(\frac{\xi}{2^{(b-a)j}} + \eta)x} \hat{g_k}(\eta) I_{\rho,m}(\xi,\eta)\, d\xi d\eta, \end{aligned}$$

where

$$g_k(x) := \chi_{I'_k} g * \Phi(x),$$

and $I_{\rho,m}$ is defined as (4.4). Since $|\xi| \simeq 1$, $I_{\rho,m}$ has rapid decay unless $|\eta| \simeq 1$. So we may insert a cut-off function $\hat{\Phi}(\eta)$ for free in the above integrand.

Pair $B_{j,m}$ with an $h \in L^\infty$,

$$\begin{aligned} \langle B_{j,m}(f,g), h\rangle = 2^{-\frac{(b-a)j}{2}} \sum_{k=1}^{2^m} \int \chi_{I_k}(x) h(x) e^{2\pi i \eta x} \\ \iint \hat{f}(\xi)\hat{\Phi}(\xi) e^{2\pi i \frac{\xi}{2^{(b-a)j}} x} \hat{\Phi}(\eta)\hat{g_k}(\eta) I_{\rho,m}(\xi,\eta)\, d\xi d\eta. \end{aligned}$$

Thanks to the cut-off $\chi_{I_k}(x)$, and we can replace $e^{2\pi i \frac{\xi}{2^{(b-a)j}} x}$ with $e^{2\pi i \frac{\xi}{2^{(b-a)j}} \alpha_k}$ using Taylor expansion. Thus essentially,

$$\langle B_{j,m}, h\rangle \sim 2^{-\frac{(b-a)j}{2}} \sum_{k=1}^{2^m} \int \check{h_k}(\eta)\Gamma_k(\eta)\hat{g_k}(\eta)\, d\eta,$$

where

$$\begin{aligned} &h_k(x) := \chi_{I_k} h(x), \text{ and} \\ &\Gamma_k(\eta) := \hat{\Phi}(\eta) \int \hat{f}(\xi)\hat{\Phi}(\xi) e^{2\pi i \frac{\xi}{2^{(b-a)j}} \alpha_k} I_{\rho,m}(\xi,\eta) d\xi. \end{aligned}$$

As before, we can replace $I_{\rho,m}$ with $2^{-\frac{m}{2}} e^{i2^m \phi(\xi,\eta)}$ and thus

$$\Gamma_k(\eta) \sim 2^{-\frac{m}{2}} \hat{\Phi}(\eta) \int \hat{f}(\xi)\hat{\Phi}(\xi) e^{i(2^m \phi(\xi,\eta) + \frac{\xi}{2^{(b-a)j}}\alpha_k)} d\xi.$$

Let $\mathcal{Q} := \{A(\xi^{\frac{b}{b-a}} + \epsilon(\xi)) + B\xi\}$, where $A, B \in \mathbb{R}$, $|A| \simeq a^m$, and $\epsilon(\xi)$ and its derivatives are $\lesssim 2^{-CN}$. Then $2^m\phi(\xi,\eta) + \frac{\xi}{2^{(b-a)j}}\alpha_k \in \mathcal{Q}$ for large N. Let $\hat{f}$ be σ-uniform in $\mathcal{Q}$. Then

$$\|\Gamma_k\|_\infty \lesssim 2^{-\frac{m}{2}}\sigma\|f\|_2,$$

and thus

$$\begin{aligned}\langle B_{j,m}(f,g),h\rangle &\lesssim 2^{-\frac{(b-a)j}{2}}\sum_{k=1}^{2^m}\|\Gamma_k\|_\infty\|h_2\|_2\|g_k\|_2\\ &\lesssim \sigma\|f\|_2\|h\|_\infty(\sum_k\|g_k\|_2^2)^{\frac{1}{2}}\\ &\lesssim \begin{cases}\sigma\|f\|_2\|g\|_2\|h\|_\infty \text{ when } (b-a)j \geq m\\ \sigma 2^{\frac{m-(b-a)j}{2}}\|f\|_2\|g\|_2\|h\|_\infty \text{ when } (b-a)j \leq m.\end{cases}\end{aligned} \tag{4.2}$$

This finishes the computation of U_σ.

Now we turn to Q_0. Let $\hat{f}(\xi) = e^{iq(\xi)}$ for some $q \in \mathcal{Q}$. Let $h \in L^\infty$ be a function supported on an interval of length $\simeq 2^{(b-a)j+m}$ as before. Define

$$\begin{aligned}\Lambda_q(g,h) &:= \langle B_{j,m},h\rangle\\ &= 2^{-\frac{(b-a)j}{2}}\iiint\hat{\Phi}(\xi)e^{i(A(\xi^{\frac{b}{b-a}}+\epsilon(\xi))+B\xi)}e^{i\xi\left(\frac{x}{2^{(b-a)j}}-2^m(t^a+\epsilon_P(t))\right)}\,d\xi\\ &\qquad g*\Phi(x-2^m(t^b+\epsilon_Q(t)))\rho(t)\,dtdx.\end{aligned}$$

Our goal is to show

$$|\Lambda_q(g,h)| \lesssim 2^{-\epsilon m}\|g\|_2\|h\|_\infty. \tag{4.3}$$

This means that $Q_0 \lesssim 2^{-\epsilon m}$. Combining this with (4.2), Proposition 4.12 will be proved by Theorem 4.14.

To prove (4.3), we will use the strategy similar to the previous cases: rescaling, stationary phase, and (local) TT*. Let

$$|\tilde{\Lambda}_q(g,h)| := \iint \mathcal{F}(y,t)g*\tilde{\Phi}\left(y - \frac{t^b+\epsilon_Q(t)}{2^{(b-a)j}}\right)\rho(t)dth(y)dy,$$

where $\hat{\tilde{\Phi}}(\xi) := \hat{\Phi}\left(\frac{\xi}{2^{(b-a)j+m}}\right)$ and

$$\mathcal{F}(y,t) := 2^{\frac{m}{2}}\int\hat{\Phi}(\xi)e^{iA\left(\xi^{\frac{b}{b-a}}+\epsilon(\xi)+\frac{2^m}{A}(y-(t^a+\epsilon_P(t))+\frac{B}{2^m})\xi\right)}\,d\xi.$$

By rescaling, (4.3) follows from the estimate

$$|\tilde{\Lambda_q}(g,h)| \lesssim 2^{-\epsilon m}\|g\|_2\|h\|_\infty \tag{4.4}$$

for any $h \in L^\infty$ supported in an interval of length $\simeq 1$.

Write

$$\mathcal{F}(y,t) = 2^{\frac{m}{2}} \int \hat{\Phi}(\xi) e^{iA\left(\xi^{\frac{b}{b-a}}+\epsilon(\xi)+C'(y-(t^a+\epsilon_P(t))+B')\xi\right)} d\xi,$$

where $C' := \frac{2^m}{A} \simeq 1$ and $B' := \frac{B}{2^m}$. For simplicity we drop C' from now on. Let $\zeta(z)$ be the solution of $(\xi^{\frac{b}{b-a}} + \epsilon(\xi) + z\xi)' = 0$ and $\beta(z) := \zeta(z)^{\frac{b}{b-a}} + \epsilon(\zeta(z)) + z\zeta(z)$. Then stationary phase methods gives that

$$\mathcal{F}(y,t) \sim e^{iA\beta(y-(t^a+\epsilon_P(t))+B')}\hat{\Phi}(\zeta(y-(t^a+\epsilon_P(t))+B')).$$

Since the term $\hat{\Phi}(\zeta(y-(t^a+\epsilon_P(t))+B'))$ can be dropped by Fourier expansion, we have

$$\tilde{\Lambda_q}(g,h) \sim \iint e^{iA\beta(y-(t^a+\epsilon_P(t))+B')} g * \tilde{\Phi}\left(y - \frac{t^b+\epsilon_Q(t)}{2^{(b-a)j}}\right)\rho(t)dth(y)dy.$$

This finishes the use of the stationary phase method. The last step is to use TT* method to obtain the decay. Change variable $s = t^b + \epsilon_Q(t)$. Define three new functions κ, l and $\tilde{\rho}$ by $t = \kappa(s)$, $l(s) = \kappa(s)^a + \epsilon_P(\kappa(s))$ and $\tilde{\rho}(s)ds = \rho(t)dt$. Then

$$\begin{aligned}\tilde{\Lambda_q}(g,h) &= \iint e^{iA\beta(y-l(s)+B')} g * \tilde{\Phi}\left(y - \frac{s}{2^{(b-a)j}}\right)\tilde{\rho}(s)ds\, h(y)dy \\ &\lesssim \|\Delta(h)\|_2\|g\|_2,\end{aligned}$$

where

$$\Delta(h)(y) := \int e^{iA\beta(y+\frac{s}{2^{(b-a)j}}-l(s)+B')} h\left(y + \frac{s}{2^{(b-a)j}}\right)\tilde{\rho}(s)\, ds.$$

It remains to show

$$\|\Delta(h)\|_2^2 \lesssim 2^{-\epsilon m}\|h\|_\infty^2. \tag{4.5}$$

A straightforward calculation gives

$$\|\Delta(h)\|_2^2 = \iiint e^{iAO_\tau(u,v)} H_\tau(u)\Theta_\tau(v)dudv\, d\tau, \tag{4.6}$$

where

$$H_\tau(u) := h(u)h\left(u + \frac{\tau}{2^{(b-a)j}}\right),$$
$$\Theta_\tau(v) := \tilde{\rho}(v)\tilde{\rho}(v+\tau),$$

and

$$O_\tau(u,v) := \beta(u - l(v) + B') - \beta\left(u + \frac{\tau}{2^{(b-a)j}} - l(v+\tau) + B'\right).$$

By the same idea in the proof of (4.7), we see that the mixed partial derivatives of $O_\tau(u,v)$ is bounded below by $C|\tau|$. By the operator version of van der Corput lemma (see for example Lemma 5.8 in [65]), we have

$$\iint e^{iAO_\tau(u,v)} H_\tau(u)\Theta_\tau(v)dudv \lesssim \min\{1, |2^m\tau|^{-\epsilon}\}\|H_\tau\|_2\|\Theta_\tau\|_2. \tag{4.7}$$

By definitions, it is easy to see that $\|H_\tau\|_2 \lesssim \|h\|_\infty^2$ and $\|\Theta_\tau\|_2 \lesssim 1$. So we can break the integral against τ in (4.6) into two parts as before: $|\tau| \leq \tau_0$ and $\tau_0 < |\tau| \lesssim 1$, and use the estimate (4.7) to obtain the desired result (4.5).

4.5 L^r estimates and the maximal function

We start to prove Proposition 4.8 and thus finish the proof of Theorem 4.2. Rewrite $T_{j,m}$ as

$$\begin{aligned} &T_{j,m}(f,g)(x) = \\ &\quad\int f * \Phi_{aj+m}\left(x - \frac{t^a + \epsilon_P(t)}{2^{aj}}\right) g * \Phi_{bj+m}\left(x - \frac{t^b + \epsilon_Q(t)}{2^{bj}}\right)\rho(t)\,dt, \end{aligned} \tag{4.1}$$

where $\Phi_k(x) := 2^k\Phi(2^kx)$. Let

$$\begin{aligned} &T^m(f,g)(x) := \\ &\sum_{j>N}\int \left| f * \Phi_{aj+m}\left(x - \frac{t^a + \epsilon_P(t)}{2^{aj}}\right) g * \Phi_{bj+m}\left(x - \frac{t^b + \epsilon_Q(t)}{2^{bj}}\right)\rho(t)\right| dt. \end{aligned} \tag{4.2}$$

It suffices to prove the boundedness of T^m with norm $\lesssim m$.

Given any measurable sets F_1, F_2, F_3 of finite measure, define

$$\Omega := \bigcup_{i=1}^{2}\left\{x : \mathfrak{m}\chi_{F_i} > C\frac{|F_i|}{|F_3|}\right\},$$

where $\mathfrak{m}$ denotes the Hardy-Littlewood maximal operator. Let $F_3' := F_3 \setminus \Omega$, which has measure no less than

$\frac{|F_3|}{2}$ when C is chosen large enough. By standard interpolation, we need to show that

$$|\langle T^m(f,g),h\rangle| \lesssim m|F_1|^{\frac{1}{p}}|F_2|^{\frac{1}{q}}|F_3|^{1-\frac{1}{r}}, \tag{4.3}$$

for all $|f| \le \chi_{F_1}$, $|g| \le \chi_{F_2}$, $p,q \in (1,\infty)$, $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$.

