



# THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### A polynomial Roth theorem on the real line

**Citation for published version:**

Durcik, P, Guo, S & Roos, J 2019, 'A polynomial Roth theorem on the real line', *Transactions of the American Mathematical Society*, vol. 371, no. 10, pp. 6973-6993. <https://doi.org/10.1090/tran/7574>

**Digital Object Identifier (DOI):**

[10.1090/tran/7574](https://doi.org/10.1090/tran/7574)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

Transactions of the American Mathematical Society

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# A POLYNOMIAL ROTH THEOREM ON THE REAL LINE

POLONA DURCIK, SHAOMING GUO, AND JORIS ROOS

ABSTRACT. For a polynomial  $P$  of degree greater than one, we show the existence of patterns of the form  $(x, x + t, x + P(t))$  with a gap estimate on  $t$  in positive density subsets of the reals. This is an extension of an earlier result of Bourgain. Our proof is a combination of Bourgain's approach and more recent methods that were originally developed for the study of the bilinear Hilbert transform along curves.

## 1. INTRODUCTION

Let  $P : \mathbb{R} \rightarrow \mathbb{R}$  be a polynomial. We will let  $\|P\|$  denote the  $\ell^1$  sum of the coefficients of  $P$ . The main result of this paper is the following.

**Theorem 1.1.** *Let  $M \geq 1$  and  $N \geq 1$  be real numbers. Let  $\varepsilon > 0$  be given and  $S$  be a measurable subset of  $[0, N]$  with  $|S| \geq \varepsilon N$ . Let  $P : \mathbb{R} \rightarrow \mathbb{R}$  be a monic polynomial of degree  $d > 1$  without constant term that satisfies  $\|P\| \leq M$ . Then there exists  $\delta(\varepsilon, M, d) > 0$  such that we can find*

$$x, x + t, x + P(t) \in S \tag{1.1}$$

with  $t > \delta(\varepsilon, M, d)N^{1/d}$  and  $\delta = \delta(\varepsilon, M, d)$  satisfies the estimate  $(\log \log \delta^{-1})^{-\frac{1}{6}} \gtrsim_{M,d} \varepsilon$ .

When  $P(t) = 2t$ , then (1.1) is a consequence of the classical Roth theorem [Rot53]. See also [Bou86] for an alternative proof and extensions to results of Szemerédi type. In the special case  $P(t) = t^d$  with  $d \neq 1$ , Theorem 1.1 is due to Bourgain [Bou88]. We extend his result to general polynomials.

A standard argument based on Lebesgue's density theorem shows that our theorem would hold trivially if we only asked for  $t$  to be positive. Thus, the main point of our result is the gap estimate giving a lower bound on  $t$ . A related work that studies the existence of certain polynomial patterns in fractal subsets of  $\mathbb{R}^n$  is [HLP15]. The Lebesgue density argument does not apply in that case, so the goal in that paper is establishing existence of certain patterns without a gap estimate.

After concluding our work on this result we were made aware that a less quantitative variant of our theorem (without the estimate on  $\delta$ ) can also be deduced from the main result in [BL96].

For the integers, the problem of searching for polynomial patterns in various sets, for instance the primes, has been studied intensively. We refer to [TZ08], [TZ16] and the references contained therein.

Our proof of Theorem 1.1 is closely related to the study of the bilinear Hilbert transform along polynomial curves. Define

$$\mathcal{H}_P(f, g)(x) := p.v. \int_{\mathbb{R}} f(x - t)g(x - P(t)) \frac{dt}{t}. \tag{1.2}$$

If  $P(t) = 2t$ , this is the classical bilinear Hilbert transform, which is the subject of Lacey and Thiele's breakthrough papers [LT97], [LT98] and has since been studied extensively. For certain nonlinear  $P$  the operator in (1.2) has recently been studied in [Li13], [Lie11], [LX16], [GX16], [Lie15]. We invite the reader to consult these papers to learn about the development of this subject.

---

*Date:* January 15, 2019.

*2010 Mathematics Subject Classification.* 05D10, 42B20.

Another closely related object is the Hilbert transform along the curve  $(t, P(t))_{t \in \mathbb{R}}$ . For a function  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ , we let

$$H_P(f)(x, y) = p.v. \int_{\mathbb{R}} f(x - t, y - P(t)) \frac{dt}{t}. \quad (1.3)$$

In fact, the operators  $\mathcal{H}_P$  and  $H_P$  share the same Fourier multiplier. We refer to [GHLR17] and the references contained therein for historical background on the study of the operator (1.3).

The principle of using estimates for multilinear singular integrals to study patterns in subsets of the Euclidean space has also been used elsewhere in the recent literature (see [CMP15], [DKR16]).

We now turn to describing the structure of the proof of Theorem 1.1. From now on  $P$  will be a fixed monic polynomial of degree  $d > 1$  satisfying  $\|P\| \leq M$  and lacking a constant term. Let  $f$  be a function on  $\mathbb{R}$  such that  $0 \leq f \leq 1$  and  $\int_0^N f \geq \varepsilon N$ . Assume that we could prove

$$\int_0^N \int_0^{N^{1/d}} f(x) f(x+t) f(x+P(t)) dt dx > \delta(\varepsilon, M, d) N^{1+\frac{1}{d}}.$$

Then Theorem 1.1 follows immediately by setting  $f = \mathbb{1}_S$ . While possibly changing  $\delta$  by multiplication with a constant depending only on  $d$ , we may assume without loss of generality that  $N = 2^{jd}$  for some  $j \in \mathbb{Z}_{\geq 0}$ . Changing variables  $x \rightarrow Nx$  and  $t \rightarrow N^{1/d}t$  and replacing  $f(x)$  by  $f(N^{-1}x)$ , we see that it suffices to show

$$\int_0^1 \int_0^1 f(x) f(x + 2^{-(d-1)jt}) f(x + 2^{-dj} P(2^j t)) dt dx > \delta(\varepsilon, M, d) \quad (1.4)$$

for all functions  $f$  with  $0 \leq f \leq 1$ ,  $\int_0^1 f \geq \varepsilon$ .

In the case  $P(t) = t^d$  with  $d \neq 1$ , Bourgain [Bou88] proved (1.4) for all  $j \in \mathbb{Z}_{\geq 0}$ . Note that in this case  $P$  is scaling-invariant in the sense that  $2^{-dj} P(2^j t) = P(t)$ . For a general polynomial, we do not know how to prove (1.4) for all  $j \in \mathbb{Z}_{\geq 0}$ . However we can prove (1.4) for sufficiently many  $j$ .

**Definition 1.2.** Let  $\Gamma_d$  be a constant depending only on the degree  $d$  that is to be determined later. A set  $\mathcal{E} = \{j_1 < j_2 < \dots\} \subset \mathbb{Z}_{\geq 0}$  is called *admissible* if

$$j_1 \leq \Gamma_d \text{ and } j_{i+1} - j_i \leq \Gamma_d \text{ for every } i \geq 1.$$

To prove Theorem 1.1, it suffices to show (1.4) for all  $j$  contained in an admissible set  $\mathcal{E}$ .

**Proposition 1.3.** *There exists an admissible set  $\mathcal{E} \subset \mathbb{Z}_{\geq 0}$  with  $0 \in \mathcal{E}$  such that for every  $\varepsilon > 0$ , we can find  $\delta > 0$  with  $(\log \log \delta^{-1})^{-\frac{1}{6}} \gtrsim_{M,d} \varepsilon$  such that for every  $0 \leq f \leq 1$  with  $\int_0^1 f \geq \varepsilon$  we have*

$$\int_0^1 \int_0^1 f(x) f(x + 2^{-(d-1)jt}) f(x + 2^{-dj} P(2^j t)) dt dx > \delta, \quad (1.5)$$

for every  $j \in \mathcal{E}$ .

Let  $\tau$  be a non-negative smooth bump function supported in  $[1/2, 2]$  with integral 1. For  $\ell \in \mathbb{Z}_{\geq 0}$  we denote  $\tau_\ell(x) = 2^\ell \tau(2^\ell x)$ . The key lemma in the proof of Proposition 1.3 is the following.

**Lemma 1.4.** *There exists  $\gamma > 0$  and admissible sets  $\mathcal{E}, \Lambda$ ,  $0 \in \mathcal{E}$ , such that for every  $(j, \ell) \in \mathcal{E} \times \Lambda$  and every test function  $g$  with  $\text{supp}(\widehat{g})$  contained in  $[2^m, 2^{m+1}]$ ,  $m \geq 0$ , we have*

$$\left\| \int_{\mathbb{R}} f(x + 2^{-(d-1)jt}) g(x + 2^{-dj} P(2^j t)) \tau_\ell(t) dt \right\|_{L^1([0,1])} \leq C_\ell 2^{-\gamma m} \|f\|_2 \|g\|_2, \quad (1.6)$$

where  $C_\ell \leq 2^{\gamma_{d,M}\ell}$  for some  $\gamma_{d,M} > 0$  depending only on  $d$  and  $M$ .

In the case that  $P$  is a monomial, Bourgain [Bou88] proved Lemma 1.4 for all  $j, \ell \in \mathbb{Z}_{\geq 0}$ . In Section 2 we show how the lemma implies Proposition 1.3. The admissible sets  $\mathcal{E}$  and  $\Lambda$  are constructed in Section 3. In Section 4 we prove Lemma 1.4.

Estimate (1.6) is related to certain estimates for the bilinear Hilbert transforms along curves that appeared in [Li13], [Lie11], [LX16], [GX16], [Lie15]. This enables us to adapt the approach used in these papers to prove Lemma 1.4. One difference to the congruent estimates for the bilinear Hilbert transform along curves is that (1.6) contains an extra scaling parameter. Another difference is that we are allowing  $P$  to have a linear term, while the methods described in the present literature for the bilinear Hilbert transform along curves cannot handle linear terms. In particular, the problem of bounding  $\mathcal{H}_P$  for  $P$  being a polynomial of degree greater than one that includes a linear term is still open.

For the proof of Lemma 1.4 we borrow a basic idea from the study of bilinear Hilbert transforms along curves: we treat a general polynomial as a perturbation of whatever monomial is dominating at each scale. However, at those scales where the linear term is dominating this turns out not to be enough. In that case we need to go one step further and see which of the remaining monomials is dominating the difference of the polynomial and its linear term.