We first remove some error terms related with Ω, Define $\Omega_k := \{x : \mathrm{dist}(x,\Omega^c) \ge 2^{-k}\}$ and let $\psi_k(x) = \chi_{\Omega_k^c} * \tilde{\psi}_k(x)$, where $\tilde{\psi} \in \mathcal{S}(\mathbb{R})$ is Fourier supported in $[-2^k, 2^k]$. It turns out that in proving (4.3) we can replace $T^m(f,g)$ with

$$\begin{aligned}(T')^m(f,g(x) := \sum_{j>N} \int &\left|\psi_{aj+m} f * \Phi_{aj+m}\left(x - \frac{t^a + \epsilon_P(t)}{2^{aj}}\right)\right| \\ &\left|\psi_{bj+m} g * \Phi_{bj+m}\left(x - \frac{t^b + \epsilon_Q(t)}{2^{bj}}\right)\rho(t)\right| dt.\end{aligned} \tag{4.4}$$

This is because the difference of these two operators has good control. See Lemma 6.3 in [65], whose proof is based on a discussion about whether $x-t$ (or $x - \frac{t^b+\epsilon_Q(t)}{2^{bj}}$) belongs to Ω or not. That proof can be easily modified to include the $x - \frac{t^a+\epsilon_P(t)}{2^{aj}}$ case. So we focus on proving the following variant of (4.3), with T^m being replaced by $(T')^m$:

$$|\langle (T')^m(f,g),h\rangle| \lesssim m|F_1|^{\frac{1}{p}}|F_2|^{\frac{1}{q}}|F_3|^{1-\frac{1}{r}}. \tag{4.5}$$

Time-frequency analysis must be employed to prove (4.5). For any integers n,j, define $I_{n,j} := [2^{-j}n, 2^{-j}(n+1))$. Let $1^*_{n,j}(x) := \chi_{I_{n,j}} * \theta_{j+m}(x)$, where $\theta_k \in \mathcal{S}(\mathbb{R})$ is Fourier supported on $[-2^{-10}2^k, 2^{-10}2^k]$. Then

$$\begin{aligned}(T')^m(f,g)(x) = \sum_{j>N} \int &\left|\sum_{n\in\mathbb{Z}} f_{n,m,j}\left(x - \frac{t^a + \epsilon_P(t)}{2^{aj}}\right)\right| \\ &\left|\sum_{n\in\mathbb{Z}} g_{n,m,j}\left(x - \frac{t^b + \epsilon_Q(t)}{2^{bj}}\right)\rho(t)\right| dt,\end{aligned} \tag{4.6}$$

where

$$f_{n,m,j}(x) := 1^*_{n,aj}\psi_{aj+m} f * \Phi_{aj+m}(x);$$

$$g_{n,m,j}(x) := 1^*_{n,aj}\psi_{bj+m} g * \Phi_{bj+m}(x).$$

Let $S_0 := \{(j,n) \in \mathbb{Z}^2 : j > N\}$. For any $S \subseteq S_0$, define $S_j := \{n \in \mathbb{Z} : (j,n) \in S\}$ and

$$\Lambda_S(f,g) := \sum_{j>N} \iint \left| \sum_{n \in S_j} f_{n,m,j}\left(x - \frac{t^a + \epsilon_P(t)}{2^{aj}}\right) \right| \left| \sum_{n \in S_j} g_{n,m,j}\left(x - \frac{t^b + \epsilon_Q(t)}{2^{bj}}\right) \right| |\rho(t)|\, dt dx \tag{4.7}$$

We aim to prove that for any finite $S \subseteq S_0$,

$$\Lambda_S(f,g) \lesssim m|F_1|^{\frac{1}{p}}|F_2|^{\frac{1}{q}}|F_3|^{1-\frac{1}{r}}, \tag{4.8}$$

from which (4.5) follows. The strategy is to organize elements in S into union of subsets called maximal trees. On each tree $\mathcal{T} \in S$, $\Lambda_{\mathcal{T}}(f,g)$ can be controlled. Let's perform some reductions on $\Lambda_{\mathcal{T}}(f,g)$ as in [65]. By a change of variable $u = x - \frac{t^b + \epsilon_Q(t)}{2^{bj}}$,

$$\Lambda_{\mathcal{T}}(f,g) = \sum_{j>N} \iint \left| \sum_{n \in \mathcal{T}_j} f_{n,m,j}\left(u - tr(t)\right) \rho(t) \right| dt \left| \sum_{n \in \mathcal{T}_j} g_{n,m,j}\left(u\right) \right| du, \tag{4.9}$$

where $tr(t) := \frac{t^a + \epsilon_P(t)}{2^{aj}} - \frac{t^b + \epsilon_Q(t)}{2^{bj}}$. Since $tr(t) \simeq \frac{t^a}{2^{aj}}$, we have

$$\int \left| \sum_{n \in \mathcal{T}_j} f_{n,m,j}\left(u - tr(t)\right) \rho(t) \right| dt \lesssim \mathfrak{m}\left(\sum_{n \in \mathcal{T}_j} f_{n,m,j} \right)(u). \tag{4.10}$$

From here the translation determined by t disappears and thus we can use the same calculations as in [65]. We omit the details. This finishes the proof of Proposition 4.8 and Theorem 4.2.

Now we show how to use Lemma 4.4 and Theorem 4.6 to obtain the boundedness of the bilinear maximal function $\mathcal{M}_{P,Q}$, proving Theorem 4.3. By triangle inequality, it suffices to consider the following operator

$$T^*(f,g)(x) := \sup_{j \in \mathbb{Z}} T_j(f,g)(x), \tag{4.11}$$

where T_j is defined as in (4.1) and f, g are non-negative. By Lemma 4.4 and symmetry, we can further assume that the supremum is taken over $j > N$ for some large N.

As before, decompose $T_j = \sum_{(m,n) \in \mathbb{Z}^2} T_{j,m,n}$ (see (4.13)). Let $E := \{|m-n| \gtrsim 1\} \cup \{\max\{m,n\} \leq 0\}$.

Using Fourier expansion and integration by parts (or Taylor expansion), it is easy to see that

$$\sup_{j>N}\left|\sum_{(m,n)\in E} T_{j,m,n}(f,g)(x)\right| \lesssim \mathfrak{m}f(x)\mathfrak{m}g(x).$$

By Hölder inequity and the boundedness of $\mathfrak{m}$, $\sup_{j>N}|\sum_{(m,n)\in E} T_{j,m,n}(f,g)|$ is bounded from $L^p \times L^q$ into L^r.

For $(m,n)\in\mathbb{Z}^2\setminus E$, we can assume without loss of generality that $m=n$. In this case we bound $\sup_{j>N}|T_{j,m,m}(f,g)(x)|$ crudely by $\sum_{j>N}|T_{j,m,m}(f,g)(x)|$. Since each $\sum_{j>N}|T_{j,m,m}|$ is bounded with $2^{-\epsilon m}$ decay in norm by Theorem 4.6, we conclude that $\sup_{j>N}|\sum_{(m,n)\in\mathbb{Z}^2\setminus E} T_{j,m,n}|$ is bounded. This finishes the proof of Theorem 4.3.

4.6 Necessary conditions

It is then natural to ask

Open Problem 4.15 *For which pairs of polynomials P and Q is the condition $r>\frac{d}{d+1}$ necessary to the L^r-boundedness of $B_{P,Q}$?*

We can further ask:

Open Problem 4.16 *Given polynomials P and Q, what is the minimal value of r that guarantees the L^r-boundedness of $B_{P,Q}$?*

The purpose of this section is to give partial answers to these questions. We show that for many pairs of polynomials, $r>\frac{d}{d+1}$ is indeed the best (up to endpoint) range for the L^r-boundedness of $B_{P,Q}$. More precisely,

Theorem 4.17 *Let P and Q be two polynomials without constant term. Assume that the operator $B_{P,Q}(f,g)(x) = \int f(x-P(t))g(x-Q(t))\frac{dt}{t}$ is bounded from $L^{p_1}\times L^{p_2}$ into L^r, $p_1,p_2\in(1,\infty)$, $\frac{1}{p_1}+\frac{1}{p_2}=\frac{1}{r}$. If $P'(t)>0$ for all $t\neq 0$, then $r\geq\frac{d}{d+1}$, where d is the correlation degree of P and Q.*

Remarks. (1) By simple arguments, the criterion in the above theorem also includes the case $P'(t)<0$ for all $t\neq 0$. By symmetry, the theorem also holds if the condition is imposed on the polynomial Q.

(2). Theorem 4.17 itself answers partially Open Problem 4.15. Together with the main theorem in [23], it also answers Open Problem 4.16 for a large range of pairs of polynomials.

(3). We believe that the criterion given in Theorem 4.17 could be weaken. The weakest condition is conjectured as follows:

Conjecture 4.18 *Let P and Q be two polynomials without constant term. Assume that the operator $B_{P,Q}(f,g)(x) = \int f(x-P(t))g(x-Q(t))\frac{dt}{t}$ is bounded from $L^{p_1} \times L^{p_2}$ into L^r for $p_1, p_2 \in (1,\infty)$ and $\frac{1}{p_1} + \frac{1}{p_2} = \frac{1}{r}$. Then $r \geq \frac{d}{d+1}$, where d is the correlation degree of P and Q, as long as not both P and Q are even.*

The criterion in Theorem 4.17 essentially says that P is strictly monotonic and thus graphically P is similar to an odd function. How to build the bridge between "odd" (in Theorem 4.17) and "not even" (in Conjecture 4.18) seems to be a very difficult problem. It is also possible that Conjecture 4.18 is in fact false, and Theorem 4.17 could be the best answer to Open Problem 4.15.

In the proof we will use C to denote a positive large constant whose value may change from line to line. Such constant may depend on the polynomials P and Q. $A \lesssim B$ is short for $A \leq CB$ and $A \ll B$ means $CA \leq B$ for some large C. χ_E will be used to denote the indicator function of the set E.

Given two polynomials P and Q without constant term, assume the correlation degree of P and Q is d. Let t_0 be a non-zero real root of $P' - Q'$ with multiplicity d. We may assume $t_0 > 0$, as the other case can be handled in a similar way. The assumption $P'(t) > 0$ for all non-zero t implies that P is strictly increasing. Therefore, we have $P(t_0) > 0$ as $P(0) = 0$. Let $0 < \delta \ll P(t_0)$ be small. We will use the following special choice of f and g:

$$\begin{cases} f = \chi_{[-\delta,\delta]}, \\ g = \chi_{[P(t_0)-Q(t_0)-\delta, P(t_0)-Q(t_0)+\delta]}. \end{cases} \tag{4.12}$$

By the boundedness of $B_{P,Q}$ and straightforward calculation, we have

$$\|B_{P,Q}(f,g)\|_r \lesssim \|f\|_{p_1}\|g\|_{p_2} \lesssim \delta^{\frac{1}{r}} \tag{4.13}$$

In what follows, we aim to get a lower bound of $\|B_{P,Q}(f,g)\|_r$ in terms of powers of δ. Recall $B_{P,Q}(f,g)(x) = \int f(x-P(t))g(x-Q(t))\frac{dt}{t}$. We will achieve our goal by properly restricting x and t in the definition of $B_{P,Q}(f,g)(x)$. Consider the interval

$$I = \left[P(t_0) - \frac{\delta^{\frac{1}{d+1}}}{A}, P(t_0) + \frac{\delta^{\frac{1}{d+1}}}{A}\right],$$

where A is a large constant to be determined later. We will only consider those $x \in I$ when calculating $\|B_{P,Q}(f,g)\|_r$. Clearly $x > 0$ when A is large. Also note that we can assume $t > 0$, otherwise we would have $P(t) < 0$ and

$$|x - P(t)| > x \geq P(t_0) - \frac{\delta^{\frac{1}{d+1}}}{A} > \delta,$$

which implies $f(x - P(t)) = 0$ and $B_{P,Q}(f, g)(x) = 0$.

Since f, g are non-negative and t is positive, we can further restrict t in order to get a lower bound $B_{P,Q}(f, g)(x)$. For any $x \in I$, define

$$J_x = \left\{t > 0 : |P(t) - x| < \frac{\delta}{2}\right\}.$$

By the definition of f (4.12), $f(x - P(t)) = 1$ when $t \in J_x$. We claim that the same holds for g:

Claim 4.19 *$g(x - Q(t)) = 1$ whenever $x \in I$ and $t \in J_x$.*

Proof: Fix $x \in I$, and let $t \in J_x$. By the definitions of I and J_x,

$$|P(t) - P(t_0)| \leq |P(t) - x| + |P(t_0) - x| \leq \frac{\delta}{2} + \frac{\delta^{\frac{1}{d+1}}}{A} \lesssim \frac{\delta^{\frac{1}{d+1}}}{A} \tag{4.14}$$

for small δ. Invoke mean value theorem,

$$|t - t_0| = |P^{-1}(P(t)) - P^{-1}(P(t_0))| = |(P^{-1})'(\xi)||P(t) - P(t_0)| \tag{4.15}$$

for some $\xi \in \mathbb{R}$. By inverse function theorem, $|(P^{-1})'(\xi)|$ can never be ∞ as P' is never 0. Therefore, (4.14) and (4.15) give that

$$|t - t_0| \lesssim |P(t) - P(t_0)| \lesssim \frac{\delta^{\frac{1}{d+1}}}{A}, \tag{4.16}$$

Now using the assumption that t_0 is a root of $P' - Q'$ with multiplicity d, we see that

$$\begin{aligned} &|x - Q(t) - (P(t_0) - Q(t_0))| \\ &\qquad \leq |x - P(t)| + |P(t) - Q(t) - (P(t_0) - Q(t_0))| \lesssim \frac{\delta}{2} + |t - t_0|^{d+1} \end{aligned}$$

By (4.16), we can bound $|x - Q(t) - (P(t_0) - Q(t_0))|$ by

$$\frac{\delta}{2} + \frac{\delta}{A^{d+1}} < \delta \tag{4.17}$$

if A is chosen large enough. Hence $g(x - Q(t)) = 1$ by the definition of g (4.12). ■

Applying mean value theorem and inverse function theorem again, we see that the measure of J_x is bounded below by

$$\left|P^{-1}\left(x+\frac{\delta}{2}\right)-P^{-1}\left(x-\frac{\delta}{2}\right)\right|\gtrsim\delta.$$

In sum, we have $f(x-P(t))g(x-Q(t))=1$ when $x\in I$ and t lies in an interval of length at least δ. Therefore,

$$|B_{P,Q}(f,g)(x)|\gtrsim\delta,\ x\in I.$$

Since the length of the interval I is $\gtrsim\delta^{\frac{1}{d+1}}$,

$$\|B_{P,Q}(f,g)\|_r\gtrsim\delta\cdot\delta^{\frac{1}{(d+1)r}}. \tag{4.18}$$

Combine (4.13) and (4.18), and we see that

$$\delta^{\frac{1}{r}}\gtrsim\delta^{1+\frac{1}{(d+1)r}}. \tag{4.19}$$

As (4.19) holds for arbitrarily small δ, we must have $r\geq\frac{d}{d+1}$, as desired.

Chapter 5

A hybrid of bilinear Hilbert transform and paraproduct

We now present the result in [25]. We prove the boundedness of a class of tri-linear operators consisting of a quasi piece of bilinear Hilbert transform whose scale equals to or dominates the scale of its linear counter part. Such type of operators is motivated by the tri-linear Hilbert transform and its curved versions.