**Notation.** Throughout this paper, we will write  $x \lesssim y$  to mean that there exists a constant  $C$  depending only on fixed parameters depending on context (such as the degree  $d$  of the polynomial  $P$ ) such that  $x \leq Cy$ . We write  $x \lesssim_p y$  to denote dependence of the implicit constant on the parameter  $p$ . Similarly, we define  $x \approx y$  to mean that  $x \lesssim y$  and  $y \lesssim x$ .  $\mathbb{1}_E$  will always denote the characteristic function of the set  $E$ .

**Acknowledgements.** The authors thank Christoph Thiele for a helpful discussion of Bourgain's approach. They also thank Pavel Zorin-Kranich for pointing out the connection to Bergelson and Leibman's work. The second and third authors are indebted to Xiaochun Li, Victor Lie and Lechao Xiao for their generosity and numerous discussions on bilinear Hilbert transforms along curves. The first author is supported by the Hausdorff Center for Mathematics. The third author is supported by the Hausdorff Center for Mathematics and the German National Academic Foundation. This material is based upon work supported by the National Science Foundation under Grant No. DMS-1440140 while the authors were in residence at the Mathematical Sciences Research Institute in Berkeley, California, during the Spring semester of 2017.

## 2. REDUCTION TO THE MAIN LEMMA

In this section we derive the estimate (1.5) from Lemma 1.4. The derivation is a straightforward adaptation of Bourgain's argument [Bou88] to our setting. Let  $\vartheta$  be a non-negative even smooth function supported on  $[-2, 2]$ , constant on  $[-1, 1]$ , and monotone on  $[1, 2]$ . We normalize it such that  $\widehat{\vartheta}(0) = 1$  and denote  $\vartheta_\ell(x) := 2^\ell \vartheta(2^\ell x)$ .

**Lemma 2.1** (Bourgain [Bou88]). *For a non-negative function  $f$  supported on  $[0, 1]$  and  $k, l \in \mathbb{Z}_{\geq 0}$  we have*

$$\int_0^1 f(f * \vartheta_k)(f * \vartheta_l) \geq c_0 \left( \int_0^1 f \right)^3$$

for some constant  $c_0 > 0$  depending only on the choice of  $\vartheta$ .

We include a proof of Lemma 2.1, which was omitted in [Bou88].

*Proof.* In this proof all intervals are dyadic, that is, of the form  $[2^k m, 2^k(m+1)]$  for  $k, m \in \mathbb{Z}$ . For  $k \in \mathbb{Z}$  we denote by  $E_k$  the dyadic martingale averages

$$E_k f = \sum_{|I|=2^{-k}} \left( \frac{1}{|I|} \int_I f \right) 1_I.$$

We claim that for any  $0 \leq k \leq \ell$  we have

$$\int_0^1 f(E_k f)(E_\ell f) \geq \left( \int_0^1 f \right)^3.$$

Once this is shown, Lemma 2.1 follows by bounding the dyadic averages pointwise from above by the continuous averages  $f * \vartheta_\ell$ .

To see the claim we first observe that for any dyadic interval  $J \subseteq [0, 1]$  and any  $k \in \mathbb{Z}$  with  $2^{-k} \leq |J|$  we have by Cauchy-Schwarz

$$\begin{aligned} \left( \frac{1}{|J|} \int_J f \right)^2 &= \left( \frac{1}{|J|} \sum_{I \subseteq J, |I|=2^{-k}} \int_I f \right)^2 \\ &\leq \frac{1}{|J|} \sum_{I \subseteq J, |I|=2^{-k}} \frac{1}{|I|} \left( \int_I f \right)^2 = \frac{1}{|J|} \int_J f(E_k f). \end{aligned}$$

Combining this estimate with Hölder's inequality we obtain

$$\begin{aligned} \left( \int_0^1 f \right)^3 &\leq \sum_{J \subseteq [0,1], |J|=2^{-\ell}} \frac{1}{|J|^2} \left( \int_J f \right)^3 \\ &\leq \sum_{J \subseteq [0,1], |J|=2^{-\ell}} \left( \frac{1}{|J|} \int_J f \right) \int_J f(E_k f) = \int_0^1 f(E_k f)(E_\ell f), \end{aligned}$$

which proves the claim.  $\square$

Now we are ready to deduce Proposition 1.3 from Lemmata 1.4 and 2.1.

*Proof of Proposition 1.3.* By localization in  $x$  we may assume that  $\text{supp}(f) \subseteq [0, 1]$ . Denote

$$I = \int_0^1 \int_0^1 f(x) f(x + 2^{-(d-1)j} t) f(x + 2^{-dj} P(2^j t)) dx dt.$$

For  $\ell', \ell, \ell'' \in \Lambda$  with  $1 \leq \ell' \leq \ell \leq \ell''$  we have

$$\begin{aligned} 2^\ell I &\geq \int_0^1 \int_0^1 f(x) f(x + 2^{-(d-1)j} t) f(x + 2^{-dj} P(2^j t)) \tau_\ell(t) dx dt \\ &= I_1 + I_2 + I_3, \end{aligned}$$

where

$$\begin{aligned} I_1 &= \int_0^1 \int_0^1 f(x) f(x + 2^{-(d-1)j} t) (f * \vartheta_{\ell'}) (x + 2^{-dj} P(2^j t)) \tau_\ell(t) dx dt, \\ I_2 &= \int_0^1 \int_0^1 f(x) f(x + 2^{-(d-1)j} t) (f * \vartheta_{\ell''} - f * \vartheta_{\ell'}) (x + 2^{-dj} P(2^j t)) \tau_\ell(t) dx dt, \\ I_3 &= \int_0^1 \int_0^1 f(x) f(x + 2^{-(d-1)j} t) (f - f * \vartheta_{\ell''}) (x + 2^{-dj} P(2^j t)) \tau_\ell(t) dx dt. \end{aligned}$$

We analyze each of the terms separately. Splitting  $f - f * \vartheta_{\ell''}$  into Littlewood-Paley pieces and applying Lemma 1.4, it follows that for some  $\sigma > 0$  we have

$$|I_3| \leq 2^{\gamma_{d,M} \ell - \sigma \ell''} \|f\|_{L^2(\mathbb{R})}^2 \leq 2^{-100} c_0 \varepsilon^3,$$

where the last inequality holds provided that  $\ell''$  is taken large enough with respect to  $\ell$ . Here  $c_0$  is the constant from Lemma 2.1.

To estimate  $I_2$  we apply the Cauchy-Schwarz inequality in  $x$ , which yields

$$\begin{aligned} |I_2| &\leq \int_0^1 \|f(x)f(x+2^{-(d-1)j}t)\|_{L_x^2} \|(f * \vartheta_{\ell''} - f * \vartheta_{\ell'})(x+2^{-dj}P(2^j t))\|_{L_x^2} \tau_{\ell}(t) dt \\ &\leq \|f * \vartheta_{\ell''} - f * \vartheta_{\ell'}\|_2. \end{aligned}$$

Passing to the last line we bounded the  $L^\infty$  norm of  $f$  and the  $L^1$  norm of  $\tau_{\ell}$  by one.

To estimate  $I_1$  we compare it with

$$\begin{aligned} I_4 &= \int_0^1 \int_0^1 f(x)f(x+2^{-(d-1)j}t)(f * \vartheta_{\ell'})(x)\tau_{\ell}(t) dx dt \\ &= \int_0^1 f(x)(f * \vartheta_{\ell'})(x)(f * \tau_{\ell+(d-1)j})(x) dx. \end{aligned}$$

Consider the difference

$$I_4 - I_1 = \int_0^1 \int_0^1 f(x)f(x+2^{-(d-1)j}t)((f * \vartheta_{\ell'})(x) - (f * \vartheta_{\ell'})(x+2^{-dj}P(2^j t)))\tau_{\ell}(t) dx dt.$$

By the mean value theorem we obtain

$$|(f * \vartheta_{\ell'})(x) - (f * \vartheta_{\ell'})(x+2^{-dj}P(2^j t))| \leq 2^{\ell'} \|f * (\vartheta')_{\ell'}\|_{\infty} |2^{-dj}P(2^j t)| \leq dM2^{\ell'-\ell+1},$$

whenever  $t$  is in the support of  $\tau_{\ell}$ . Choosing  $\ell$  large enough with respect to  $\ell'$  gives

$$|I_4 - I_1| \leq 2^{-100} c_0 \varepsilon^3.$$

We return to analyzing the term  $I_4$ , which we write as

$$I_4 = \left( \int_0^1 f(x)(f * \vartheta_{\ell'})(x)((f * \tau_{\ell+(d-1)j})(x) - (f * \vartheta_{\ell'+(d-1)j})(x)) dx \right) \quad (2.1)$$

$$+ \left( \int_0^1 f(x)(f * \vartheta_{\ell'})(x)(f * \vartheta_{\ell'+(d-1)j})(x) dx \right) \quad (2.2)$$

By Lemma 2.1, the term (2.2) is bounded from below by  $c_0 \varepsilon^3$ . For (2.1) we use the triangle inequality and Young's convolution inequality to estimate

$$\begin{aligned} &\|f * \tau_{\ell+(d-1)j} - f * \vartheta_{\ell'+(d-1)j}\|_2 \\ &\leq \|(f * \tau_{\ell+(d-1)j} * \vartheta_{\ell''+(d-1)j}) - (f * \vartheta_{\ell'+(d-1)j} * \tau_{\ell+(d-1)j})\|_2 \end{aligned} \quad (2.3)$$

$$+ \|\tau_{\ell+(d-1)j} - (\tau_{\ell+(d-1)j} * \vartheta_{\ell''+(d-1)j})\|_1 \quad (2.4)$$

$$+ \|\vartheta_{\ell'+(d-1)j} - (\vartheta_{\ell'+(d-1)j} * \tau_{\ell+(d-1)j})\|_1 \quad (2.5)$$

By another application of Young's convolution inequality in (2.3) and rescaling in (2.4) and (2.5), we bound the last display by

$$\|(f * \vartheta_{\ell''+(d-1)j}) - (f * \vartheta_{\ell'+(d-1)j})\|_2 + \|\tau_{\ell} - (\tau_{\ell} * \vartheta_{\ell''})\|_1 + \|\vartheta_{\ell'} - (\vartheta_{\ell'} * \tau_{\ell})\|_1.$$