5.1 Introduction

5.1.1 Background

In a pair of breakthrough papers [60, 61], Lacey and Thiele proved the boundedness property of the bilinear Hilbert transform (BHT)

$$B(f_1, f_2)(x) = p.v. \int f_1(x-t) f_2(x+t) \frac{1}{t}\, dt.$$

Many interesting results about multilinear operators have been established in the spirit of Lacey-Thiele's method. However, L^p-boundedness of tri-linear Hilbert transform (THT)

$$T(f_1, f_2, f_3)(x) = p.v. \int f_1(x-t) f_2(x-2t) f_3(x-3t) \frac{1}{t}\, dt.$$

is still unknown. One difficulty arises from certain non-linear issue hidden in the trilinear structure. This is one of the main reasons motivating Li to study BHT along curves [64], say

$$H_\Gamma(f_1, f_2)(x) = p.v. \int f_1(x-t) f_2(x-t^d) \frac{1}{t}\, dt, \text{ where } d \geq 2 \text{ is an integer.}$$

In [64], H_Γ is split into two operators according to the efficiency of some oscillatory integral estimate (stationary phase vs. non-stationary phase). One of the two operators is a paraproduct of the form $\Pi_\Gamma(f_1, f_2) = \sum_k f_{1k} f_{2k}$ [63] that is more complex than the classical Coifman-Meyer paraproduct [18]. Although it turns out Π_Γ is slightly simpler than BHT, the proof of its boundedness already requires sophisticated multi-scale time-frequency analysis that is essential in the study of BHT. Hence it is reasonable to expect that tri-linear analogues of the paraproduct Π_Γ would be easier to handle than THT, but at the same time the study of such tri-linear operators could provide some new insights to THT.

The definition of tri-linear correspondence of $\Pi_\Gamma(f_1, f_2)$ was given in [26], where the author and Li introduced the following class of operators $T^{\alpha,\beta}$ that can be viewed a hybrid of BHT and paraproduct:

$$T^{\alpha,\beta}(f_1, f_2, f_3)(x) = \sum_{k\in\mathbb{Z}} H^{\alpha,k}(f_1, f_2)(x) f_3^{\beta,k}(x), \tag{5.1}$$

where

$$\begin{cases} H^{\alpha,k}(f_1, f_2)(x) = \iint_{\mathbb{R}^2} \widehat{f_1}(\xi_1)\widehat{f_2}(\xi_2) e^{2\pi i(\xi_1+\xi_2)x} \widehat{\Phi_1}\left(\frac{\xi_1-\xi_2}{2^{\alpha k}}\right) d\xi_1 d\xi_2, \\ \\ f^{\beta,k}(x) = \int_{\mathbb{R}} \widehat{f}(\xi) e^{2\pi i \xi x} \widehat{\Phi_2}\left(\frac{\xi}{2^{\beta k}}\right) d\xi. \end{cases} \tag{5.2}$$

Here α, β are non-zero positive real numbers, and various conditions (about smoothness, support, etc) can be imposed on the cut-off functions $\widehat{\Phi_1}$ and $\widehat{\Phi_2}$.

$T^{\alpha,\beta}$ is closely related with THT along curves. For example, one promising way to prove the boundedness of $T_C(f_1, f_2, f_3)(x) = p.v. \int f_1(x-t) f_2(x+t) f_3(x-t^d) \frac{dt}{t}$ is to study $T^{1,d}$ first (See [64] for a similar approach in the bilinear setting). The following theorem is proved in [26].

Theorem 5.1 ([26], Theorem 1.2) *Let Φ_1 and Φ_2 be smooth functions satisfying supp $\widehat{\Phi_1} \subseteq [9, 10]$ and supp $\widehat{\Phi_2} \subseteq [-1, 1]$. Assume $\alpha = \beta \neq 0$. Then the operator $T^{\alpha,\beta}$ defined by (5.1)(5.2) is bounded from $L^{p_1} \times L^{p_2} \times L^{p_3}$ to L^p, $\frac{1}{p} = \frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{p_3}$, whenever $(p_1, p_2, p_3) \in D = \{(p_1, p_2, p_3) \in (1, \infty)^3 : \frac{1}{p_1} + \frac{1}{p_2} < \frac{3}{2}\}$.*

Remarks. (1) Strictly speaking, this theorem is proved in [26] only in the case $\alpha = \beta = 1$, but this restriction is inessential. The proof given in [26] works for any homogeneous-scale case.

(2) The intervals $[9, 10]$ and $[-1, 1]$ in the assumptions of Theorem 5.1 are not essential. The point is that $\widehat{\Phi_1}$ should be supported away from 0 and $\widehat{\Phi_2}$ should be supported near 0.

(3) We conjectured that the condition $\alpha = \beta$ can be dropped in the above theorem, but the proof given in [26] relies on the homogeneity of the scales. Let us briefly analyze the difficulties in the case $\alpha \neq \beta$ here. Assume $0 < \alpha < \beta$ and let $k \geq 2$ be an integer. After wave packet decomposition, the tile associated with $f_3^{\beta,k}$ dominates the other two tiles (associated with f_1 and f_2) in frequency space as supp $\widehat{f_3^{\beta,k}}$ has a much larger scale $2^{\beta k}$. This will also introduce a long tile for the fourth function f_4 in the 4-linear form $\langle T^\alpha(f_1, f_2, f_3), f_4\rangle$: see Figure 5.1. As there are two long tiles and one of them contains the origin,

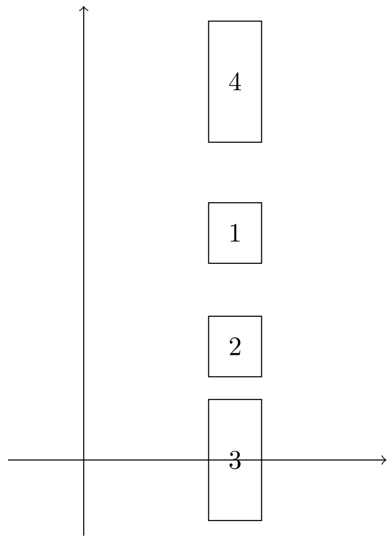

Figure 5.1: Tile structure of $T^{\alpha,\beta}$, $\alpha < \beta$, $k \geq 2$

the situation is difficult to handle even we use telescoping techniques that are powerful in some uniform estimates ([42, 62, 104]).

5.1.2 Main results and application

The purpose of this paper is to investigate other instances of $T^{\alpha,\beta}$, including some non-homogeneous-scale cases. We would like to switch the roles of $\widehat{\Phi_1}$ and $\widehat{\Phi_2}$, i.e. assume that $\widehat{\Phi_1}$ is supported near the origin and $\widehat{\Phi_2}$ is supported away from 0 (instead of the other way around in Theorem 5.1). In this case, $H^{\alpha,k}$ is no longer a piece of BHT at certain scale: we may call it a *quasi piece* of BHT. Surprisingly we can obtain the same range of boundedness as before, even in some cases with non-homogeneous scales (See Theorem 5.3 below). More precisely, we have

Theorem 5.2 *Let Φ_1 and Φ_2 be smooth bump functions satisfying supp $\widehat{\Phi_1} \subseteq [-1,1]$ and supp $\widehat{\Phi_2} \subseteq [9,10]$. Let $\alpha = \beta \neq 0$. Then the operator $T^{\alpha,\beta}$ defined by* (5.1)(5.2) *is bounded from $L^{p_1} \times L^{p_2} \times L^{p_3}$ to L^p for any $(p_1,p_2,p_3) \in D = \{(p_1,p_2,p_3) \in (1,\infty)^3 : \frac{1}{p_1} + \frac{1}{p_2} < \frac{3}{2}\}$, $\frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{p_3} = \frac{1}{p}$.*

The proof of Theorem 5.2 uses Lacey-Thiele's ideas about BHT. However, it should be noted that because of the quasi pieces of BHT, the 4-tile structure of the operator $T^{\alpha,\alpha}$ quite different from the tri-tile structure of BHT (see Figure 5.3 for a comparison): the loss of one tile (1-tile and 2-tile are identical) forces us to mainly work with only two tiles as opposed to three tiles in BHT. The presence of a Littlewood-Paley piece (3-tile), however, will be of great help (see the proof of Proposition 5.10).

Using Theorem 5.2 together with Theorem 5.1, we can derive the boundedness property of positive truncations of $T^{\alpha,\beta}$ in some non-homogeneous-scale cases.

Theorem 5.3 *Let Φ_1 and Φ_2 be smooth bump functions satisfying supp $\widehat{\Phi_1} \subseteq [-1,1]$ and supp $\widehat{\Phi_2} \subseteq [9,10]$. Assume $\alpha > \beta > 0$. Define a positive truncation of $T^{\alpha,\beta}$ by*

$$T_N^{\alpha,\beta}(f_1,f_2,f_3)(x) = \sum_{k\geq N} H^{\alpha,k}(f_1,f_2)(x) f_3^{\beta,k}(x),\ N\in\mathbb{N}, \tag{5.3}$$

where $H^{\alpha,k}$ and $f_3^{\beta,k}$ are given in (5.2). Then for any $N \geq 10\alpha/\beta$, the operator $T_N^{\alpha,\beta}$ is bounded from $L^{p_1}\times L^{p_2}\times L^{p_3}$ into L^p for any $(p_1,p_2,p_3)\in D = \{(p_1,p_2,p_3)\in(1,\infty)^3 : \frac{1}{p_1}+\frac{1}{p_2}<\frac{3}{2}\}$, $\frac{1}{p_1}+\frac{1}{p_2}+\frac{1}{p_3}=\frac{1}{p}$.

Remarks. (1) The choice of intervals $[-1,1]$ and $[9,10]$ in the above two theorems are not important. The key is that $\widehat{\Phi_1}$ should be supported near 0 and $\widehat{\Phi_2}$ should be supported away from 0.

(2) One of anticipated applications of Theorem 5.3 is to use boundedness of $T_N^{d,1}$ to prove that of one prototype of THT along polynomial curves

$$T^C(f_1,f_2,f_3)(x) = p.v.\int_{-1}^{1} f_1(x-t)f_2(x-t^d)f_3(x+t^d)\,\frac{dt}{t}.$$

Just like the relationship between $H_\Gamma(f_1,f_2)(x) = p.v.\int f_1(x-t)f_2(x-t^d)\frac{1}{t}\,dt$ and the paraproduct $\Pi_\Gamma(f_1,f_2) = \sum_k f_{1k}f_{2k}$ studied in [63], T^C can be written as the sum of finitely many operators of the form $T_N^{\alpha,\beta}$ (plus some other terms). The condition $N\geq 10\alpha/\beta$ in Theorem 5.3 is assumed only for technical reasons and it does not affect the application as each scale of T^C (after the standard dyadic decomposition $\frac{1}{t}=\sum_k \rho_k(t)$) is trivially bounded. The reason that we only consider the positive truncation instead of $T^{\alpha,\beta}$ itself is that $|t|\leq 1$ in the definition of T^C.

(3) Under the assumptions on $\widehat{\Phi_1}$ and $\widehat{\Phi_2}$ in Theorem 5.3, Figure 5.2 illustrates the worst case of the tri-tile structure of $T_N^{\alpha,\beta}$ with $\alpha>\beta$ at any positive scale k. The two identical long tiles seems to be very problematic. The key to resolve this issue is to reduce the study of $T_N^{\alpha,\beta}$ with $\alpha>\beta$ to that of $T^{\beta,\beta}$ (homogeneous case) by a telescoping argument. The details are provided in Section 5.6.

5.1.3 Notations

Throughout the paper we will use C to denote a positive constant whose value may change from line to line. We may add one or more subscripts to C to emphasize dependence of C. $A\lesssim B$ is short for $A\leq CB$ and $A\lesssim_N B$ means $A\leq C_N B$. If $A\lesssim B$ and $B\lesssim A$, then we write $A\simeq B$. χ_E and $|E|$ will be used to denote the characteristic function and the Lebesgue measure of the set E, respectively.

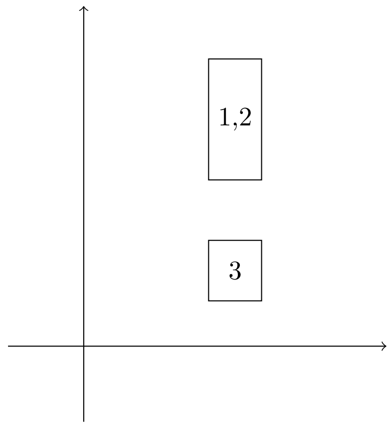

Figure 5.2: tri-tile structure of $T_N^{\alpha,\beta}$, $\alpha > \beta > 0$, $k \geq 1$

5.2 Reduction to model form

The goal of this section is to reduce Theorem 5.2 to the study of a model form using standard wave packet decomposition process. For notational convenience, we assume $\alpha = \beta = 1$ in the proof. The general case can be handled the same way.

Let $\mathcal{S}(\mathbb{R})$ denote the class of Schwartz functions on $\mathbb{R}$. Given $f_j \in \mathcal{S}(\mathbb{R})$, $j \in \{1,2,3,4\}$, consider the 4-linear form Λ associated with $T^{1,1}$

$$\begin{aligned}\Lambda(f_1,f_2,f_3,f_4) &:= \int T^{1,1}(f_1,f_2,f_3)(x)\overline{f_4}(x)\,dx \\ &= \sum_{k\in\mathbb{Z}} \iiint \widehat{f_1}(\xi_1)\widehat{f_2}(\xi_2)\widehat{f_3}(\xi_3)\widehat{\Phi_1}\left(\frac{\xi_1-\xi_2}{2^k}\right)\widehat{\Phi_2}\left(\frac{\xi_3}{2^k}\right)\overline{\widehat{f_4}}(\xi_1+\xi_2+\xi_3)\,d\xi_1 d\xi_2 d\xi_3,\end{aligned} \tag{5.1}$$

where supp $\widehat{\Phi_1} \subseteq [-1,1]$ and supp $\widehat{\Phi_2} \subseteq [9,10]$.