By the mean value theorem, the second and third term are bounded from above by  $2^{-100} c_0 \varepsilon^3$  provided  $\ell''$  is chosen large enough with respect to  $\ell$ , and  $\ell$  large enough with respect to  $\ell'$ . This in turn bounds (2.1) from above by

$$\|f * \vartheta_{\ell''+(d-1)j} - f * \vartheta_{\ell'+(d-1)j}\|_2 + 2^{-99} c_0 \varepsilon^3.$$

From the estimates for the terms  $I_1, I_2, I_3, I_4$  and  $I_4 - I_1$  we obtain

$$c_0 \varepsilon^3 \leq 2^{\ell} I + \|f * \vartheta_{\ell'} - f * \vartheta_{\ell''}\|_2 + \|f * \vartheta_{\ell'+(d-1)j} - f * \vartheta_{\ell''+(d-1)j}\|_2 + 2^{-90} c_0 \varepsilon^3$$

Therefore, we either have  $I > 2^{-\ell-10} c_0 \varepsilon^3$ , or

$$\|f * \vartheta_{\ell'} - f * \vartheta_{\ell''}\|_2 + \|f * \vartheta_{\ell'+(d-1)j} - f * \vartheta_{\ell''+(d-1)j}\|_2 > 2^{-10} c_0 \varepsilon^3.$$

By the preceding discussion we can construct a sequence  $\{\ell_0 < \ell_1 < \dots < \ell_k < \dots\} \subseteq \Lambda$ , which is independent of  $f$  and  $j$  and satisfies  $\ell_{k+1} \leq C\ell_k$  for some sufficiently large constant  $C$  that depends on  $M, \varepsilon, d$  such that for each  $k$  either

$$I > 2^{-\ell_{k+1}-10} c_0 \varepsilon^3$$

or

$$\|f * \vartheta_{\ell_k} - f * \vartheta_{\ell_{k+1}}\|_2 + \|f * \vartheta_{\ell_k+(d-1)j} - f * \vartheta_{\ell_{k+1}+(d-1)j}\|_2 > 2^{-10} c_0 \varepsilon^3. \quad (2.6)$$

Observe that for any  $K \geq 0$  one has

$$\begin{aligned} & \sum_{k=0}^K \left( \|f * \vartheta_{\ell_k} - f * \vartheta_{\ell_{k+1}}\|_2^2 + \|f * \vartheta_{\ell_k+(d-1)j} - f * \vartheta_{\ell_{k+1}+(d-1)j}\|_2^2 \right) \\ & \leq C_0 \|f\|_2^2 \leq C_0 \end{aligned} \quad (2.7)$$

with  $C_0$  independent of  $K$  and  $f$ . Let us fix  $K > C_0 2^{100} c_0^{-2} \varepsilon^{-6}$ . If (2.6) holds for all  $0 < k \leq K$ , then (2.7) yields  $K \leq C_0 2^{100} c_0^{-2} \varepsilon^{-6}$ , which is a contradiction. Thus, for some  $0 \leq k \leq K$  we necessarily have  $I > 2^{-\ell_{k+1}-10} c_0 \varepsilon^3$ . Together with  $\ell_{k+1} \leq C\ell_k$  and  $\ell_0 \leq \Gamma_d$  this gives the lower estimate on  $I$  claimed in Proposition 1.3.  $\square$

### 3. CONSTRUCTION OF ADMISSIBLE SETS

In this section we construct the admissible sets  $\mathcal{E}$  and  $\Lambda$ . We write  $P(t) = t^d + a_{d-1}t^{d-1} + \dots + a_2t^2 + a_1t$  and let  $\Gamma_0$  be a large number depending only on  $d$ , say  $\Gamma_0 = 2^{100d}$ . The precise value of  $\Gamma_0$  is irrelevant. Define

$$\mathcal{J}_r = \{k \in \mathbb{Z} : |a_r(2^k)^r| > \Gamma_0 |a_{r'}(2^k)^{r'}| \text{ for every } r' \neq r\}$$

for  $r = 1, \dots, d$  and similarly,

$$\mathcal{J}_{1,r} = \mathcal{J}_1 \cap \{k \in \mathbb{Z} : |a_r(2^k)^r| > \Gamma_0 |a_{r'}(2^k)^{r'}| \text{ for every } r' \notin \{1, r\}\}$$

for  $r = 2, \dots, d$ . Roughly speaking,  $\mathcal{J}_r$  can be understood as the set of dyadic scales  $k$ , where the  $r$ th power monomial dominates the behavior of the polynomial (and its derivatives). We further denote

$$\mathcal{J}_{\text{good}} = \bigcup_{2 \leq r \leq d} \mathcal{J}_r \cup \mathcal{J}_{1,r} \quad \text{and} \quad \mathcal{J}_{\text{bad}} = \mathbb{Z} \setminus \mathcal{J}_{\text{good}}.$$

Then the following variant of a lemma of Li and Xiao [LX16] holds.

**Lemma 3.1.** *We have*

$$|\mathcal{J}_{\text{bad}}| \leq \Gamma_d. \quad (3.1)$$

Here  $\Gamma_d$  is a constant that depends only on  $d$ .

*Proof.* This lemma is a slight variant of Lemma 2.1 in [LX16]. We include the proof for the sake of completeness. The claim is that

$$|\mathbb{Z} \setminus \bigcup_{1 \leq r \leq d} \mathcal{J}_r| \leq \Gamma_d. \quad (3.2)$$

Estimate (3.1) then follows from applying this estimate first to the polynomial  $P$  and then to the polynomial  $t \mapsto P(t) - a_1t$ . To prove this estimate, we define  $\mathcal{J}(r', r'')$  to be the collection of integers  $k$  with

$$|a_{r'}| 2^{r'k} \Gamma_0 \geq |a_{r''}| 2^{r''k} \geq |a_{r'}| 2^{r'k} \Gamma_0^{-1}.$$

It is not difficult to see that  $|\mathcal{J}(r', r'')| \leq 4\Gamma_0$ . Moreover,

$$\left( \mathbb{Z} \setminus \bigcup_{1 \leq r \leq d} \mathcal{J}_r \right) \subset \bigcup_{1 \leq r' < r'' \leq d} \mathcal{J}_{r', r''}.$$

This proves (3.2) with  $\Gamma_d = 4d^2\Gamma_0$ .  $\square$

The good and bad sets for the rescaled polynomial  $t \mapsto 2^{-dj}P(2^j t)$  are simply given by shifts of the good and bad sets for  $P$ . Accordingly, we define

$$\mathcal{J}_r^{(j)} = \mathcal{J}_r - j$$

and similarly  $\mathcal{J}_{1,r}^{(j)}$ ,  $\mathcal{J}_{\text{good}}^{(j)}$  and  $\mathcal{J}_{\text{bad}}^{(j)}$ .

Now we construct the admissible sets  $\mathcal{E}$  and  $\Lambda$ . The set  $\mathcal{E}$  will be chosen as a suitable subset of

$$\mathcal{E}_0 = \{2\Gamma_d j : j \in \mathbb{Z}_{\geq 0}\},$$

where  $\Gamma_d$  is the constant from Lemma 3.1. We claim that the set

$$\Lambda_0 = \mathbb{Z}_{\geq 0} \setminus \bigcup_{j \in \mathcal{E}_0} -\mathcal{J}_{\text{bad}}^{(j)}$$

is admissible. Indeed, looking at residue classes modulo  $2\Gamma_d$ , we note that the cardinality of

$$\left( \bigcup_{j \in \mathcal{E}_0} -\mathcal{J}_{\text{bad}}^{(j)} \bmod 2\Gamma_d \right) = \left( -\mathcal{J}_{\text{bad}} \bmod 2\Gamma_d \right)$$

is at most  $\Gamma_d$  by Lemma 3.1. This proves the claim. By construction, we have that

$$-\ell \in \mathcal{J}_{\text{good}}^{(j)}$$

holds for every  $j \in \mathcal{E}_0$  and  $\ell \in \Lambda_0$ . That is, the polynomial  $t \mapsto 2^{-dj}P(2^j t)$  behaves like a monomial on the annulus  $|t| \in [2^{-\ell}, 2^{-\ell+1}]$ .

For  $r = 1, \dots, d$ , we let  $b_r$  be the unique integer such that<sup>1</sup>

$$|a_r| \in [2^{b_r}, 2^{b_r+1}). \quad (3.3)$$

For reasons that will become clear in Section 4 (see (4.4) and the discussion below (4.22)) we require the condition

$$|b_r + (r-1)(j-\ell)| \geq 10^{10d} \text{ for every } 2 \leq r \leq d. \quad (3.4)$$

to hold. Now we just pick

$$0 = j_0 < \ell_0 < j_1 < \ell_1 < \dots$$

with  $j_i \in \mathcal{E}_0$ ,  $\ell_i \in \Lambda_0$  such that (3.4) holds (this is possible by looking at residue classes modulo  $2\Gamma_d$  again). Then we set  $\mathcal{E} = \{j_0, j_1, \dots\}$  and  $\Lambda = \{\ell_0, \ell_1, \dots\}$ .

#### 4. THE MAIN ARGUMENT

In this section we prove Lemma 1.4. Let  $(j, \ell) \in \mathcal{E} \times \Lambda$ , where  $\mathcal{E}$  and  $\Lambda$  are the sets constructed in the previous section. Then there exists  $1 \leq d_0 \leq d$  such that  $-\ell \in \mathcal{J}_{d_0}^{(j)}$ . Thus we have by definition that

$$|a_{d_0}(2^{j-\ell})^{d_0}| > \Gamma_0 |a_r(2^{j-\ell})^r| \quad (4.1)$$

for all  $r \neq d_0$ . In other words, the  $d_0$ th order monomial dominates the absolute value of  $P(x)$  at the scale  $x \approx 2^{j-\ell}$ . Note that the same automatically holds for all derivatives of the polynomial. Recall the definition of  $b_r$  from (3.3). By (4.1) and since  $a_d = 1$ , we have the lower bound

$$b_{d_0} \geq (j-\ell)(d-d_0). \quad (4.2)$$

Notice that if we are in the case  $m \leq 100d\ell$ , then the desired estimate (1.6) will follow simply from Minkowski's and Hölder's inequalities, since the right hand side of (1.6) is allowed to depend on  $\ell$  as indicated in Lemma 1.4.