To simplify the 4-linear form above, we use the wave packet decomposition. Choose a $\psi \in \mathcal{S}(\mathbb{R})$ such that supp$\widehat{\psi} \subseteq [0,1]$ and

$$\sum_{l\in\mathbb{Z}} \widehat{\psi}\left(\xi - \frac{l}{2}\right) = 1 \text{ for any } \xi \in \mathbb{R}.$$

Define

$$\widehat{\psi_{k,l}}(\xi) := \widehat{\psi}\left(\frac{\xi - 2^{k-1}l}{2^k}\right) \text{ for } (k,l) \in \mathbb{Z}^2.$$

Pick a non-negative $\varphi \in \mathcal{S}(\mathbb{R})$ with supp$\widehat{\varphi} \subseteq [-1,1]$ and $\widehat{\varphi}(0) = 1$. Let

$$\varphi_k(x) := 2^k\varphi(2^k x), k \in \mathbb{Z}.$$

For every $(k,n) \in \mathbb{Z}^2$, denote $I_{k,n} := [2^{-k}n, 2^{-k}(n+1))$. Then for each scale $k \in \mathbb{Z}$ and any function $f \in \mathcal{S}(\mathbb{R})$, we have

$$f = \sum_{(n,l)\in\mathbb{Z}^2} f_{k,n,l}, \tag{5.2}$$

where

$$f_{k,n,l}(x) := \chi^*_{I_{k,n}}(x) f * \psi_{k,l}(x), \text{ and} \tag{5.3}$$

$$\chi^*_I(x) := \chi_I * \varphi_k(x) \text{ for any interval } I. \tag{5.4}$$

In sum, $f_{k,n,l}$ is well-localized, as supp $\widehat{f_{k,n,l}} \subseteq [2^k(\frac{l}{2}-1), 2^k(\frac{l}{2}+2)]$ and $f_{k,n,l}$ is essentially supported on $I_{k,n}$ in the sense that

$$|f_{k,n,l}(x)| \lesssim_{N,M} \left(1+\frac{\operatorname{dist}(x,I_{k,n})}{|I_{k,n}|}\right)^{-N} \frac{1}{|I_{k,n}|}\int |f(y)| \left(1+\frac{|x-y|}{|I_{k,n}|}\right)^{-M} dy. \tag{5.5}$$

Now we apply the decomposition (5.2) to all the four functions in (5.1) and obtain

$$\Lambda(f_1,f_2,f_3,f_4) = \sum_{\substack{k\in\mathbb{Z}\\(n_1,n_2,n_3,n_4)\in\mathbb{Z}^4\\(l_1,l_2,l_3,l_4)\in\mathbb{Z}^4}} \iiint \widehat{(f_1)_{k,n_1,l_1}}(\xi_1)\widehat{(f_2)_{k,n_2,l_2}}(\xi_2)\widehat{(f_3)_{k,n_3,l_3}}(\xi_3)$$
$$\widehat{\Phi_1}\left(\frac{\xi_1-\xi_2}{2^k}\right)\widehat{\Phi_2}\left(\frac{\xi_3}{2^k}\right)\overline{\widehat{(f_4)_{k,n_4,l_4}}}(\xi_1+\xi_2+\xi_3)\,d\xi_1 d\xi_2 d\xi_3.$$

By the support of functions, each term in the sum is non-zero only when

$$\begin{cases} \xi_i \in [2^k(\frac{l_i}{2}-1), 2^k(\frac{l_i}{2}+2)] \text{ for } i=1,2,3; \\ |\xi_1-\xi_2| \lesssim 2^k, |\xi_3| \in [9\cdot 2^k, 10\cdot 2^k); \\ \xi_1+\xi_2+\xi_3 \in [2^k(\frac{l_4}{2}-1), 2^k(\frac{l_4}{2}+2)]. \end{cases}$$

These imply that

$$\begin{cases} |l_2 - l_1| \lesssim 1; \\ |l_3 - 9| \lesssim 1; \\ |l_4 - (2l_1 - 18)| \lesssim 1. \end{cases}$$

In other words, among the four parameters l_1, l_2, l_3, l_4 only one is free, say l_1. Without loss of generality we can fix a dependence relation between l_2, l_3, l_4 and l_1. Then drop the cut-off functions by the Fourier expansion trick and ignore the fast decay terms so that $\Lambda(f_1, f_2, f_3, f_4)$ becomes essentially as

$$\sum_{\substack{k,l_1 \\ n_1,n_2,n_3,n_4}} \int (f_1)_{k,n_1,l_1}(x)(f_2)_{k,n_2,l_2}(x)(f_3)_{k,n_3,l_3}(x)\overline{(f_4)_{k,n_4,l_4}}(x)\, dx.$$

Since $(f_j)_{k,n_j,l_j}$ is almost supported in $I_{k,n_j} = [2^{-k}n_j, 2^{-k}(n_j+1))$, there is not too much loss to assume $n_1 = n_2 = n_3 = n_4$ due to the fast decay in other cases. Therefore, the original 4-linear form has been simplified to the following model form (we still use Λ to denote the model 4-linear form by an abuse of notation):

$$\Lambda(f_1, f_2, f_3, f_4) = \sum_{(k,n,l)\in\mathbb{Z}^3} \int \prod_{j=1}^{4} (f_j)_{k,n,l_j}(x)\, dx. \tag{5.6}$$

Here $l_1 = l$, $l_2 = l$, $l_3 = 18$ and $l_4 = 2l + 18$.

We will prove directly that T is of restricted weak type (see [78] for the definition) when (p_1, p_2, p_3) is in a smaller range $D_0 := \{(p_1, p_2, p_3) : 1 < p_1, p_2 < 2, \frac{1}{p_1} + \frac{1}{p_2} < \frac{3}{2}, p_3 \in (1, \infty)\}$. More precisely, we will prove

Theorem 5.4 *Let $(p_1, p_2, p_3) \in D_0$. For any measurable sets F_1, F_2, F_3, F of finite measure, there exists measurable set $F' \subseteq F$ with $|F'| \geq \frac{1}{2}|F|$ such that Λ defined in (5.6) satisfy*

$$|\Lambda(f_1, f_2, f_3, f_4)| \lesssim |F_1|^{\frac{1}{p_1}} |F_2|^{\frac{1}{p_2}} |F_3|^{\frac{1}{p_3}} |F'|^{\frac{1}{p'}}; \tag{5.7}$$

for every $|f_1| \leq \chi_{F_1}$, $|f_2| \leq \chi_{F_2}$, $|f_3| \leq \chi_{F_3}$ and $|f_4| \leq \chi_{F'}$. Here $\frac{1}{p'} := 1 - (\frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{p_3})$.

To prove Theorem 5.4 we pick up an arbitrary finite subset $S \subset \mathbb{Z}^3$ and aim to obtain (5.7) for

$$\Lambda_S(f_1, f_2, f_3, f_4) := \sum_{(k,n,l)\in S} \int \prod_{j=1}^{4} (f_j)_{k,n,l_j}(x)\, dx, \tag{5.8}$$

provided the bound does not depend on the set S. We can also assume $|F| = 1$ by dilation invariance. Next we make the geometric structure of Λ_S clearer. To each tuple $s = (k, n, l) \in \mathbb{Z}^3$ we assign a time-interval

$I_s := I_{k,n}$ and four frequency-intervals ω_{s_j}, $j \in \{1,2,3,4\}$, representing the localization of functions in the time-frequency space. More precisely, I_s and ω_{s_j}'s satisfy:

$$(f_j)_{k,n,l_j}(x) \text{ is dominated by} \tag{5.9}$$
$$C_{N,M}\left(1+\frac{\operatorname{dist}(x,I_s)}{|I_s|}\right)^{-N}\frac{1}{|I_s|}\int |f_j(y)|\left(1+\frac{|x-y|}{|I_s|}\right)^{-M}dy$$
$$\text{The Fourier transform of } (f_j)_{k,n,l_j} \text{is supported on } \omega_{s_j}. \tag{5.10}$$

Definition 5.5 *We call* $s=(k,n,l)$ *a 4-tile (or simply a tile) as it corresponds to 4 single-tiles* $s_j := I_s \times \omega_{s_j}$, $j \in \{1,2,3,4\}$. *Write* $f_{s_j} := f_{k,n,l_j}$ *for simplicity.*

We can take finitely many sparse subsets of S and transform ω_{s_j}'s by fixed affine mappings if needed (since only relative locations of Fourier supports matter) so that I_s and ω_{s_j}'s enjoy nice geometric properties as follows:

$$\omega_{s_1} = \omega_{s_2}; \tag{5.11}$$
$$|\omega_{s_1}| = |\omega_{s_3}| = |\omega_{s_4}| = C|I_s|^{-1}; \tag{5.12}$$
$$\operatorname{dist}(\omega_{s_1},\omega_{s_4}) = |\omega_{s_1}|; \tag{5.13}$$
$$c(\omega_{s_1}) > c(\omega_{s_4}), \text{ where } c(I) \text{ is the center of the interval } I; \tag{5.14}$$
$$\{I_s\}_{s\in S} \text{ is a grid (defined below)}; \tag{5.15}$$
$$\{\omega_{s_1}\cup\omega_{s_4}\}_{s\in S} \text{ is a gird}; \tag{5.16}$$
$$\omega_{s_i} \subsetneq J \text{ for some } i \in \{1,4\}, J := \omega_{s'_1}\cup\omega_{s'_2}\cup\omega_{s'_4}, s' \in S \Rightarrow \tag{5.17}$$
$$\omega_{s_j} \subseteq J \text{ for all } j \in \{1,4\}.$$

Here a grid is defined as a set of intervals having the property that if two different elements intersect then one must contain the other and the larger interval is at least twice as long as the smaller one. See [60] for a detailed construction of the time and frequency intervals.

From now on we fix a finite set of tiles $S \subset \mathbb{Z}^3$ and assume the tiles satisfy (5.9)-(5.17). See Figure 5.3 for a comparison between the tile structure of $T^{1,1}$ and that of BHT.

Theorem 5.4 has been reduced to the following theorem.

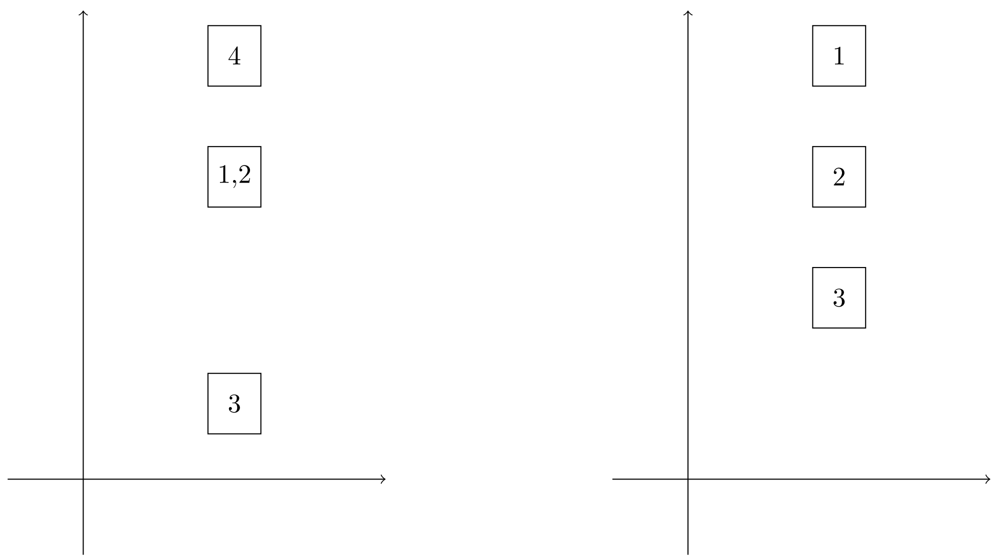

Figure 5.3: 4-tile of $T^{1,1}$ vs. tri-tile of BHT

Theorem 5.6 *Let $p > 1$ be arbitrary. Given any $(p_1, p_2, p_3) \in D_0$ with $p_3 \geq p$ and any sets of finite measure F_1, F_2, F_3, F with $|F| = 1$, there exists $F' \subseteq F$ with $|F'| \geq \frac{1}{2}$ such that*

$$|\Lambda_S(f_1, f_2, f_3, f_4)| \lesssim |F_1|^{\frac{1}{p_1}} |F_2|^{\frac{1}{p_2}} |F_3|^{\frac{1}{p_3}}$$

for every $|f_1| \leq \chi_{F_1}$, $|f_2| \leq \chi_{F_2}$, $|f_3| \leq \chi_{F_3}$ and $|f_4| \leq \chi_{F'}$.

5.3 Proof of Theorem 5.2

In this section we prove Theorem 5.6 and hence Theorem 5.2, using some propositions whose proof will be given in subsequent sections. Fix $p > 1$, $(p_1, p_2, p_3) \in D_0 = \{(p_1, p_2, p_3) : 1 < p_1, p_2 < 2, \frac{1}{p_1} + \frac{1}{p_2} < \frac{3}{2}, p_3 \in (1, \infty)\}$ with $p_3 > p$, and measurable sets F_1, F_2, F_3, F with $|F| = 1$. Let $\mathcal{M}$ denote the maximal operator. Define the exceptional set

$$\Omega := \left(\bigcup_{j=1}^{2} \{x : \mathcal{M}(\chi_{F_j})(x) > C|F_j|\} \right) \bigcup \left\{ x : \mathcal{M}(\chi_{F_3})(x) > C|F_3|^{\frac{1}{p}} \right\}.$$

Then $|\Omega| \leq \frac{1}{4}$ when C is large enough. Set $F' := F \setminus \Omega$ so that $|F'| \geq \frac{1}{2}$. For any dyadic number $\mu \geq 1$, define

$$S^{\mu} := \left\{ s \in S : 1 + \frac{\text{dist}(I_s, \Omega^c)}{|I_s|} \simeq \mu \right\}. \tag{5.1}$$

Then it suffices to obtain the estimate

$$|\Lambda_{S^\mu}(f_1, f_2, f_3, f_4)| \lesssim \mu^{-2}|F_1|^{\frac{1}{p_1}}|F_2|^{\frac{1}{p_2}}|F_3|^{\frac{1}{p_3}} \text{ for any dyadic } \mu \geq 1. \tag{5.2}$$

The main idea to obtain (5.2) is to group the tiles in S^μ appropriately, aiming to establish orthogonality among groups. The following definitions are needed.

Definition 5.7 *Let $j \in \{1, 4\}$. Given two 4-tiles s and s', we write $s_j < s'_j$ if $I_s \subseteq I_{s'}$ and $\omega_{s_j} \supseteq \omega_{s'_j}$. We call $T \subseteq S$ a j-**tree** if there exists a $t \in T$ such that $s_j < t_j$ for all $s \in T$. t is called the **top** of T and denote $I_T := I_t$. We call $T \subseteq S$ a **tree** (with top t) if for any $s \in T$ we have $I_s \subseteq I_t$ and $\omega_{s_j} \supseteq \omega_{t_j}$ for some $j \in \{1, 4\}$.*

It is easy to see that any tree is a union of a 1-tree and a 4-tree.