In the rest of this section, we always assume that  $m > 100d\ell$ . The claim in Lemma 1.4 about the dependence of the constant on  $\ell$  and  $M$  is easily seen by an inspection of the proof. In order to simplify notation, we will not make any further comments on

<sup>1</sup>We assume without loss of generality that  $a_r \neq 0$ .



this issue and merely indicate the dependence of inequalities on  $\ell$  by writing  $\lesssim_\ell$  and similarly for  $M$ . By  $P_k f$  we denote the frequency projection defined by

$$\widehat{P_k f}(\xi) = \widehat{f}(\xi) \mathbb{1}_k(\xi),$$

where  $\mathbb{1}_k(\xi) = \mathbb{1}_{\{|\xi| \in 2^k[1,2)\}}$ . Then the quantity we need to bound can be written as the  $L^1([0, 1])$  norm of

$$\sum_{k \in \mathbb{Z}} \int_{\mathbb{R}} (P_k f)(x + 2^{-(d-1)j}t) g(x + 2^{-dj} P(2^j t)) \tau_\ell(t) dt. \quad (4.3)$$

Passing to the Fourier side and looking at potential critical points of the phase we expect the main contribution to come from the case when

$$k = b_{d_0} + (d_0 - 1)(j - \ell) + m.$$

This motivates us to write

$$m_0 = b_{d_0} + (d_0 - 1)(j - \ell). \quad (4.4)$$

Observe that (4.2) implies the following lower bound on  $m_0$ :

$$m_0 \geq (d - 1)(j - \ell). \quad (4.5)$$

Now we write (4.3) as

$$\sum_{k \in \mathbb{Z}} \int_{\mathbb{R}} P_{m+m_0+k} f(x + 2^{-(d-1)j}t) g(x + 2^{-dj} P(2^j t)) \tau_\ell(t) dt.$$

Let us first consider the case that  $|k|$  is large, say greater than  $10^d$ . Due to the lack of critical points in the phase we do not expect a large contribution from this term. We dualize using  $h \in L^\infty([0, 1])$  and consider

$$\sum_{|k| > 10^d} \int_0^1 \int_{\mathbb{R}} P_{m+m_0+k} f(x + 2^{-(d-1)j}t) g(x + 2^{-dj} P(2^j t)) \tau_\ell(t) h(x) dt dx.$$

By Fourier inversion, this can be written as (up to a universal constant)

$$\sum_{|k| > 10^d} \iint_{\mathbb{R}^2} \widehat{f}(\xi) \widehat{g}(\eta) \mathbb{1}_{m+m_0+k}(\xi) \widehat{h}(-\xi - \eta) \left( \int_{\mathbb{R}} e^{i\xi 2^{-(d-1)j-\ell}t + i\eta 2^{-dj} P(2^j t)} \tau(t) dt \right) d(\xi, \eta).$$

Note that the  $t$ -derivative of the phase in the integral over  $t$  is  $\gtrsim 2^{-\ell d} 2^m$  (actually it is even larger for positive  $k$ , but we don't need to make use of any decay in  $k$ ). This follows using (4.1) and (4.5) and that  $|k|$  is large. Therefore, integration by parts and an application of the Cauchy-Schwarz inequality to the integration in  $(\xi, \eta)$  yields that the previous display is bounded as

$$\begin{aligned} &\lesssim_\ell 2^{-m} \left( \iint_{\mathbb{R}^2} |\widehat{g}(\eta)|^2 \sum_{|k| > 10^d} \left| \widehat{f}(\xi) \mathbb{1}_{m+m_0+k}(\xi) \right|^2 d(\xi, \eta) \right)^{1/2} \\ &\quad \times \left( \iint_{\mathbb{R}^2} |\widehat{h}(-\xi - \eta)|^2 \mathbb{1}_m(\eta) d(\xi, \eta) \right)^{1/2}. \end{aligned}$$

Here we have used orthogonality of the functions  $\mathbb{1}_k$ . The previous display is

$$\leq 2^{-\frac{m}{2}} \|f\|_2 \|g\|_2 \|h\|_\infty,$$

where we have estimated  $\|\widehat{h}\|_2 = \|h\|_{L^2([0,1])} \leq \|h\|_\infty$ .

Thus it remains to treat the case when  $|k|$  is small (bounded by a constant depending only on  $d$ ). Without loss of generality we set  $k = 0$  to simplify notation. So for the remainder of this section, we will be concerned with the verification of the inequality

$$\left\| \int_{\mathbb{R}} f(x + 2^{-(d-1)j}t)g(x + 2^{-dj}P(2^j t))\tau_{\ell}(t)dt \right\|_{L^1([0,1])} \lesssim_{\ell} 2^{-\gamma m} \|f\|_2 \|g\|_2$$

for some positive  $\gamma > 0$ ,  $\widehat{f}$  supported in  $2^{m+m_0} \leq |\xi| \leq 2^{m+m_0+1}$  and  $\widehat{g}$  supported in  $2^m \leq |\xi| \leq 2^{m+1}$ . We first perform a few preliminary manipulations in order to streamline the argument. Changing variables  $x \rightarrow 2^{-m-m_0}x$  we see that it suffices to show

$$\begin{aligned} & \left\| \int_{\mathbb{R}} f(2^{-m-m_0}x + 2^{-(d-1)j}t)g(2^{-m-m_0}x + 2^{-dj}P(2^j t))\tau_{\ell}(t)dt \right\|_{L^1([0,2^{m_0+m}])} \\ & \lesssim_{\ell} 2^{m+m_0} 2^{-\gamma m} \|f\|_2 \|g\|_2, \end{aligned}$$

By a rescaling of  $f$  and  $g$  it suffices to show

$$\left\| \int_{\mathbb{R}} f(x + \lambda t)g(2^{-m_0}x + \lambda Q(t))\tau(t)dt \right\|_{L^1([0,2^{m_0+m}])} \lesssim_{\ell} 2^{\frac{m_0}{2}} 2^{-\gamma m} \|f\|_2 \|g\|_2,$$

where  $\widehat{f}$  and  $\widehat{g}$  are supported in the annulus  $1 \leq |\xi| \leq 2$  and we have set

$$\lambda = 2^{m+m_0-(d-1)j-\ell} \quad \text{and} \quad Q(t) = 2^{-m_0+\ell-j}P(2^{j-\ell}t). \quad (4.6)$$

By (4.5) we have

$$\lambda \gtrsim_{\ell} 2^m. \quad (4.7)$$

Also note that  $Q$  is well normalized in the sense that  $\|Q\|_{C^d([1/2,2])} \approx 1$ . Dualizing the  $L^1([0,2^{m_0+m}])$  norm using  $h \in L^{\infty}([0,2^{m_0+m}])$  it is enough to show that

$$\left| \int_{\mathbb{R}} \int_{\mathbb{R}} f(x + \lambda t)g(2^{-m_0}x + \lambda Q(t))h(x)\tau(t)dtdx \right| \lesssim_{\ell} 2^{\frac{m_0}{2}} 2^{-\gamma m} \|f\|_2 \|g\|_2 \|h\|_{\infty} \quad (4.8)$$

By Hölder's inequality it suffices to verify that the trilinear estimate

$$\left| \int_{\mathbb{R}} \int_{\mathbb{R}} f(x + \lambda t)g(2^{-m_0}x + \lambda Q(t))h(x)\tau(t)dtdx \right| \lesssim_{\ell} 2^{-\gamma m} 2^{-\frac{1}{2}m} \|f\|_2 \|g\|_2 \|h\|_2 \quad (4.9)$$

holds for  $f, g$  with Fourier support in  $1 \leq |\xi| \leq 2$ . Applying the Fourier inversion formula to  $f$  and  $g$ , the integral on the left hand side of the previous display can be written as (up to a universal constant)

$$\iint_{\mathbb{R}^2} \widehat{f}(\xi)\widehat{g}(\eta)\widehat{h}(-\xi - 2^{-m_0}\eta) \left( \int_{\mathbb{R}} e^{i\lambda(t\xi + Q(t)\eta)} \tau(t)dt \right) d(\xi, \eta).$$

We denote the phase function of the integral in  $t$  by

$$\Phi_{\xi, \eta}(t) = t\xi + \eta Q(t). \quad (4.10)$$

In the following we will always assume that the equation

$$\Phi'_{\xi, \eta}(t_c) = 0$$

has a unique solution  $t_c = t_c(\xi, \eta) \in (1/2, 2)$ . In the case of multiple solutions, each of them can be isolated by adding an appropriate cutoff function in  $t$  (which we silently include into  $\tau$ ) and each is then treated in the exact same way. If on the other hand there is no such solution, we can integrate by parts and apply the Cauchy-Schwarz inequality similarly as above to obtain the desired bound (also see the discussion below (4.12)). Let us denote the dual phase function by

$$\Psi(\xi, \eta) = \Phi_{\xi, \eta}(t_c) = t_c \xi + \eta Q(t_c). \quad (4.11)$$

Note that the existence of a critical point depends on the variables  $\xi$  and  $\eta$  (in the case  $d_0 > 1$  it actually depends only on  $\xi/\eta$ ). Before we proceed we state an incarnation of the stationary phase principle that will be invoked various times during the argument.

**Lemma 4.1.** *Suppose that there exists a unique  $t_0 \in (1/2, 2)$  such that  $\phi'(t_0) = 0$ . Assume  $\phi''(t_0) \neq 0$  and that  $\tau$  is supported in  $[1/2, 2]$ . Then*

$$\int_{\mathbb{R}} e^{i\lambda\phi(t)} \tau(t) dt = \lambda^{-1/2} e^{i\lambda\phi(t_0)} (c|\phi''(t_0)|^{-\frac{1}{2}} \tau(t_0) + R(\lambda)),$$

with  $R^{(r)}(\lambda) = O(\lambda^{-r-\frac{1}{2}})$ . Here,  $c$  is a universal constant and the estimates of the remainder term  $R$  depend only on finitely many derivatives of  $\phi$  and  $\tau$ .

The proof is standard and follows from [Ste93, Ch. VIII.1, Prop. 3], combined with appropriate integration by parts.

We now distinguish three cases. In the first two cases we assume that the dominating monomial is nonlinear, i.e.  $d_0 \geq 2$ . In the third case we assume that the linear term is dominating, that is,  $d_0 = 1$ .

**4.1. Case I:  $d_0 \geq 2$  and  $|m_0| \leq (1-\kappa)m$ .** Here  $\kappa$  is a small, positive universal constant that is to be determined later. In this case we follow the approach of [Lil3, Section 5], and use the  $TT^*$  method to obtain (4.9).