Definition 5.8 *For any $P \subseteq S$ and $f \in \mathcal{S}(\mathbb{R})$, define*

$$\boldsymbol{size}_j(P, f) := \sup_{\substack{T \subseteq P \\ T \text{ is a 4-tree}}} \left(\frac{1}{|I_T|} \sum_{s \in T} \|f_{s_j}\|_2^2 \right)^{\frac{1}{2}}, j = 1 \text{ or } 2;$$

$$\boldsymbol{size}_4(P, f) := \sup_{\substack{T \subseteq P \\ T \text{ is a 1-tree}}} \left(\frac{1}{|I_T|} \sum_{s \in T} \|f_{s_4}\|_2^2 \right)^{\frac{1}{2}}.$$

Sizes can be controlled using the proposition below, whose proof will be given in Section 5.4.

Proposition 5.9 *Fix a dyadic number $\mu \geq 1$. For any $P \subseteq S^\mu$, $j \in \{1, 2, 4\}$ and $f \in \mathcal{S}(\mathbb{R})$,*

$$size_j(P, f) \lesssim_M \sup_{s \in P} \left(\frac{1}{|I_s|} \|f\|_{L^1(\mu I_s)} + \mu^{-M} \inf_{y \in \mu I_s} Mf(y) \right).$$

If tiles form a tree, then we can control the corresponding 4-form by sizes, as suggested by the following proposition.

Proposition 5.10 *Let $T \subseteq S^\mu$ be a tree. Then*

$$|\Lambda_T(f_1, f_2, f_3, f_4)| \lesssim \mu |I_T| \prod_{j \in \{1,2,4\}} size_j(T, f_j) |F_3|^{\frac{1}{p_3}}.$$

Proof: First assume T is a 1-tree. By Cauchy-Schwartz inequality, we have

$$|\Lambda_T(f_1,f_2,f_3,f_4)| \leq \int \sup_{s\in T}|(f_1)_{s_1}| \sup_{s\in T}|(f_2)_{s_2}| \left(\sum_{s\in T}|(f_3)_{s_3}|^2\right)^{\frac{1}{2}} \left(\sum_{s\in T}|(f_4)_{s_4}|^2\right)^{\frac{1}{2}}$$
$$\leq |I_T| \sup_{s\in T}\|(f_1)_{s_1}\|_\infty \sup_{s\in T}\|(f_2)_{s_2}\|_\infty \left(\frac{1}{|I_T|}\sum_{s\in T}\|(f_3)_{s_3}\|_2^2\right)^{\frac{1}{2}} \left(\frac{1}{|I_T|}\sum_{s\in T}\|(f_4)_{s_4}\|_2^2\right)^{\frac{1}{2}}.$$

Using the structure of the 1-tree and the definition of S^μ,

$$\left(\frac{1}{|I_T|}\sum_{s\in T}\|(f_3)_{s_3}\|_2^2\right)^{\frac{1}{2}} \lesssim \mu \min\{1, |F_3|^{\frac{1}{p}}\} \leq \mu |F_3|^{\frac{1}{p_3}} \tag{5.3}$$

Combine the above two estimates and can bound $|\Lambda_T(f_1,f_2,f_3,f_4)|$ by

$$\mu |I_T| \sup_{s\in T}\|(f_1)_{s_1}\|_\infty \sup_{s\in T}\|(f_2)_{s_2}\|_\infty \ \mathrm{size}_4(T,f_4)|F_3|^{\frac{1}{p_3}}.$$

It remains to prove that for $i=1$ or $i=2$, $\|(f_i)_{s_i}\|_\infty \lesssim \mathrm{size}_i(T,f_i)$ for any $s\in T$. We will only consider $i=1$ case as the other case can be handled similarly. We just need to prove the estimate

$$\|(f_1)_{s_1}\|_\infty \lesssim \|(f_1)_{s_1}\|_2 |I_s|^{-\frac{1}{2}} \tag{5.4}$$

since $\{s\}$ is a 4-tree. To prove (5.4), recall for $s=(k,n,l)$, $(f_1)_{s_1}(x) = \chi^*_{I_{k,n}}(x) f_1 * \psi_{k,l}(x)$, where $\psi_{k,l}(x) = 2^k\psi(2^k x)e^{-2\pi i \frac{l}{2}x}$. Let b be a real number such that $|\frac{l}{2}-b| = 2^k$ and define $\widetilde{(f_1)_{s_1}}(x) := e^{2\pi i b x}(f_1)_{s_1}(x)$. Then $\widetilde{(f_1)_{s_1}}'(x) = \gamma (f_1)_{s_1}(x)$ for some $\gamma \lesssim 2^k$. Hence

$$\|(f_1)_{s_1}\|_\infty = \|\widetilde{(f_1)_{s_1}}\|_\infty \lesssim \sqrt{\|\widetilde{(f_1)_{s_1}}\|_2 \|\widetilde{(f_1)_{s_1}}'\|_2} \lesssim 2^{\frac{k}{2}}\|(f_1)_{s_1}\|_2 \lesssim \|(f_1)_{s_1}\|_2 |I_s|^{-\frac{1}{2}}$$

as desired.

Now assume T is a 4-tree. By similar arguments, we have

$$
\begin{aligned}
|\Lambda_T(f_1,f_2,f_3,f_4)| &\le \int \left(\sum_{s\in T} |(f_1)_{s_1}|^2\right)^{\frac12} \sup_{s\in T} |(f_2)_{s_2}| \left(\sum_{s\in T} |(f_3)_{s_3}|^2\right)^{\frac12} \sup_{s\in T} |(f_4)_{s_4}| \\
&\le |I_T| \left(\frac{1}{|I_T|}\sum_{s\in T} \|(f_1)_{s_1}\|_2^2\right)^{\frac12} \sup_{s\in T} \|(f_2)_{s_2}\|_\infty \left(\frac{1}{|I_T|}\sum_{s\in T} \|(f_3)_{s_3}\|_2^2\right)^{\frac12} \sup_{s\in T} \|(f_4)_{s_4}\|_\infty \\
&\lesssim \mu |I_T| \ \mathrm{size}_1(T,f_1) \sup_{s\in T} \|(f_2)_{s_2}\|_\infty \sup_{s\in T} \|(f_4)_{s_4}\|_\infty |F_3|^{\frac{1}{p_3}} \\
&\lesssim \mu |I_T| \prod_{j\in\{1,2,4\}} \mathrm{size}\ _j(T,f_j) |F_3|^{\frac{1}{p_3}}.
\end{aligned}
$$

This finishes the proof of Proposition 5.10. ■

The following proposition provides the algorithm to select trees and group tiles.

Proposition 5.11 *Let $f \in L^2$. Suppose for some $j \in \{1,2,4\}$ and $P \subseteq S$, we have*

$$
size_j(P,f) \le \sigma \|f\|_2 \text{ for some dyadic number } \sigma = 2^n, n \in \mathbb{Z}.
$$

Then we can decompose $P = P' \cup P''$ such that

$$
size_j(P',f) \le \frac{\sigma}{2}\|f\|_2 \tag{5.5}
$$

and P'' is a union of trees T in some collection $\mathfrak{F}$ with $\sum_{T\in\mathfrak{F}} |I_T| \lesssim \frac{1}{\sigma^2}$.

The proof of this organization proposition will be postponed to Section 5.5.

Now we are ready to prove our goal (5.2). By Proposition 5.9 and the definition of S^μ, we have

$$
\mathrm{size}_j(S^\mu, f_j) \lesssim \begin{cases} \mu |F_j| \text{ when } j = 1,2; \\ \mu^{-M} \text{ for any large } M > 0 \text{ when } j = 4. \end{cases} \tag{5.6}
$$

Iterate the organization algorithm Proposition 5.11 for all $j = 1,2,4$ simultaneously, and we can decompose S^μ as

$$
S^\mu = \bigcup_{\substack{\sigma \text{ is a} \\ \text{dyadic number}}} S_\sigma,
$$

where

$$\text{size}_j(S_\sigma, f_j) \lesssim \begin{cases} \min\{\mu|F_j|, \sigma|F_j|^{\frac{1}{2}}\} \text{ when } j = 1, 2; \\ \min\{\mu^{-M}, \sigma\} \text{ for any large } M > 0 \text{ when } j = 4, \end{cases} \tag{5.7}$$

and $S_\sigma = \cup_{T\in\mathfrak{F}_\sigma} T$ is a union of tees with $\sum_{T\in\mathfrak{F}_\sigma} |I_T| \lesssim \frac{1}{\sigma^2}$.

Using this decomposition and the estimate on a single tree (Proposition 5.10), we see that

$$\begin{aligned} |\Lambda_{S^\mu}(f_1,f_2,f_3,f_4)| &\lesssim \sum_{\sigma \text{ is dyadic}} \sum_{T\in\mathfrak{F}_\sigma} |\Lambda_T(f_1,f_2,f_3,f_4)| \\ &\lesssim \mu \sum_\sigma \sum_{T\in\mathfrak{F}_\sigma} |I_T| \prod_{j\in\{1,2,4\}} \text{size }_j(T,f_j)|F_3|^{\frac{1}{p_3}} \\ &\lesssim \mu^3|F_3|^{\frac{1}{p_3}} \sum_\sigma \frac{1}{\sigma^2} \min\{|F_1|, \sigma|F_1|^{\frac{1}{2}}\} \min\{|F_2|, \sigma|F_2|^{\frac{1}{2}}\} \min\{\mu^{-M}, \sigma\}. \end{aligned}$$

Apply the elementary inequality $\min\{X,Y\} \le X^\theta Y^{1-\theta}$, and we can bound $|\Lambda_{S^\mu}(f_1,f_2,f_3,f_4)|$ by

$$\mu^3|F_3|^{\frac{1}{p_3}} \sum_\sigma \frac{1}{\sigma^2} \sigma^{2\left(1-\frac{1}{p_1}\right)+2\left(1-\frac{1}{p_2}\right)} |F_1|^{\frac{1}{p_1}} |F_2|^{\frac{1}{p_2}} \min\{\mu^{-M}, \sigma\} \lesssim \mu^{-2}|F_1|^{\frac{1}{p_1}}|F_2|^{\frac{1}{p_2}}|F_3|^{\frac{1}{p_3}},$$

where we used the fact $\frac{1}{p_1} + \frac{1}{p_2} < \frac{3}{2}$ in the last inequality. This proves (5.2).

5.4 Size estimates

In this section, we prove Proposition 5.9. The proofs of some variants of this proposition already appear in [26] and [78]. For the convenience of the reader, we include the details here. First we need the following lemma which is another form of the John-Nirenberg inequality.

Lemma 5.12 *For any $P \subseteq S$ and $f \in \mathcal{S}(\mathbb{R})$,*

$$\begin{aligned} size_j(P,f) &\lesssim \sup_{\substack{T\subseteq P \\ T is\ a\ 4\text{-}tree}} \frac{1}{|I_T|} \left\| \left(\sum_{s\in T} \frac{\|f_{s_j}\|_2^2}{|I_s|} \chi_{I_s} \right)^{\frac{1}{2}} \right\|_{1,\infty}, \ j \in \{1,2\}, \\ size_4(P,f) &\lesssim \sup_{\substack{T\subseteq P \\ T is\ a\ 1\text{-}tree}} \frac{1}{|I_T|} \left\| \left(\sum_{s\in T} \frac{\|f_{s_j}\|_2^2}{|I_s|} \chi_{I_s} \right)^{\frac{1}{2}} \right\|_{1,\infty}. \end{aligned}$$

Proof: Fix $j \in \{1,2,4\}$, $P \subseteq S$ and $f \in \mathcal{S}(\mathbb{R})$. Let $T \subseteq P$ be an i-tree for some $i \in \{1,4\}$ with $i \neq j$ such that

$$\text{size}_j(P,f) = \left(\frac{1}{|I_T|}\sum_{s\in T}\|f_{s_j}\|_2^2\right)^{\frac{1}{2}}$$

For simplicity write $a_s := \|f_{s_j}\|_2$ for $s \in T$ and we aim to show

$$\left(\frac{1}{|I_T|}\sum_{s\in T}{a_s}^2\right)^{\frac{1}{2}} \lesssim \frac{1}{|I_T|}\left\|\left(\sum_{s\in T}\frac{{a_s}^2}{|I_s|}\chi_{I_s}\right)^{\frac{1}{2}}\right\|_{1,\infty}. \tag{5.1}$$

Denote the left-hand side (LHS) and the right-hand side (RHS) of (5.1) by A and B, respectively. Let C be a large constant and define the set

$$E := \left\{x : \left(\sum_{s\in T}\frac{{a_s}^2}{|I_s|}\chi_{I_s}(x)\right)^{\frac{1}{2}} > CB\right\} \subseteq I_T. \tag{5.2}$$

By the definition of weak 1 norm,

$$|E| \leq \frac{B|I_T|}{CB} = \frac{|I_T|}{C} \tag{5.3}$$

Write E as a joint union of intervals $E = \bigcup_{I^m\in\mathcal{J}^M} I^m$, where $\mathcal{J}^M$ is the set of maximal elements in

$$\mathcal{J} := \left\{ I = I_{s_0} \text{ for some } s_0 \in T : \left(\sum_{s\in T, I_s\supseteq I}{a_s}^2|I_s|^{-1}\right)^{\frac{1}{2}} > CB\right\}. \tag{5.4}$$

By the definition of A,

$$A^2|I_T| = \sum_{s\in T}{a_s}^2 = \int_E\sum_{s\in T}\frac{{a_s}^2}{|I_s|}\chi_{I_s} + \int_{I_T\setminus E}\sum_{s\in T}\frac{{a_s}^2}{|I_s|}\chi_{I_s} =: H + K. \tag{5.5}$$

Use the decomposition $E = \bigcup_{I^m\in\mathcal{J}^M} I^m$ to split H further as

$$H = \sum_{I^m\in\mathcal{J}^M}\int_{I^m}\sum_{s\in T, I_s\supsetneq I^m}\frac{{a_s}^2}{|I_s|}\chi_{I_s} + \sum_{I^m\in\mathcal{J}^M}\int_{I^m}\sum_{s\in T, I_s\subseteq I^m}\frac{{a_s}^2}{|I_s|}\chi_{I_s} =: H_1 + H_2. \tag{5.6}$$

Since each I^m is maximal in $\mathcal{J}$ defined by (5.4),

$$H_1 \leq \sum_{I^m\in\mathcal{J}^M}(CB)^2|I^m| = (CB)^2|E| \leq (CB)^2|I_T|. \tag{5.7}$$