As a consequence of Lemma 4.1 we have

$$\int_{\mathbb{R}} e^{i\lambda\Phi_{\xi,\eta}(t)} \tau(t) dt = \lambda^{-1/2} e^{i\lambda\Psi(\xi,\eta)} a(\xi, \eta) + R_{\xi,\eta}(\lambda), \quad (4.12)$$

where  $a(\xi, \eta)$  is a smooth and compactly supported function and the remainder term  $R_{\xi,\eta}$  satisfies

$$|R_{\xi,\eta}(\lambda)| \lesssim \lambda^{-1},$$

where the implied constant depends only on  $d$ . In the case that  $\xi, \eta$  are such that there exists no critical point (and therefore  $\Psi(\xi, \eta)$  is not well-defined), we may set  $a(\xi, \eta) = 0$ . As a consequence, that case is also handled by the remainder term. The reader should also keep this convention in mind for the remaining applications of stationary phase later in this section. We will not address this issue anymore from now on.

The function  $a(\xi, \eta)$  depends on all our parameters, however does so in a harmless way. For instance, we have  $|a(\xi, \eta)| \approx 1$  (and this information suffices for our purposes). The remainder term in (4.12) can be treated by an application of the Cauchy-Schwarz inequality in  $(\xi, \eta)$ . Indeed, we obtain

$$\left| \iint_{\mathbb{R}^2} \widehat{f}(\xi) \widehat{g}(\eta) \widehat{h}(-\xi - 2^{-m_0}\eta) R_{\xi,\eta}(\lambda) d(\xi, \eta) \right| \lesssim_{\ell} 2^{-m} \|f\|_2 \|g\|_2 \|h\|_2. \quad (4.13)$$

Here we used that  $\lambda^{-1} \lesssim_{\ell} 2^{-m}$ . Turning our attention to the main term, it now remains prove that

$$\left| \iint_{\mathbb{R}^2} \widehat{f}(\xi) \widehat{g}(\eta) e^{i\lambda\Psi(\xi,\eta)} a(\xi, \eta) \widehat{h}(-\xi - 2^{-m_0}\eta) d(\xi, \eta) \right| \lesssim_{\ell} 2^{-\gamma m} \|f\|_2 \|g\|_2 \|h\|_2. \quad (4.14)$$

Changing variables from  $\xi$  to  $\xi - 2^{-m_0}\eta$  and applying the Cauchy-Schwarz inequality to separate the function  $h$ , we see that it suffices to show

$$\left\| \int_{\mathbb{R}} \widehat{f}(\xi - 2^{-m_0}\eta) \widehat{g}(\eta) a(\xi - 2^{-m_0}\eta, \eta) e^{i\lambda\Psi(\xi - 2^{-m_0}\eta, \eta)} d\eta \right\|_{L^2_{\xi}(\mathbb{R})} \lesssim_{\ell} 2^{-\gamma m} \|f\|_2 \|g\|_2. \quad (4.15)$$

Expanding the square of the  $L^2$  norm on the left hand-side gives

$$\begin{aligned} & \int_{\mathbb{R}} \left( \int_{\mathbb{R}} \widehat{f}(\xi - 2^{-m_0}\eta) \widehat{g}(\eta) a(\xi - 2^{-m_0}\eta, \eta) e^{i\lambda\Psi(\xi - 2^{-m_0}\eta, \eta)} d\eta \right) \\ & \quad \times \left( \int_{\mathbb{R}} \overline{\widehat{f}(\xi - 2^{-m_0}\eta') \widehat{g}(\eta') a(\xi - 2^{-m_0}\eta', \eta')} e^{-i\lambda\Psi(\xi - 2^{-m_0}\eta', \eta')} d\eta' \right) d\xi. \end{aligned}$$

The change of variables

$$\eta' \rightarrow \eta - \alpha, \quad \xi - 2^{-m_0}\eta \rightarrow \xi \quad (4.16)$$

transforms the left hand-side of (4.15) into

$$\iiint_{\mathbb{R}^3} F_\alpha(\xi) G_\alpha(\eta) \chi_\alpha(\xi, \eta) e^{i\lambda \Xi_\alpha(\xi, \eta)} d(\eta, \xi, \alpha),$$

where

$$\begin{aligned} F_\alpha(\xi) &= \widehat{f}(\xi) \overline{\widehat{f}(\xi + 2^{-m_0} \alpha)} \\ G_\alpha(\eta) &= \widehat{g}(\eta) \overline{\widehat{g}(\eta - \alpha)}, \\ \chi_\alpha(\xi, \eta) &= a(\xi, \eta) \overline{a(\xi + 2^{-m_0} \alpha, \eta - \alpha)}, \text{ and} \\ \Xi_\alpha(\xi, \eta) &= \Psi(\xi, \eta) - \Psi(\xi + 2^{-m_0} \alpha, \eta - \alpha). \end{aligned} \tag{4.17}$$

We split the integration in  $\alpha$  over the regions  $|\alpha| \leq \alpha_0$  and  $|\alpha| \geq \alpha_0$ , where  $\alpha_0 > 0$  is to be determined later. If  $|\alpha| \leq \alpha_0$  we simply use the triangle inequality and Cauchy-Schwarz in  $(\xi, \eta)$  to estimate

$$\int_{|\alpha| \leq \alpha_0} \iint_{\mathbb{R}^2} |F_\alpha(\xi) G_\alpha(\eta) \chi_\alpha(\xi, \eta)| d(\eta, \xi) d\alpha \lesssim \alpha_0 \|f\|_2^2 \|g\|_2^2. \tag{4.18}$$

If  $|\alpha| \geq \alpha_0$ , the idea is to exploit the cancellation caused by the oscillation of the phase function  $\Xi_\alpha$ . Our claim is that since  $|m_0|$  is large (recall (3.4)) we have

$$|\partial_\xi \partial_\eta \Xi_\alpha(\xi, \eta)| \gtrsim \alpha. \tag{4.19}$$

By implicit differentiation we have

$$\partial_\eta t_c = -\frac{Q'(t_c)}{\eta Q''(t_c)}$$

Moreover we compute

$$\partial_\xi \Psi(\xi, \eta) = t_c + \partial_\xi t_c \cdot \xi + \partial_\xi t_c \cdot \eta Q'(t_c) = t_c,$$

and as a consequence,

$$\partial_\eta \partial_\xi \Psi(\xi, \eta) = \partial_\eta t_c = -\frac{Q'(t_c)}{\eta Q''(t_c)}.$$

Let us set  $H(\xi, \eta) = \partial_\eta \partial_\xi \Psi(\xi, \eta)$ . By the mean value theorem we have

$$H(\xi, \eta) - H(\xi + 2^{-m_0} \alpha, \eta - \alpha) = \nabla H(\tilde{\xi}, \tilde{\eta}) \cdot (-2^{-m_0} \alpha, \alpha), \tag{4.20}$$

where  $(\tilde{\xi}, \tilde{\eta})$  is some convex combination of  $(\xi, \eta)$  and  $(\xi + 2^{-m_0} \alpha, \eta - \alpha)$ . We claim that

$$|\partial_\xi H(\xi, \eta)| \approx_d 1 \text{ and } |\partial_\eta H(\xi, \eta)| \approx_d 1. \tag{4.21}$$

Indeed, we compute

$$\partial_\xi H(\xi, \eta) = \frac{1}{\eta^2 Q''} \left( 1 - \frac{Q''' Q'}{(Q'')^2} \right) \text{ and } \partial_\eta H(\xi, \eta) = \frac{Q'}{\eta^2 Q''} \left( 2 - \frac{Q' Q'''}{(Q'')^2} \right).$$

It is clear that  $|\nabla H(\xi, \eta)| \lesssim 1$ . To obtain the lower bounds, we need to study the fraction

$$\frac{Q' Q'''}{(Q'')^2} = \frac{P'(2^{j-\ell} t_c) P'''(2^{j-\ell} t_c)}{P''(2^{j-\ell} t_c)^2}$$

Since  $2^{j-\ell} t$  is in the range where the  $d_0$ th monomial dominates (i.e. (4.1) holds), we have that this is bounded by

$$\frac{(d_0 + h)(d_0(d_0 - 1)(d_0 - 2) + h)}{(d_0(d_0 - 1) - h)^2}$$

To see this recall that if a monomial dominates the absolute value of the polynomial at a certain scale, then it also dominates the absolute value of the derivatives of the polynomial at that scale. Here,  $h > 0$  can be made as small as we please by making

$\Gamma_0$  larger, if necessary. By continuity, we can make  $h$  small enough such that the above fraction is only  $\frac{1}{100(d_0-1)}$  away from

$$\frac{d_0 - 2}{d_0 - 1}$$

Thus,

$$1 - \frac{Q'Q'''}{(Q'')^2} \geq 1 - \frac{d_0 - 2}{d_0 - 1} - \frac{1}{100(d_0 - 1)} = \frac{99}{100} \cdot \frac{1}{d_0 - 1} \gtrsim_d 1. \quad (4.22)$$

We have therefore proven (4.21). Recall that we have chosen  $j \in \mathcal{E}$  and  $\ell \in \Lambda$ . Thus,  $|m_0|$  is large in the sense that (3.4) holds. Using this we obtain from (4.20) that (4.19) holds. This is because in the case that  $m_0$  is large, the inner product on the right hand side of (4.20) is dominated by the second component and if  $-m_0$  is large, then it is dominated by the first component. Now we use the following well known fact.

**Lemma 4.2** (Hörmander). *Let  $a, \varphi$  be smooth functions in  $\mathbb{R}^2$ ,  $\varphi$  real-valued and  $\lambda > 1$ . Also denote*

$$T_\lambda f(x) = \int_{\mathbb{R}} e^{i\lambda\varphi(x,y)} a(x,y) f(y) dy.$$

*Assume that  $\partial_x \partial_y \varphi(x,y) \neq 0$  in the support of  $a$ . Then we have*

$$|\langle T_\lambda f, g \rangle| \leq C \lambda^{-1/2} \|f\|_2 \|g\|_2.$$

This is a dualized version of the  $L^2$  endpoint of [Hor73, Thm. 1.1]. The proof is simply by  $TT^*$  and stationary phase. Applying this result to our situation we conclude

$$\begin{aligned} & \left| \int_{|\alpha| \geq \alpha_0} \int_{\mathbb{R}} \int_{\mathbb{R}} F_\alpha(\xi) G_\alpha(\eta) e^{i\lambda \Xi_\alpha(\xi, \eta)} d\eta d\xi d\alpha \right| \\ & \lesssim_\ell (2^m \alpha_0)^{-\frac{1}{2}} \int_{|\alpha| \geq \alpha_0} \|F_\alpha\|_2 \|G_\alpha\|_2 d\alpha \lesssim_\ell (2^m \alpha_0)^{-\frac{1}{2}} 2^{\frac{m_0}{2}} \|f\|_2^2 \|g\|_2^2, \end{aligned} \quad (4.23)$$

where the last inequality follows by Cauchy-Schwarz applied to the integration in  $\alpha$ . Thus, if  $|m_0| \leq (1 - \kappa)m$  for some fixed small absolute constant  $\kappa > 0$ , then by letting  $\alpha_0 = 2^{-\frac{\kappa m}{2}}$  we see that our desired estimate (4.14) holds with  $\gamma = \kappa/8$ .

**4.2. Case II:  $d_0 \geq 2$  and  $|m_0| > (1 - \kappa)m$ .** Here we apply a  $\sigma$ -uniformity argument in the spirit of [Li13] and [LX16]. Alternatively, one can also follow the approach of [Lie11] which does not use the concept of  $\sigma$ -uniformity. Before we start, let us briefly review the basic setup of  $\sigma$ -uniformity.

**Definition 4.3** ( $\sigma$ -uniformity). Let  $\sigma \in (0, 1)$ ,  $\mathcal{I} \subset \mathbb{R}$  a bounded interval and  $\mathcal{U}(\mathcal{I})$  a non-trivial subset of  $L^2(\mathcal{I})$  such that  $\sup_{u \in \mathcal{U}(\mathcal{I})} \|u\|_2 < \infty$ . A function  $f \in L^2(\mathcal{I})$  is called  $\sigma$ -uniform in  $\mathcal{U}(\mathcal{I})$  if

$$\left| \int_{\mathcal{I}} f(x) \overline{u(x)} dx \right| \leq \sigma \|f\|_{L^2(\mathcal{I})}, \text{ for every } u \in \mathcal{U}(\mathcal{I}).$$

The main result on  $\sigma$ -uniformity is the following.

**Lemma 4.4** ([Li13]). *Let  $\mathcal{L}$  be a bounded sublinear functional from  $L^2(\mathcal{I})$  to  $\mathbb{C}$ , and  $S_\sigma$  be the set of all functions that are  $\sigma$ -uniform in  $\mathcal{U}(\mathcal{I})$ . Denote*

$$A_\sigma = \sup\{|\mathcal{L}(f)|/\|f\|_{L^2(\mathcal{I})} : f \in S_\sigma, f \neq 0\},$$

and

$$K = \sup_{u \in \mathcal{U}(\mathcal{I})} \frac{|\mathcal{L}(u)|}{\|u\|_2}.$$

Then

$$\|\mathcal{L}\| \leq \max\{A_\sigma, 2\sigma^{-1}K\}.$$

In the following we will apply the lemma to the functional

$$\mathcal{L}(g) = \int_{\mathbb{R}} \int_{\mathbb{R}} f(x + \lambda t) g(2^{-m_0} x + \lambda Q(t)) h(x) \tau(t) dt dx. \quad (4.24)$$

Our goal is to prove (4.8). The interval  $\mathcal{I}$  will mean either  $[1/2, 2]$  or  $[-2, -1/2]$ .

We also define

$$\mathcal{U}(\mathcal{I}) = \{\eta \mapsto a(\xi, \eta) e^{i\alpha\eta + i\lambda\Psi(\xi, \eta)} : \alpha \in \mathbb{R}, 2^{-10} \leq |\xi| \leq 2^{10}\},$$

where  $\lambda, \Psi$  are as defined in (4.6), (4.11) and  $a(\xi, \eta)$  is a compactly supported smooth function that is to be determined later.

First we assume that  $\widehat{g}|_{\mathcal{I}}$  is  $\sigma$ -uniform in  $\mathcal{U}(\mathcal{I})$ . Localizing in the spatial variable  $x$ , we write  $\mathcal{L}(g)$  as

$$\sum_{0 \leq \iota < 2^m} \int_{\mathbb{R}} \int_{\mathbb{R}} (\mathbb{1}_{J_\iota} f)(x + \lambda t) g(2^{-m_0} x + \lambda Q(t)) \tau(t) (\mathbb{1}_{I_\iota} h)(x) dt dx,$$

where  $I_\iota = 2^{m_0}[\iota, \iota + 1]$  and

$$J_\iota = [\iota 2^{m_0} + \frac{1}{2}\lambda, (\iota + 1)2^{m_0} + 2\lambda]$$

We introduce this cumbersome notation because we need to keep track of the spatial localization of  $f$  for technical reasons that become clear at the end of the argument. Denote  $f_\iota = \mathbb{1}_{J_\iota} f$  and  $h_\iota = \mathbb{1}_{I_\iota} h$ . Passing to the Fourier side, we obtain

$$\sum_{0 \leq \iota < 2^m} \iiint_{\mathbb{R}^3} \widehat{f}_\iota(\xi) e^{ix\xi} \widehat{g}(\eta) e^{i2^{-m_0}x\eta} \left( \int_{\mathbb{R}} \tau(t) e^{i\lambda\Phi_{\xi, \eta}(t)} dt \right) h_\iota(x) d(x, \xi, \eta), \quad (4.25)$$

where  $\Phi_{\xi, \eta}$  is as defined in (4.10). Due to the localization in  $x$  we can replace  $e^{i2^{-m_0}x\eta}$  by the constant  $e^{i2^{-m_0}\alpha_\iota\eta}$ , where  $\alpha_\iota$  is an arbitrary point chosen from  $I_\lambda$ . More precisely, we write

$$e^{i2^{-m_0}x\eta} = e^{i2^{-m_0}\alpha_\iota\eta} \sum_{s=0}^{\infty} \frac{i^s}{s!} (2^{-m_0}(x - \alpha_\iota))^s \eta^s.$$

Plugging this into (4.25) we then proceed to treat every term of the Taylor expansion separately. However, since the treatment is the same for each of them we will here only show the argument for the term  $s = 0$  for simplicity of notation. Thus we are left with bounding

$$\sum_{0 \leq \iota < 2^m} \iiint_{\mathbb{R}^3} \widehat{f}_\iota(\xi) e^{ix\xi} \widehat{g}(\eta) e^{i2^{-m_0}\alpha_\iota\eta} \left( \int_{\mathbb{R}} \tau(t) e^{i\lambda\Phi_{\xi, \eta}(t)} dt \right) h_\iota(x) d(x, \xi, \eta).$$

Appealing again to the stationary phase principle in the form of (4.12) leaves us with having to estimate

$$2^{-\frac{m}{2}} \sum_{0 \leq \iota < 2^m} \int_{\mathbb{R}} \left( \int_{\mathbb{R}} \widehat{g}(\eta) a(\xi, \eta) e^{i2^{-m_0}\alpha_\iota\eta + i\lambda\Psi(\xi, \eta)} d\eta \right) \widehat{f}_\iota(\xi) \widehat{h}_\iota(-\xi) d\xi,$$

where  $a(\xi, \eta)$  is a compactly supported smooth function. Recall that  $a(\xi, \eta) = 0$  if there is no critical point and the remainder term from Lemma 4.1 is treated as in (4.13). Now we apply the definition of  $\sigma$ -uniformity and Cauchy-Schwarz to bound the last expression by

$$2^{-\frac{m}{2}} \sigma \sum_{0 \leq \iota < 2^m} \|f_\iota\|_2 \|h_\iota\|_2 \|g\|_2. \quad (4.26)$$

Another application of the Cauchy-Schwarz inequality yields

$$\sum_{0 \leq \iota < 2^m} \|f_\iota\|_2 \|h_\iota\|_2 \leq \left( \sum_{0 \leq \iota < 2^m} \|f_\iota\|_2^2 \right)^{1/2} \left( \sum_{0 \leq \iota < 2^m} \|h_\iota\|_2^2 \right)^{1/2}.$$

This is where we need to make use of the spatial localization of  $f_\ell$  on the interval  $J_\ell$ . We have

$$\sum_{0 \leq \ell < 2^m} \|f_\ell\|_2^2 = \int_{\mathbb{R}} \left( \sum_{0 \leq \ell < 2^m} \mathbb{1}_{J_\ell}(x) \right) |f(x)|^2 dx \lesssim \max(1, 2^{m-(d-1)j-\ell}) \|f\|_2^2.$$

Thus, (4.26) is bounded by

$$2^{-\frac{m}{2}} \sigma \max(1, 2^{(m-(d-1)j-\ell)/2}) \|f\|_2 \|g\|_2 \|h\|_2 \lesssim_{\ell, M} \sigma 2^{\frac{m_0}{2}} 2^{\frac{\kappa}{2}m} \|f\|_2 \|g\|_2 \|h\|_\infty, \quad (4.27)$$

where in the last estimate we have used that  $|m_0| > (1 - \kappa)m$  and that  $h$  is supported on  $[0, 2^{m+m_0}]$ . This finishes the estimate for the case when  $\widehat{g}|_{\mathcal{I}}$  is  $\sigma$ -uniform in  $\mathcal{U}(\mathcal{I})$ .