For each $I^m \in \mathcal{J}^M$, $\{s \in T : I_s \subseteq I^m\}$ is still an i-tree by the grid structure. So the definition of $size_j(P, f)$ and (5.3) give

$$H_2 = \sum_{I^m \in \mathcal{J}^M} |I^m| \left(\frac{1}{|I^m|} \sum_{s \in T, I_s \subseteq I^m} {a_s}^2 \right) \leq \sum_{I^m \in \mathcal{J}^M} |I^m| A^2 = A^2 |E| \leq A^2 \frac{|I_T|}{C} \tag{5.8}$$

Since the integrand in K is dominated by CB by (5.2), we have

$$K \leq (CB)^2 |I_T|. \tag{5.9}$$

Putting (5.5)-(5.9) together, we obtain

$$A^2 |I_T| = H_1 + H_2 + K \leq (CB)^2 |I_T| + A^2 \frac{|I_T|}{C} + (CB)^2 |I_T|, \tag{5.10}$$

from which we obtain $A \lesssim B$. This proves (5.1) and thus Lemma 5.12. ■

We now turn to the proof of Proposition 5.9. Without loss of generality, assume $j = 1$. By Lemma 5.12, it suffices to show for any 4-tree T,

$$\left\| \left(\sum_{s \in T} \frac{\|f_{s_1}\|_2^2}{|I_s|} \chi_{I_s} \right)^{\frac{1}{2}} \right\|_{1,\infty} \lesssim_M \|f\|_{L^1(\mu I_T)} + \mu^{-M} \inf_{y \in \mu I_T} \mathcal{M} f(y) |I_T|. \tag{5.11}$$

Write $f = f\chi_{\mu I_T} + f\chi_{(\mu I_T)^c}$. LHS of (5.11) is bounded by

$$\left\| \left(\sum_{s \in T} \frac{\|(f\chi_{\mu I_T})_{s_1}\|_2^2}{|I_s|} \chi_{I_s} \right)^{\frac{1}{2}} \right\|_{1,\infty} + \left\| \left(\sum_{s \in T} \frac{\|(f\chi_{(\mu I_T)^c})_{s_1}\|_2^2}{|I_s|} \chi_{I_s} \right)^{\frac{1}{2}} \right\|_{1} =: I + II.$$

By the conditions (5.11)-(5.17) of the tiles, in a 4-tree, s_1 tiles are Littlewood-Paley pieces as illustrated in Figure 5.4. Thus term I is bounded by $C\|f\|_{L^1(\mu I_T)}$ since the discrete square-function operator is of weak type $(1, 1)$ by the L^2 estimate and Calderón-Zygmund decomposition.

Using the fact l^2 norm is no more than l^1 norm, we estimate II by

$$\sum_{s \in T} \|(f\chi_{(\mu I_T)^c})_{s_1}\|_2 |I_s|^{\frac{1}{2}}.$$

It remains to show

$$\sum_{s \in T} \|(f\chi_{(\mu I_T)^c})_{s_1}\|_2 |I_s|^{\frac{1}{2}} \lesssim_M \mu^{-M} \inf_{y \in \mu I_T} \mathcal{M} f(y) |I_T|. \tag{5.12}$$

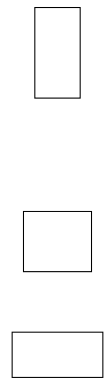

Figure 5.4: s_1 tiles in a 4-tree

Using (5.9) we see that control the function $|(f\chi_{(\mu I_T)^c})_{s_1}(x)|$ is bounded above by

$$\left(1+\frac{\operatorname{dist}(I_s,(\mu I_T)^c)}{|I_s|}\right)^{-N}\left(1+\frac{\operatorname{dist}(x,I_s)}{|I_s|}\right)^{-N}\inf_{y\in\mu I_T}\mathcal{M}f(y).$$

Hence $\sum_{s\in T}\|(f\chi_{(\mu I_T)^c})_{s_1}\|_2|I_s|^{\frac{1}{2}}$ is dominated by

$$\inf_{y\in\mu I_T}\mathcal{M}f(y)\sum_{s\in T}|I_s|\left(1+\frac{\operatorname{dist}(I_s,(\mu I_T)^c)}{|I_s|}\right)^{-N}\lesssim_M \mu^{-M}\inf_{y\in\mu I_T}\mathcal{M}f(y)|I_T|,$$

as desired.

5.5 Organizing tiles

We provide the proof of Proposition 5.11 in this section. Without loss of generality, let $j=1$. By the assumptions of Proposition 5.11,

$$\sup_{\substack{T\subseteq P\\ T\text{ is a 4-tree}}}\left(\frac{1}{|I_T|}\sum_{s\in T}\|f_{s_1}\|_2^2\right)^{\frac{1}{2}}\leq\sigma\|f\|_2. \tag{5.1}$$

Now we begin the tree selection algorithm. Initially set $S_0=P$ and $\mathfrak{F}=\emptyset$. Let

$$\mathfrak{F}_0=\left\{T\subseteq S_0: T\text{ is a 4-tree such that }\left(\frac{1}{|I_T|}\sum_{s\in T}\|f_{s_1}\|_2^2\right)^{\frac{1}{2}}\geq\frac{\sigma}{2}\|f\|_2\right\}. \tag{5.2}$$

If $\mathfrak{F}_0\neq\emptyset$, then take T_1 to be the 4-tree in $\mathfrak{F}_0$ with top t such that $c(\omega_{t_4})\geq c(\omega_{t_4'})$ for any other $T\in\mathfrak{F}_0$ with top t'. Let

$$\begin{cases} T_1^{(4)} := \text{maximal 4-tree in } S_0 \text{ with top } t, \\ T_1^{(1)} := \text{maximal 1-tree in } S_0 \text{ with top } t, \\ T_1^* := T_1^{(1)} \cup T_1^{(4)} \text{ (This is a tree with top } t). \end{cases}$$

Update S_0 and $\mathfrak{F}$ by setting $S_0 := S_0 \setminus T_1^*$ and $\mathfrak{F} := \mathfrak{F} \cup \{T_1^*\}$.

Repeat this algorithm until there is no 4-tree in the updated S_0 satisfying

$$\left(\frac{1}{|I_T|} \sum_{s \in T} \|f_{s_1}\|_2^2 \right)^{\frac{1}{2}} \geq \frac{\sigma}{2} \|f\|_2.$$

When the algorithm terminates, we obtain

$$S_0 = P \setminus \{T_1^*, T_2^*, \ldots, T_l^*\},$$

$$\mathfrak{F} = \{T_1^*, T_2^*, \ldots, T_l^*\}.$$

Simply let $P' = S_0$ and $P'' = \cup_{T \in \mathfrak{F}} T$. Then Clearly $\text{size}_1(P', f) \leq \frac{\sigma}{2} \|f\|_2$.

Now we turn to the proof of $\sum_{T \in \mathfrak{F}} |I_T| \lesssim \frac{1}{\sigma^2}$. We can assume that each $T \in \mathfrak{F}$ is a 4-tree. By the definition of $\mathfrak{F}_0$ (5.2), for any $T \in \mathfrak{F}$,

$$\left(\frac{1}{|I_T|} \sum_{s \in T} \|f_{s_1}\|_2^2 \right)^{\frac{1}{2}} \geq \frac{\sigma}{2} \|f\|_2. \tag{5.3}$$

Therefore,

$$\sum_{T \in \mathfrak{F}} |I_T| \lesssim \frac{1}{\sigma^2 \|f\|_2^2} \sum_{T \in \mathfrak{F}} \sum_{s \in T} \|f_{s_1}\|_2^2.$$

It will suffice to prove

$$\sum_{T \in \mathfrak{F}} \sum_{s \in T} \|f_{s_1}\|_2^2 \lesssim \|f\|_2^2. \tag{5.4}$$

For each 4-tile s, define an operator A_s by $A_s f(x) = f_{s_1}(x)$. By Cauchy-Schwartz inequality,

$$\sum_{T \in \mathfrak{F}} \sum_{s \in T} \|f_{s_j}\|_2^2 = \left\langle \sum_{T \in \mathfrak{F}} \sum_{s \in T} A_s^* A_s f, f \right\rangle \leq \left\| \sum_{T \in \mathfrak{F}} \sum_{s \in T} A_s^* A_s f \right\|_2 \|f\|_2.$$

Hence (5.4) follows from the following estimate:

$$\left\|\sum_{T\in\mathfrak{F}}\sum_{s\in T}A_s^*A_sf\right\|_2\lesssim\left(\sum_{T\in\mathfrak{F}}\sum_{s\in T}\|f_{s_j}\|_2^2\right)^{\frac{1}{2}}. \tag{5.5}$$

To prove (5.5), write

$$(\text{LHS of }(5.5))^2=\sum_{T,T'\in\mathfrak{F}}\sum_{\substack{s\in T\\ s'\in T'}}\langle A_s^*A_sf,A_{s'}^*A_{s'}f\rangle=I+II,$$

where

$$\begin{cases} I:=\sum_{T\neq T'\in\mathfrak{F}}\sum_{\substack{s\in T\\ s'\in T'}}\langle A_s^*A_sf,A_{s'}^*A_{s'}f\rangle, \\ II:=\sum_{T\in\mathfrak{F}}\sum_{s,s'\in T}\langle A_s^*A_sf,A_{s'}^*A_{s'}f\rangle. \end{cases}$$

Therefore, (5.5) follows from the estimate

$$\max\{I,II\}\lesssim\sum_{T\in\mathfrak{F}}\sum_{s\in T}\|f_{s_1}\|_2^2. \tag{5.6}$$

We will only provide the estimate for I, as II is easier to control so we omit the proof. Apply Cauchy-Schwartz inequality,

$$I\leq\sum_{T\neq T'\in\mathfrak{F}}\sum_{\substack{s\in T\\ s'\in T'}}\|A_sf\|_2\|A_sA_{s'}^*\|\|A_{s'}f\|_2.$$

Hence (5.6) is a consequence of the inequality below.

$$\sum_{T\neq T'\in\mathfrak{F}}\sum_{\substack{s\in T\\ s'\in T'}}\|A_sf\|_2\|A_sA_{s'}^*\|\|A_{s'}f\|_2\lesssim\sum_{T\in\mathfrak{F}}\sum_{s\in T}\|f_{s_1}\|_2^2. \tag{5.7}$$

The following estimate for $\|A_sA_{s'}^*\|$ is the key to sum up all the terms in the LHS of (5.7).

Claim 5.13 $\|A_sA_{s'}^*\|\neq 0$ *only when* $\omega_{s_1}\cap\omega_{s'_1}\neq\emptyset$. *Moreover,*

$$\|A_sA_{s'}^*\|\lesssim_N\frac{|I_{s'}|^{\frac{1}{2}}}{|I_s|^{\frac{1}{2}}}\left(1+\frac{dist(I_s,I_{s'})}{|I_s|}\right)^{-N}\qquad if\quad \omega_{s_1}\subseteq\omega_{s'_1}. \tag{5.8}$$

Proof: Write $A_sA_{s'}^*f(x)=\int K(x,y)f(y)\,dy$, where $K(x,y)=\chi_{I_s}^*(x)\chi_{I_{s'}}^*(y)\widetilde{\psi_{s'_j}}*\psi_{s_j}(x-y)$, $\psi_{s_j}:=\psi_{k,l_j}$ for $s=(k,n,l)$ and $\widetilde{g}(x):=\overline{g(-x)}$ for any function g. Note that $\widetilde{\psi_{s'_j}}*\psi_{s_j}(t)=\int\overline{\widehat{\psi_{s'}}}(\xi)\widehat{\psi_s}(\xi)e^{2\pi i\xi t}\,d\xi$ is non-zero only when $\omega_{s_j}\cap\omega_{s'_j}\neq\emptyset$ by (5.3) and (5.10). Assume $\omega_{s_1}\subseteq\omega_{s'_1}$. By the definitions of χ_I^* (5.4) and $\psi_{k,l}$ and using the triangle inequality $(1+|a|)^{-1}+(1+|b|)^{-1}\leq(1+|a+b|)^{-1}$,

$$|K(x,y)| \lesssim_N \left(1+\frac{\operatorname{dist}(x,I_s)}{|I_s|}\right)^{-2N}\left(1+\frac{\operatorname{dist}(y,I_{s'})}{|I_{s'}|}\right)^{-N}$$
$$\frac{1}{|I_s||I_{s'}|}\int\left(1+\frac{|x-y-z|}{|I_{s'}|}\right)^{-2N}\left(1+\frac{|z|}{|I_s|}\right)^{-N}dz$$
$$\lesssim_N \left(1+\frac{\operatorname{dist}(I_s,I_{s'})}{|I_s|}\right)^{-N}\frac{1}{|I_s|}\left(1+\frac{\operatorname{dist}(x,I_s)}{|I_s|}\right)^{-N}.$$

Hence

$$\int |K(x,y)|dx \lesssim_N \left(1+\frac{\operatorname{dist}(I_s,I_{s'})}{|I_s|}\right)^{-N}. \tag{5.9}$$

Similarly,

$$\int |K(x,y)|dy \lesssim_N \left(1+\frac{\operatorname{dist}(I_s,I_{s'})}{|I_s|}\right)^{-N}\frac{|I_{s'}|}{|I_s|}. \tag{5.10}$$

(5.9) and (5.10) imply (5.8) by Schur's lemma. ■

By the claim and symmetry, in the proof of (5.7) we will assume without loss of generality $\omega_{s_1} \subseteq \omega_{s'_1}$. We will also assume that $\omega_{s_1} \subsetneq \omega_{s'_1}$, as the case $\omega_{s_1} = \omega_{s'_1}$ can be handled the same way. Under these assumptions, (5.7) has been reduced to

$$\sum_{T\neq T'\in\mathfrak{F}}\sum_{\substack{s\in T, s'\in T'\\ \omega_{s_1}\subsetneq\omega_{s'_1}}}\|A_s f\|_2\|A_sA_{s'}^*\|\|A_{s'}f\|_2 \lesssim \sum_{T\in\mathfrak{F}}\sum_{s\in T}\|f_{s_1}\|_2^2. \tag{5.11}$$

Since $\{s\}$ is a 4-tree and $\text{size}_1(f,P)\le\sigma\|f\|_2$,

$$\|A_s f\|_2 \le |I_s|^{\frac{1}{2}}\sigma\|f\|_2. \tag{5.12}$$

Also notice that by (5.3)

$$\sigma\|f\|_2 \lesssim \left(|I_T|^{-1}\sum_{s_0\in T}\|f_{(s_0)_1}\|_2^2\right)^{\frac{1}{2}}. \tag{5.13}$$

Combine (5.12) and (5.13), and we see that

$$\|A_s f\|_2 \lesssim |I_s|^{\frac{1}{2}}|I_T|^{-\frac{1}{2}}\left(\sum_{s_0\in T}\|f_{(s_0)_1}\|_2^2\right)^{\frac{1}{2}}. \tag{5.14}$$

Similarly,

$$\|A_{s'}f\|_2 \lesssim |I_{s'}|^{\frac{1}{2}}|I_T|^{-\frac{1}{2}}\left(\sum_{s_0\in T}\|f_{(s_0)_1}\|_2^2\right)^{\frac{1}{2}}. \tag{5.15}$$

Using (5.14) and (5.15), LHS of (5.11) is bounded by

$$\sum_{T\in\mathbf{T}}\left(\sum_{s_0\in T}\|f_{(s_0)_1}\|_2^2\right)\left(\sum_{\substack{s\in T,T'\neq T\\ s'\in T',\omega_{s_1}\subsetneq\omega_{s'_1}}}|I_s|^{\frac{1}{2}}|I_{s'}|^{\frac{1}{2}}|I_T|^{-1}\|A_sA_{s'}^*\|\right).$$

Therefore, (5.11) will be established once we show that for any $T\in\mathfrak{F}$,

$$\sum_{\substack{s\in T,T'\neq T\\ s'\in T',\omega_{s_1}\subsetneq\omega_{s'_1}}}|I_s|^{\frac{1}{2}}|I_{s'}|^{\frac{1}{2}}|I_T|^{-1}\|A_sA_{s'}^*\|\lesssim 1.$$

By (5.8), this can be reduced to the estimate that for any $T\in\mathfrak{F}$,

$$\sum_{\substack{s\in T,T'\neq T\\ s'\in T',\omega_{s_1}\subsetneq\omega_{s'_1}}}\left(1+\frac{\operatorname{dist}(I_s,I_{s'})}{|I_s|}\right)^{-N}|I_{s'}|\lesssim|I_T|. \tag{5.16}$$

To prove (5.16), we need a crucial observation.

Claim 5.14 *If $T_1\neq T_2\in\mathfrak{F}$, $s\in T_1$, and $s'\in T_2$, then*

$$\omega_{s_1}\subseteq\omega_{s'_1}\Rightarrow I_{s'}\cap I_{T_1}=\emptyset.$$

Proof: Let t and t' denote the top of T_1 and T_2 respectively. Assume otherwise $I_{s'}\cap I_{T_1}\neq\emptyset$. Then $I_{s'}\subseteq I_t$. By (5.17) and the definition of tree, $\omega_{s'_1}\supseteq\omega_{s_4}\supseteq\omega_{t_4}$. Then T_1 is selected before T_2 as $c(\omega_{t_4})>c(\omega_{t'_4})$. However, $s'_1<t_1$ indicates that s' should be selected together with T_1 according to the algorithm (See Figure 5.5). This contradicts with the assumption that $s'\in T_2$. ■

Now we are ready to prove (5.16). It is easy to see that

$$\text{LHS of (5.16)}\lesssim\sum_{s\in T}\sum_{\substack{T'\neq T\\ s'\in T',\omega_{s_1}\subsetneq\omega_{s'_1}}}\int_{I_{s'}}\left(1+\frac{\operatorname{dist}(I_s,x)}{|I_s|}\right)^{-N}dx.$$

By Claim 5.14, $I_{s'}$'s are pairwise disjoint and the union of these intervals is contained in $(I_T)^c$. Therefore,

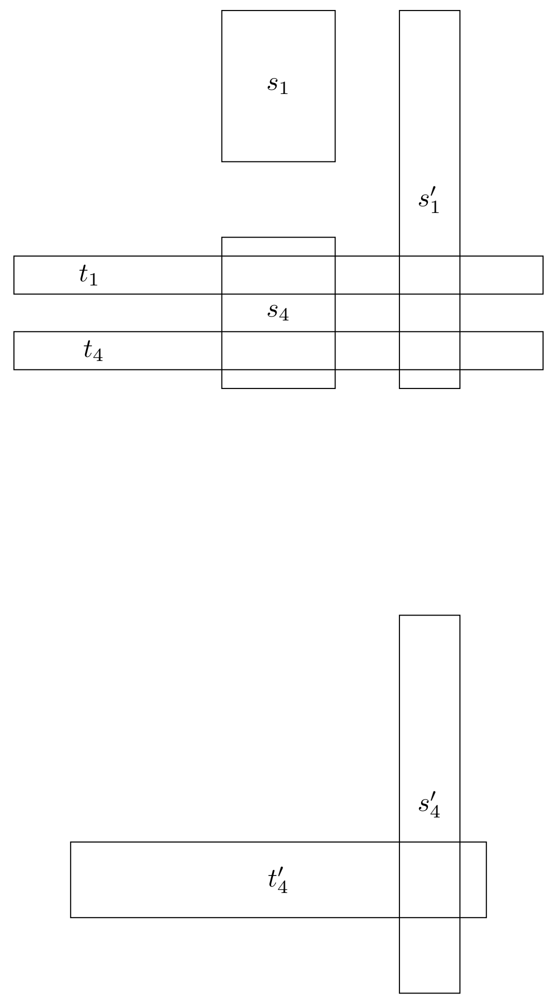

Figure 5.5: a crucial geometric observation

$$\sum_{s\in T} \sum_{\substack{T'\neq T \\ s'\in T', \omega_{s_1} \subsetneq \omega_{s'_1}}} \int_{I_{s'}} \left(1+\frac{\text{dist}(I_s,x)}{|I_s|}\right)^{-N} dx$$

$$\leq \sum_{s\in T} \int_{(I_T)^c} \left(1+\frac{\text{dist}(I_s,x)}{|I_s|}\right)^{-N} dx \lesssim \sum_{s\in T} \left(1+\frac{\text{dist}(I_s,(I_T)^c)}{|I_s|}\right)^{-N} |I_s|$$

Using the tree structure of T and the grid structure of tiles, it is easy to see that

$$\sum_{s\in T} \left(1+\frac{\text{dist}(I_s,(I_T)^c)}{|I_s|}\right)^{-N} |I_s| \lesssim |I_T|.$$

This proves (5.16).

5.6 Telescoping

We prove Theorem 5.3 by a telescoping argument. In what follows, $[x]$ will be used to denote the integer part of $x \in \mathbb{R}$.

Since $k \geq N \geq 10\alpha/\beta$, $[\frac{\beta}{\alpha}k]$ is large and essentially we have

$$T_N^{\alpha,\beta}(f_1, f_2, f_3)(x) = \sum_{k \geq N} H^{\alpha,k}(f_1, f_2)(x) f_3^{\beta,k}(x) =: A + B,$$

where

$$\begin{cases} A := \sum_{k\geq N} \left(\sum_{j=0}^{[(1-\frac{\beta}{\alpha})k]-1} \left(H^{\alpha,k-j}(f_1,f_2)(x) - H^{\alpha,k-j-1}(f_1,f_2)(x) \right) \right) f_3^{\beta,k}(x), \\ B := \sum_{k\geq N} H^{\alpha,[\frac{\beta}{\alpha}k]}(f_1,f_2)(x) f_3^{\beta,k}(x) = \sum_{k\geq N} H^{\beta,k}(f_1,f_2)(x) f_3^{\beta,k}(x). \end{cases}$$

B has a much better tile structure than $T_N^{\alpha,\beta}$: See Figure 5.2 and Figure 5.6 for a comparison. Since B is a part of $T^{\beta,\beta}$ and the proof of Theorem 5.2 is valid for any collection of scales k, boundedness of B is obtained.

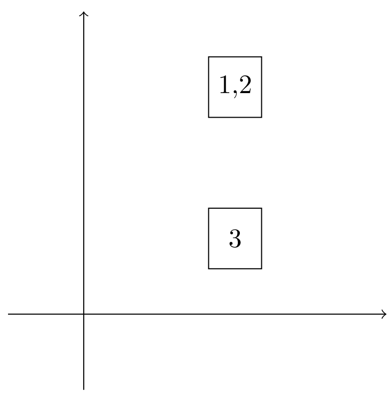

Figure 5.6: tri-tile structure of $T^{\beta,\beta}$

It remains to analyze the operator A. By a change of variable $k \to k + j$, we can write $A = I + II$, where

$$\begin{cases} I := \sum_{k\geq N} \left(H^{\alpha,k}(f_1,f_2)(x) - H^{\alpha,k-1}(f_1,f_2)(x) \right) \left(\sum_{j=0}^{[(\frac{\alpha}{\beta}-1)k]} f_3^{\beta,k+j}(x) \right), \\ II := \sum_{(k,j)\in P} \left(H^{\alpha,k}(f_1,f_2)(x) - H^{\alpha,k-1}(f_1,f_2)(x) \right) f_3^{\beta,k+j}(x). \end{cases}$$

Here P is a finite set of indices, and II should be considered as an error term, whose boundedness follows from Hölder and Lacey-Thiele's Theorem ([60, 61]). To prove the boundedness of the main term I, first note that

$$\sum_{j=0}^{[(\frac{\alpha}{\beta}-1)k]} f_3^{\beta,k+j}(x) = f_3^{\alpha k}(x) - f_3^{\beta k}(x),$$

where

$$f^l(x) := \int \hat{f}(\xi)\phi_0\left(\frac{\xi}{2^l}\right) d\xi,\ l \in \mathbb{R},$$

for some bump function ϕ_0 supported in $[-1, 1]$. Hence we can write I as the difference of two parts:

$$I = \sum_{k\geq 1}\left(H^{\alpha,k}(f_1, f_2)(x) - H^{\alpha,k-1}(f_1, f_2)(x)\right) f_3^{\alpha k}(x) -$$
$$\sum_{k\geq 1}\left(H^{\alpha,k}(f_1, f_2)(x) - H^{\alpha,k-1}(f_1, f_2)(x)\right) f_3^{\beta,k}(x).$$

Note that $H^{\alpha,k}(f_1, f_2)(x) - H^{\alpha,k-1}(f_1, f_2)(x)$ is a piece of BHT at scale k. Since $\alpha > \beta$ and $k > 0$, the supports of $\widehat{f_3^{\alpha k}}$ and $\widehat{f_3^{\beta k}}$ are at most as large as $2^{\alpha k}$. We can introduce a fourth function and do the wave packet decomposition to f_1, f_2, f_4. Then the tiles associated with these functions have structures similar to that of the tri-tiles as in the study of BHT. Therefor, the proof of Theorem 5.1 given in [26] still applies to I, and we omit the details. This finishes the proof of Theorem 5.3.

References

[1] R. C. Baker, G. Harman, J. Pintz, *The difference between consecutive primes, II,* Proc. London Math. Soc. (**3**) **83** (2001) 532-562.

[2] A. Balog, J. Pelikán, J. Pintz, E. Szemerédi *Difference sets without th powers*, Acta Math. Hungar. 65 (1994), no. 2, 165-187.

[3] V. Bergelson, *Ergodic Ramsey theoryan update*, Ergodic theory of $\mathbb{Z}^d$ actions (Warwick, 19931994), 1-61, London Math. Soc. Lecture Note Ser., 228, Cambridge Univ. Press, Cambridge, 1996.

[4] V. Bergelson, A. Leibman, *Polynomial extensions of van der Waerden's and Szemerédi's theorems*, J. Amer. Math. Soc. 9 (1996), no. 3, 725-753.

[5] D. Bilyk, L. Grafakos, *Distributional estimates for the bilinear Hilbert transform*, J. Geom. Anal. 16 (2006), no. 4, 563-584.

[6] D. Bilyk, L. Grafakos *A new way of looking at distributional estimates; applications for the bilinear Hilbert transform*, Collect. Math. 2006, Vol. Extra, 141-169.

[7] E. Bombieri, *On exponential sums in finite fields*, Amer. J. Math. 88 1966 71-105

[8] D. Bose, S. Madan, P. Mohanty, S. Shrivastava, *Relations between bilinear multipliers on $\mathbb{R}^n$, $\mathbb{T}^n$ and* $\mathbb{Z}^n$, Proc. Indian Acad. Sci. Math. Sci. **119** (2009), no. **4**, 501-512.

[9] J. Bourgain, *A nonlinear version of Roth's theorem for sets of positive density in the real line*, J. Analyse Math. 50 (1988), 169-181.

[10] J. Bourgain *Pointwise ergodic theorems for arithmetic sets*, with an appendix by the author, Harry Furstenberg, Yitzhak Katznelson and Donald S. Ornstein. Inst. Hautes tudes Sci. Publ. Math. No. 69 (1989), 5-45.

[11] J. Bourgain, M.C. Chang, *Nonlinear Roth type theorems in finite fields*, Israel J. Math. 221 (2017), no. 2, 853-867.

[12] L. Carlitz, S. Uchiyama, *Bounds for exponential sums*, Duke Math. J. 24 (1957), 37-41.

[13] R. D. Carmichael, *On Euler's ϕ-function*, Bull. Amer. Math. Soc. **13** (1907), 241-243.

[14] R. D. Carmichael, *Note on Euler's ϕ-function*, Bull. Amer. Math. Soc. **28** (1922), 109-110.

[15] H. Cramér, *On the order of magnitude of the difference between consecutive prime numbers*, Acta Arith, **2**: 23-46 (1936).

[16] M. Christ, X. Li, T. Tao, C. Thiele, *On multilinear oscillatory integrals, nonsingular and singular*, Duke Math. J. 130 (2005), no. 2, 321-351.

[17] M. Christ, A. Nagel, E.M. Stein, S. Wainger, *Singular and maximal Radon transforms: analysis and geometry*, Ann. of Math. (2) 150 (1999), no. 2, 489-577.

[18] R. Coifman, Y. Meyer *Ondelettes et oprateurs. III.*, Hermann, Paris, 1991.

[19] P. Deligne, *La conjecture de Weil. I.*, Inst. Hautes Études Sci. Publ. Math. No. 43 (1974), 273-307.

[20] P. Deligne *Cohomologie étale* ($SGA4\frac{1}{2}$), Lecture Notes in Math. 569, Springer, Berlin (1977)

[21] F. Di Plinio, C. Thiele, *Endpoint bounds for the bilinear Hilbert transform*, Trans. Amer. Math. Soc. 368 (2016), no. 6, 3931-3972.