It remains to treat the case when  $\widehat{g}|_{\mathcal{I}} \in \mathcal{U}(\mathcal{I})$ . We substitute  $x \rightarrow 2^{m+m_0}x - \lambda t$  in in (4.24) and arrive at

$$2^{m+m_0} \int_{\mathbb{R}} f(2^{m+m_0}x) \int_{\mathbb{R}} g(2^m x - 2^{-m_0} \lambda t + \lambda Q(t)) h(2^{m+m_0}x - \lambda t) \tau(t) dt dx$$

Applying Hölder's inequality we bound this by  $\|f\|_2 \|T(g, h)\|_2$ , where

$$T(g, h)(x) = 2^{\frac{m+m_0}{2}} \int_{\mathbb{R}} g(2^m x - 2^{-m_0} \lambda t + \lambda Q(t)) h(2^{m+m_0}x - \lambda t) \tau(t) dt.$$

Using the Fourier inversion formula we see that this is equal to<sup>2</sup>

$$2^{\frac{m+m_0}{2}} \int_{\mathbb{R}} \int_{\mathbb{R}} \widehat{g}(\eta) e^{i\eta\lambda(2^m \lambda^{-1}x - 2^{-m_0}t + Q(t))} h(2^{m+m_0}x - \lambda t) \tau(t) d\xi dt.$$

Using our assumption on  $\widehat{g}$  this becomes

$$2^{\frac{m+m_0}{2}} \int_{\mathbb{R}} \left( \int_{\mathbb{R}} a(\xi, \eta) e^{i\lambda(\eta z_{x,t} + \Psi(\xi, \eta))} d\eta \right) h(2^{m+m_0}x - \lambda t) \tau(t) dt.$$

Here we have set

$$z_{x,t} = \alpha + 2^m \lambda^{-1}x - 2^{-m_0}t + Q(t), \quad (4.28)$$

where  $\alpha \in \mathbb{R}$  is arbitrary and  $\xi$  is a parameter comparable to one whose precise value is irrelevant. We would like to apply the stationary phase principle. The phase function is

$$\tilde{\Phi}_{x,t,\xi}(\eta) = \eta z_{x,t} + \Psi(\xi, \eta).$$

Similarly as above, we may assume that the equation

$$z_{x,t} + \partial_\eta \Psi(\xi, \eta_c) = 0.$$

has a unique solution  $\eta_c = \eta_c(x, t, \xi) \in \mathcal{I}$ . Recall that  $\partial_\eta \Psi(\xi, \eta) = Q(t_c(\xi, \eta))$ . Let us write  $t_c = t_c(\xi, \eta_c(x, t, \xi))$  in the following. Set

$$\tilde{\Psi}_\xi(x, t) = \tilde{\Phi}_{x,t,\xi}(\eta_c) = \eta_c z_{x,t} + \Psi(\xi, \eta_c)$$

Since

$$\Psi(\xi, \eta_c) = t_c \xi + \eta_c Q(t_c)$$

we have

$$\tilde{\Psi}_\xi(x, t) = t_c \xi = \xi Q^{-1}(-z_{x,t}). \quad (4.29)$$

Calculate

$$|\partial_\eta^2 \Psi(\xi, \eta)| = \left| \frac{(Q')^2}{\eta Q''} \right| \approx_d 1.$$

Thus, by stationary phase the major contribution to  $T(g, h)(x)$  is

$$2^{\frac{m_0}{2}} \int_{\mathbb{R}} \tilde{a}_\xi(x, t) e^{i\lambda \tilde{\Psi}_\xi(x,t)} h(2^{m+m_0}x - \lambda t) \tau(t) dt,$$

<sup>2</sup>Up to a universal constant.

for some compactly supported smooth function  $\tilde{a}_\xi$ . Expanding the square of the  $L^2$  norm of this expression and changing variables gives

$$2^{m_0} \int_{\mathbb{R}} \int_{\mathbb{R}} \left( \int_{\mathbb{R}} e^{i\lambda(\tilde{\Psi}_\xi(x,t) - \tilde{\Psi}_\xi(x,t+s))} \tilde{a}_\xi(x,t) \overline{\tilde{a}_\xi(x,t+s)} \right. \\ \left. \times h(2^{m+m_0}x - \lambda t) \overline{h(2^{m+m_0}x - \lambda(t+s))} \tau(t) \overline{\tau(t+s)} dt \right) dx ds.$$

The plan is to integrate by parts in the  $t$  variable. Set

$$H(x) = h(2^{m+m_0}x) \overline{h(2^{m+m_0}x - \lambda s)}, \quad \Theta(t) = \tau(t) \overline{\tau(t+s)}$$

and  $\chi_\xi(x,t) = \tilde{a}_\xi(x,t) \overline{\tilde{a}_\xi(x,t+s)}$ . Then our quantity equals

$$2^{m_0} \int_{\mathbb{R}} \int_{\mathbb{R}} \left( \int_{\mathbb{R}} e^{i\lambda(\tilde{\Psi}_\xi(x,t) - \tilde{\Psi}_\xi(x,t+s))} \Theta(t) H(x - 2^{-m-m_0}\lambda t) dt \right) \chi_\xi(x,t) dx ds.$$

Changing variables  $x \mapsto x + 2^{-m-m_0}\lambda t$  we get

$$2^{m_0} \int_{\mathbb{R}} \int_{\mathbb{R}} \left( \int_{\mathbb{R}} e^{i\lambda(\tilde{\Psi}_\xi(x+2^{-m-m_0}\lambda t,t) - \tilde{\Psi}_\xi(x+2^{-m-m_0}\lambda t,t+s))} \Theta(t) b dt \right) H(x) \\ \times \chi_\xi(x + 2^{-m-m_0}\lambda t, t) dx ds. \quad (4.30)$$

From (4.28) and (4.29) we see that

$$\tilde{\Psi}_\xi(x + 2^{-m-m_0}\lambda t, t) - \tilde{\Psi}_\xi(x + 2^{-m-m_0}\lambda t, t+s)$$

is equal to

$$\xi \cdot (Q^{-1}(-2^m \lambda^{-1} x - Q(t)) - Q^{-1}(-2^m \lambda^{-1} x - 2^{-m_0} s - Q(t+s))) \quad (4.31)$$

Set

$$\vartheta(x,t) = Q^{-1}(-2^m \lambda^{-1} x - Q(t))$$

Writing  $\vartheta = \vartheta(x,t)$  we have

$$\partial_t^2 \vartheta(x,t) = -\frac{Q'(t)^2}{Q'(\vartheta)} \left( \frac{Q''(t)}{Q'(t)^2} + \frac{Q''(\vartheta)}{Q'(\vartheta)^2} \right)$$

We claim that

$$|\partial_t^2 \vartheta(x,t)| \gtrsim_\ell |x|. \quad (4.32)$$

At this point we may assume without loss of generality that  $Q$  is an odd polynomial. This is justified since it suffices to handle the cases that  $Q$  is even or odd and the even case follows in the same way. With that assumption we have

$$|\partial_t^2 \vartheta(x,t)| \approx \left| \frac{Q''(t)}{Q'(t)^2} - \frac{Q''(Q^{-1}(Q(t) + \tilde{x}))}{Q'(Q^{-1}(Q(t) + \tilde{x}))^2} \right| \text{ with } \tilde{x} = 2^m \lambda^{-1} x.$$

Now (4.32) follows immediately from the mean value theorem.

By applying the mean value theorem again, together with the fact that  $2^{-m_0}|s|$  is much smaller compared to  $|s|$ , we obtain that the  $t$  derivative of (4.31) is bounded from below by  $|xs|$ . Hence, by using the triangle inequality on small subsets around the origin in the variables  $x$  and  $s$  and integration by parts on the complement we obtain that

$$|(4.30)| \lesssim_\ell 2^{m_0} 2^{-\beta m} \|h\|_\infty^2 \quad (4.33)$$

for any  $0 < \beta < 1$ . Here we have used that both  $x$  and  $s$  take values in intervals of lengths which can be bounded by constants depending only on  $d$  and  $\ell$ .

Combining estimates (4.27) and (4.33) we obtain from Lemma 4.4 that

$$|\mathcal{L}(g)| \lesssim_{\ell, M} 2^{\frac{m_0}{2}} \max\{\sigma 2^{\frac{\kappa}{2}m}, \sigma^{-1} 2^{-\frac{\beta}{2}m}\} \|f\|_2 \|g\|_2 \|h\|_\infty.$$

Choosing  $\kappa$  small enough,  $\sigma = 2^{-\kappa m}$  and  $\beta = \frac{1}{2}$ , we can bound this by

$$\lesssim 2^{\frac{m_0}{2} - \frac{\kappa}{2}m} \|f\|_2 \|g\|_2 \|h\|_\infty.$$

Thus we have finished the proof of (4.8).



**4.3. Case III:**  $d_0 = 1$ . By construction of the admissible sets  $\mathcal{E}$  and  $\Lambda$ , there exists  $d_1 \neq 1$  such that  $-\ell \in \mathcal{J}_{1,d_1}^{(j)}$ . That is, the linear term dominates at the dyadic scale  $-\ell$  and the  $d_1$ th power monomial dominates the remaining, nonlinear monomials at that scale. We have

$$|a_{d_1}(2^{j-\ell})^{d_1}| > \Gamma_0 |a_r(2^{j-\ell})^r| \text{ for every } r \notin \{1, d_1\}. \quad (4.34)$$

Note that  $m_0 = b_1$  and let us write

$$Q(t) = 2^{-b_1} a_1 t + R(t).$$

For convenience let us assume that  $a_1 = 2^{b_1}$ . This does not affect the argument we give, but simplifies notation. From (4.34) and (4.6) we see that  $\|R\|_{C^d([1/2,2])} \approx 2^{q_0}$ , where

$$0 > q_0 := b_{d_1} + (d_1 - 1)(j - \ell) - b_1 > (d - 1)(j - \ell) - b_1. \quad (4.35)$$

Also, since the linear term is dominating,

$$(j - \ell)(d - 1) < b_1 \leq \log(M).$$

Since  $j \geq 0$  we therefore have that  $|b_1|$  and  $|q_0|$  are comparable to one up to constants depending on  $d, \ell, M$ . Applying the stationary phase principle in the form of Lemma 4.1 and discarding the remainder term based on the same argument that led to (4.13), it now remains to prove

$$\left| \int_{\mathbb{R}} \int_{\mathbb{R}} \widehat{f}(\xi) \widehat{g}(\eta) e^{i\lambda\Psi(\xi,\eta)} a(\xi, \eta) \widehat{h}(-\xi - 2^{-b_1}\eta) d\xi d\eta \right| \lesssim 2^{-\gamma m} \|f\|_2 \|g\|_2 \|h\|_2 \quad (4.36)$$

for some positive  $\gamma > 0$  and  $\widehat{f}, \widehat{g}$  supported in the annulus  $1 \leq |\xi| \leq 2$ . Here  $a$  is a smooth and compactly supported cutoff function and we keep in mind that in the support of  $a$  we have  $|\xi + \eta| \approx 2^{q_0}$ , since this is a necessary condition for the existence of a stationary point.