[22] D. Dong, On a discrete bilinear singular operator, *C. R. Math. Acad. Sci. Paris*, 355 (2017), no. 5, 538-542

[23] D. Dong, *On bilinear Hilbert transform along two polynomials*, https://arxiv.org/abs/1708.01326, submitted.

[24] D. Dong, *Full range boundedness of bilinear Hilbert transform along certain nonlinear polynomials.*

[25] D. Dong, *Quasi pieces of bilinear Hilbert transform incorporated into a paraproduct*, J Geom Anal (2018). https://doi.org/10.1007/s12220-018-9987-4

[26] D. Dong, X. Li, *On a hybrid of bilinear Hilbert transform and paraproduct*, Acta Math. Sin. (Engl. Ser.) 34 (2018), no. 1, 29-41

[27] D. Dong, X. Li, W. Sawin, *Improved estimates for polynomial Roth type theorems in finite fields*, https://arxiv.org/abs/1709.00080

[28] D. Dong, X. Meng, *Discrete bilinear Radon transforms along arithmetic functions with many common values*, Bull. Lond. Math. Soc., https://doi.org/10.1112/blms.12127.

[29] P. Durcik, S. Guo, J. Roos *A polynomial Roth theorem on the real line*, https://arxiv.org/abs/1704.01546

[30] J. Ellenberg, D. Gijswijt *On large subsets of $\mathbb{F}_q^n$ with no three-term arithmetic progression*, Ann. of Math. (2) 185 (2017), no. 1, 339-343.

[31] D. Fan, X. Li *A bilinear oscillatory integral along parabolas*, Positivity 13 (2009), no. 2, 339-366.

[32] E. Fabes, *Singular integrals and partial differential equations of parabolic type*, Studia Math. 28 1966/1967 81-131.

[33] C. Fefferman, *Pointwise convergence of Fourier series*, Ann. of Math. (2) 98 (1973), 551-571.

[34] K. Ford, *The number of solutions of $\phi(x) = m$*, Ann. of Math.(**2**) **150** (1999), no. **1**, 283-311.

[35] K. Ford, B. Green, S. Konyagin, J. Maynard, and T. Tao, *Long gaps between primes*, J. Amer. Math. Soc. 31 (2018), no. 1, 65-105.

[36] N. Frantzikinakis, B. kra, *Polynomial averages converge to the product of integrals*, Israel J. Math. 148 (2005), 267-276.

[37] É. Fouvry, E. Kowalski, P. Michel, *A study in sums of products* Philos. Trans. Roy. Soc. A 373 (2015), no. 2040, 20140309, 26 pp.

[38] H. Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Analyse Math. 31 (1977), 204-256.

[39] H. Furstenberg, *Nonconventional ergodic averages. The legacy of John von Neumann (Hempstead, NY, 1988), 43-56*, Proc. Sympos. Pure Math., 50, Amer. Math. Soc., Providence, RI, 1990.

[40] W.T. Gowers, *A new proof of Szemerédi's theorem for arithmetic progressions of length four*, Geom. Funct. Anal. 8 (1998), no. 3, 529-551.

[41] W.T. Gowers, *A new proof of Szemerédi's theorem*, Geom. Funct. Anal. 11 (2001), no. 3, 465-588

[42] L. Grafakos, X. Li, *Uniform bounds for the bilinear Hilbert transforms. I*, Ann. of Math. (2) 159 (2004), no. 3, 889-933.

[43] L. Grafakos and X. Li, *The disc as a bilinear multiplier*, Amer. J. Math. 128 (2006), no. 1, 91-119.

[44] B. Green, T. Tao, *The primes contain arbitrarily long arithmetic progressions*, Ann. of Math. (2) 167 (2008), no. 2, 481-547.

[45] J. Guo, L. Xiao, *Bilinear Hilbert transforms associated with plane curves*, J. Geom. Anal. 26 (2016), no. 2, 967-995.

[46] G. H. Hardy and E. M. Wright. *An introduction to the theory of numbers.* Fourth Ed., Oxford University Press, London, 1960.

[47] M. Hamel, N. Lyall, A. Rice *Improved bounds on Sárközy's theorem for quadratic polynomials* Int. Math. Res. Not. IMRN 2013, no. 8, 1761-1782.

[48] D. R. Heath-Brown and H. Iwaniec, *On the difference between consecutive primes*, Invent. Math. **55** (1979) 49-69.

[49] G. Hoheisel, *Primzahlprobleme in der Analysis*, Sitzunsberichte der Kniglich Preussischen Akademie der Wissenschaften zu Berlin. **33**: 3-11, 1930.

[50] L. Hörmander, *Oscillatory integrals and multipliers on* FL^p Ark. Mat. 11(1973), 1-11.

[51] B. Host, B. Kra, *Nonconventional ergodic averages and nilmanifolds*, Ann. of Math. (2) 161 (2005), no. 1, 397-488.

[52] Y. Hu, X. Li, *Discrete Fourier restriction associated with Schrödinger equations*, Rev. Mat. Iberoam. 30 (2014), no. 4, 1281-1300.

[53] A. Ionescu, S. Wainger, L^p *boundedness of discrete singular Radon transforms*, J. Amer. Math. Soc. 19 (2006), no. 2, 357-383.

[54] A. Ionescu, A. Magyar, E.M. Stein, S. Wainger, *Discrete Radon transforms and applications to ergodic theory*, Acta Math. 198 (2007), no. 2, 231-298.

[55] H. Iwaniec and M. Jutila, *Primes in short intervals*, Ark. Mat. **17** (1979) 167-176.

[56] B. Krause, *Polynomial Ergodic Averages Converge Rapidly: Variations on a Theorem of Bourgain*, https://arxiv.org/abs/1402.1803

[57] N. Katz, *Sommes exponentielles* Astérisque, 79. Société Mathématique de France, Paris, 1980. 209 pp.

[58] N. Katz, *Estimates for "singular" exponential sums.* Internat. Math. Res. Notices 1999, no. 16, 875-899.

[59] E. Kowalski, P. Michel, W. Sawin, *Bilinear forms with Kloosterman sums and applications* Ann. of Math. (2) 186 (2017), no. 2, 413-500.

[60] M. Lacey, C. Thiele, L^p *estimates on the bilinear Hilbert transform for* $2 < p < \infty$, Ann. Math. **146** (1997), 693-724.

[61] M. Lacey, C. Thiele, *On Calderón's conjecture*, Ann. of Math. (2) 149 (1999), no. 2, 475-496.

[62] X. Li, *Uniform bounds for the bilinear Hilbert transforms. II*, Rev. Mat. Iberoam. 22 (2006), no. 3, 1069-1126.

[63] X. Li, *Uniform estimates for some paraproducts*, New York J. Math. 14 (2008), 145-192.

[64] X. Li, *Bilinear Hilbert transforms along curves I: The monomial case*, Anal. PDE 6 (2013), no. 1, 197-220.

[65] X. Li, L. Xiao, *Uniform estimates for bilinear Hilbert transform and bilinear maximal functions associated to polynomials*, Amer. J. Math., 138 (2016), No. 4, 907-962.

[66] V. Lie, *On the boundedness of the bilinear Hilbert transform along "non-flat" smooth curves*, Amer. J. Math. 137 (2015), no. 2, 313-363.

[67] V. Lie, *On the Boundedness of The Bilinear Hilbert Transform along "non-flat" smooth curves. The Banach triangle case* ($L^r, 1 \leq r < \infty$), Rev. Mat. Iberoam., 34 (2018), no. 1, 331-353.

[68] J. Lucier, *Intersective sets given by a polynomial*, Acta Arith. 123 (2006), no. 1, 57-95.

[69] M. Lyall, Á. Magyar *Polynomial configurations in difference sets*, J. Number Theory 129 (2009), no. 2, 439-450.

[70] M. Lyall, Á. Magyar *Simultaneous polynomial recurrence*, Bull. Lond. Math. Soc. 43 (2011), no. 4, 765-785.

[71] J. Maynard, *Small gaps between primes*, Ann. of Math. (**2**) **181** (2015), no. **1**, 383-413.

[72] M. Mirek, *Square function estimates for discrete Radon transforms*, Anal. PDE 11 (2018), no. 3, 583-608

[73] M. Mirek, E.M. Stein, B. Trojan, $l^p(\mathbb{Z}^d)$*-estimates for discrete operators of Radon type: Maximal functions and vector-valued estimates*, https://arxiv.org/abs/1512.07518

[74] M. Mirek, E.M. Stein, B. Trojan, $l^p(\mathbb{Z}^d)$*-estimates for discrete operators of Radon type: Variational estimates*, Invent. Math. 209 (2017), no. 3, 665-748

[75] M. Mirek, B. Trojan, *Cotlar's ergodic theorem along the prime numbers*, J. Fourier Anal. Appl. 21 (2015), no. 4, 822-848.

[76] M. Mirek, B. Trojan, *Discrete maximal functions in higher dimensions and applications to ergodic theory*, Amer. J. Math., 138 (2016), No. 6, 1495–1532.

[77] M. Mirek, B. Trojan, P. Zorin-Kranich, *Variational estimates for averages and truncated singular integrals along the prime numbers*, Trans. Amer. Math. Soc. 369 (2017), no. 8, 5403-5423.

[78] C. Muscalu, W. Schlag, *Classical and multilinear harmonic analysis. Vol. II*, Cambridge Studies in Advanced Mathematics, 138. Cambridge University Press, Cambridge, 2013.

[79] C. Muscalu, T. Tao, C. Thiele, *Multi-linear operators given by singular multipliers*, J. Amer. Math. Soc. 15 (2002), no. 2, 469-496.

[80] A. Nagel, N. Riviere, S. Wainger, *On Hilbert transforms along curves*, Bull. Amer. Math. Soc. 80 (1974), 106-108.

[81] A. Nagel, N. Riviere, S. Wainger, *On Hilbert transforms along curves. II*, Amer. J. Math. 98 (1976), no. 2, 395-403.

[82] A. Nagel, N. Riviere, S. Wainger, *A maximal function associated to the curve* (t, t^2), Proc. Nat. Acad. Sci. U.S.A. 73 (1976), no. 5, 1416-1417.

[83] S. Peluse, *Three-term polynomial progressions in subsets of finite fields*, https://arxiv.org/abs/1707.05977

[84] S. Peluse, *On the polynomial Szemerédi's theorem in finite fields*, https://arxiv.org/abs/1802.02200

[85] L. Pierce, *Discrete fractional Radon transforms and quadratic forms*, Duke Math. J. **161** (2012), no. **1**, 69-106.

[86] J. Pintz, W. Steiger, E. Szemerédi *On sets of natural numbers whose difference set contains no squares*, J. London Math. Soc. (2) 37 (1988), no. 2, 219-231.

[87] C. Pomerance, *Popular values of Euler's function*, Mathematika, **27** (1980), no. **1**, 84-89.

[88] M. Riesz, *Sur les fonctions conjuguées*, Math. Z. (1928), 27 (1) 218-244.

[89] K.F. Roth, *On certain sets of integers*, J. London Math. Soc. 28, (1953). 104-109

[90] A. Sárközy, *On difference sets of sequences of integers. I.*, Acta Math. Acad. Sci. Hungar. 31 (1978), no. 1-2, 125-149.

[91] W. Schramm, *The Fourier transform of functions of the greatest common divisor*, Integers, A**50 8** (2008).

[92] S. Slijepčević, *A polynomial Sárközy-Furstenberg theorem with upper bounds*, Acta Math. Hungar. 98 (2003), no. 1-2, 111-128.

[93] E.M. Stein, *Some problems in harmonic analysis*, Proc. Internat. Congress Math (Nice, 1970), vol.1, Gauthier-Villars, Paris, 1971, pp.173-190.

[94] E.M. Stein, *Maximal functions. II. Homogeneous curves*, Proc. Nat. Acad. Sci. U.S.A. 73 (1976), no. 7, 2176-2177.

[95] E.M. Stein, S. Wainger, *The estimation of an integral arising in multiplier transformations*, Studia Math. 35 1970 101–104.

[96] E.M. Stein, S. Wainger, *Maximal functions associated to smooth curves*, Proc. Nat. Acad. Sci. U.S.A. 73 (1976), no. 12, 4295-4296.

[97] E.M. Stein, S. Wainger, *Problems in harmonic analysis related to curvature*, Bull. Amer. Math. Soc. 84 (1978), no. 6, 1239-1295.

[98] E.M. Stein, S. Wainger, *Discrete analogues of singular Radon transforms*, Bull. Amer. Math. Soc. (N.S.) 23 (1990), no. 2, 537-544.

[99] E. Szemerédi, *On sets of integers containing no four elements in arithmetic progression*, Acta Math. Acad. Sci. Hungar. 20 1969 89-104.

[100] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. 27 (1975), 199-245.

[101] T. Tao *Expanding polynomials over finite fields of large characteristic, and a regularity lemma for definable sets*, Contr. to Disc. Math. 10 (2014), no 1. 22-98.

[102] T. Tao, *Cancellation for the multilinear Hilbert transform*, Collect. Math. 67 (2016), no. 2, 191-206.

[103] T. Tao, T. Ziegler, *The primes contain arbitrarily long polynomial progressions*, Acta Math. 201 (2008), no. 2, 213-305.

[104] C. Thiele, *A uniform estimate*, Ann. of Math. (2) 156 (2002), no. 2, 519-563

[105] C. Thiele, *Wave packet analysis*, CBMS Regional Conference Series in Mathematics, 105

[106] A. Weil, *On the Riemann hypothesis in function fields*, Proc. Nat. Acad. Sci. U. S. A. 27, (1941). 345-347.

[107] T. Ziegler, *Universal characteristic factors and Furstenberg averages*, J. Amer. Math. Soc. 20 (2007), no. 1, 53-97.

[108] Y. Zhang, *Bounded gaps between primes*, Ann. of Math. **179** (2014), no. **3**, 1121-1174.

[109] P. Zorin-Kranich, *Variation estimates for averages along primes and polynomials*, J. Funct. Anal. 268 (2015), no. 1, 210-238.

[110] P. Zorin-Kranich, *Cancellation for the simplex Hilbert transform*, Math. Res. Lett. 24 (2017), no. 2, 581-592.