**Claim 4.5.** *There exist  $\gamma > 0$  and an integer  $C_d$  depending only on the degree  $d$  and intervals  $W_1, \dots, W_{C_d} \subset \mathbb{R}$  of length at most  $2^{-\gamma m}$ , such that whenever  $\xi/\eta \notin W_\iota$  for all  $\iota$ , we have*

$$|\partial_\xi \partial_\eta (2^{-b_1} \partial_\xi \Psi - \partial_\eta \Psi)| \gtrsim 2^{-\gamma m}. \quad (4.37)$$

*Implicit constants depend on  $d, \ell$  and  $M$ .*

This claim will be proven at the end of this section. A condition of the form (4.37) first appeared in the work of Li [Li13], see also Xiao [Xi17] and Gressman and Xiao [GX16]. Let  $\tilde{a} : \mathbb{R} \rightarrow \mathbb{R}$  be a smooth bump function that is equal to one on each enlarged interval  $2W_\iota$  such that  $\|\tilde{a}\|_{C^4} \lesssim 2^{4\gamma m}$ . We bound the left hand side of (4.36) by the sum of

$$\int_{\mathbb{R}} \int_{\mathbb{R}} |\widehat{f}(\xi) \widehat{g}(\eta) a(\xi, \eta) \tilde{a}(\xi/\eta) \widehat{h}(-\xi - 2^{-b_1}\eta)| d\xi d\eta \quad (4.38)$$

and

$$\left| \int_{\mathbb{R}} \int_{\mathbb{R}} \widehat{f}(\xi) \widehat{g}(\eta) e^{i\lambda\Psi(\xi,\eta)} a(\xi, \eta) (1 - \tilde{a}(\xi/\eta)) \widehat{h}(-\xi - 2^{-b_1}\eta) d\xi d\eta \right|. \quad (4.39)$$

By the Cauchy-Schwarz inequality we obtain

$$(4.38) \lesssim_{l,M} 2^{-\frac{\gamma m}{2}} \|f\|_2 \|g\|_2 \|h\|_2.$$

It remains to control (4.39). Here we follow a similar argument as in Case I. Write  $b(\xi, \eta) = a(\xi, \eta) (1 - \tilde{a}(\xi/\eta))$ . Then  $\|b\|_{C^4} \lesssim 2^{4\gamma m}$ . After applying a change of variables and the Cauchy-Schwarz inequality, we conclude that it is enough to prove

$$\left\| \int_{\mathbb{R}} \widehat{f}(\xi - 2^{-b_1}\eta) \widehat{g}(\eta) b(\xi - 2^{-b_1}\eta, \eta) e^{i\lambda\Psi(\xi - 2^{-b_1}\eta, \eta)} d\eta \right\|_{L_\xi^2}^2 \lesssim 2^{-\gamma m} \|f\|_2^2 \|g\|_2^2. \quad (4.40)$$

By the triangle inequality, it suffices to prove (4.40) with a better gain  $2^{-6\gamma m}$  in place of  $2^{-\gamma m}$  for every function  $g$  with  $\widehat{g}$  supported on an interval of length  $2^{-2\gamma m}$ . We expand the square on the left hand side of (4.40). After a change of variable, we obtain

$$\int_{|\alpha| \leq 2^{-2\gamma m}} \iint_{\mathbb{R}^2} e^{i\lambda[\Psi(\xi, \eta) - \Psi(\xi + 2^{-b_1}\alpha, \eta - \alpha)]} F_\alpha(\xi) G_\alpha(\eta) a_\alpha(\xi, \eta) d\xi d\eta d\alpha, \quad (4.41)$$

where  $F_\alpha, G_\alpha$  are as in (4.17), and  $a_\alpha$  is some new compactly supported amplitude. By the mean value theorem we see that

$$\left| \partial_\xi \partial_\eta \left( \Psi(\xi, \eta) - \Psi(\xi + 2^{-b_1}\alpha, \eta - \alpha) \right) \right| \gtrsim 2^{-\gamma' m} |\alpha|. \quad (4.42)$$

Now we are ready to apply Lemma 4.2 to bound (4.41) in the same way as we did in (4.23). This concludes the proof of (4.40).

*Proof of Claim 4.5.* Recall that  $t_c(\xi, \eta)$  is defined via

$$\xi + \eta + \eta R'(t_c) = 0. \quad (4.43)$$

We write  $\rho = R'(t_c) = -\frac{\xi + \eta}{\eta}$ . Recall that

$$\Psi(\xi, \eta) = t_c \cdot (\xi + \eta) + \eta R(t_c). \quad (4.44)$$

Direct computation shows that the expression  $2^{-b_1} \cdot \partial_\xi^2 \partial_\eta \Psi(\xi, \eta) - \partial_\xi \partial_\eta^2 \Psi(\xi, \eta)$  equals

$$\frac{\eta(2\xi + 2^{-b_1}\eta)(R''(t_c))^2 + \xi(\xi + 2^{-b_1}\eta)R'''(t_c)}{\eta^4 R''(t_c)^3}.$$

We have

$$\begin{aligned} \eta(2\xi + 2^{-b_1}\eta) &= -2\eta^2(\rho + (1 - \frac{1}{2}2^{-b_1})), \text{ and} \\ \xi(\xi + 2^{-b_1}\eta) &= \eta^2(\rho + 1)(\rho + (1 - 2^{-b_1})). \end{aligned}$$

Therefore, the left hand side of the claimed inequality is comparable (up to constants depending on  $d, \ell, M$ ) to the absolute value of

$$-2(\rho + (1 - 2^{-b_1-1}))(R''(t_c))^2 + (\rho + 1)(\rho + 1 - 2^{-b_1})R'''(t_c), \quad (4.45)$$

which equals

$$-2(R'(t_c) + (1 - 2^{-b_1-1}))(R''(t_c))^2 + (R'(t_c) + 1)(R'(t_c) + 1 - 2^{-b_1})R'''(t_c),$$

As a function of  $t_c$ , this is a polynomial of degree  $3d - 5$  with bounded coefficients (by a constant depending only on  $d, \ell, M$ ) and leading coefficient away from zero. This implies the claim.  $\square$

## REFERENCES

- [BL96] V. Bergelson, A. Leibman, *Polynomial extensions of van der Waerden's and Szemerédi's theorems*, J. Amer. Math. Soc. **9** (1996), 725–753.
- [Bou86] J. Bourgain, *A Szemerédi type theorem for sets of positive density in  $\mathbb{R}^k$* , Israel J. Math. **54** (1986), no. 3, 307–316.
- [Bou88] J. Bourgain, *A nonlinear version of Roth's theorem for sets of positive density in the real line*, J. Anal. Math. **50** (1988), 169–181.
- [CMP15] B. Cook, Á. Magyar, M. Pramanik, *A Roth type theorem for dense subsets of  $\mathbb{R}^d$*  (2015), to appear in Bull. London Math. Soc., available at arXiv:1511.06010.
- [DKR16] P. Durcik, V. Kovač, L. Rimanić, *On side lengths of corners in positive density subsets of the Euclidean space* (2016), to appear in Int. Math. Res. Not., available at arXiv:1609.09056.
- [GX16] P. Gressman and L. Xiao, *Maximal decay inequalities for trilinear oscillatory integrals of convolution type*. J. Func. Anal., 271, No. 12 (2016), 3695–3726.
- [GHLR17] S. Guo, J. Hickman, V. Lie, J. Roos, *Maximal operators and Hilbert transforms along variable non-flat homogeneous curves*, Proc. London Math. Soc. **115** (2017), no. 1, 177–219.
- [GX16] J. Guo, L. Xiao, *Bilinear Hilbert transforms associated with plane curves*, J. Geom. Anal. **26** (2016), no. 2, 967–995.
- [HLP15] K. Henriot, I. Laba, M. Pramanik. *On polynomial configurations in fractal sets*, Anal. PDE **9** (2016), no. 5, 1153–1184.

- [Hor73] L. Hörmander, *Oscillatory integrals and multipliers on  $FL^p$* , Ark. Mat. **11** (1973), no. 1, 1–11.
- [LT97] M. Lacey and C. Thiele,  *$L^p$  estimates for the bilinear Hilbert transform for  $2 < p < \infty$* , Ann. of Math (2), **146** (1997), no. 3, 693–724.
- [LT98] M. Lacey and C. Thiele, *On Calderón’s conjecture*, Ann. of Math (2), **149** (1999), no. 2, 475–496.
- [Li13] X. Li, *Bilinear Hilbert transforms along curves I: The monomial case*, Anal. PDE **6** (2013), no. 1, 197–220.
- [LX16] X. Li and L. Xiao, *Uniform estimates for bilinear Hilbert transforms and bilinear maximal functions associated to polynomials*, Amer. J. Math. **138** (2016), no. 4, 907–962.
- [Lie11] V. Lie, *On the boundedness of the bilinear Hilbert transform along “non-flat” smooth curves*, Amer. J. Math. **137** (2015), no. 2, 313–363.
- [Lie15] V. Lie, *The boundedness of the Bilinear Hilbert Transform along “non-flat” smooth curves. The Banach triangle case ( $L^r, 1 \leq r < \infty$ )* (2015), to appear in Rev. Mat. Iberoam., available at arXiv:1512.09356.
- [Rot53] K. Roth, *On certain sets of integers*, J. Lond. Math. Soc. **28** (1953), no. 1, 104–109.
- [Ste93] E. M. Stein, *Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals*, Princeton University Press, Princeton, 1993.
- [TZ08] T. Tao and T. Ziegler, *The primes contain arbitrarily long polynomial progressions*, Acta Math. **201** (2008), no. 2, 213–305.
- [TZ16] T. Tao and T. Ziegler, *Polynomial patterns in the primes*. arXiv:1603.07817.
- [Xi17] L. Xiao, *Sharp Estimates for Trilinear Oscillatory Integrals and an Algorithm of Two-dimensional Resolution of Singularities*. Rev. Mat. Ibero., **33**, No. 1 (2017), 67–116.

POLONA DURCIK, UNIVERSITY OF BONN, ENDENICHER ALLEE 60, 53115 BONN, GERMANY  
*E-mail address:* durcik@math.uni-bonn.de

SHAOMING GUO, INDIANA UNIVERSITY BLOOMINGTON, 831 E THIRD ST, BLOOMINGTON, IN 47405, USA

CURRENT ADDRESS: DEPARTMENT OF MATHEMATICS, THE CHINESE UNIVERSITY OF HONG KONG, HONG KONG, CHINA  
*E-mail address:* shaomingguo2018@gmail.com

JORIS ROOS, UNIVERSITY OF BONN, ENDENICHER ALLEE 60, 53115 BONN, GERMANY  
*E-mail address:* jroos@math.uni-bonn.de