

POLYNOMIAL CONFIGURATIONS IN DENSE SUBSETS OF THE PRIME LATTICE

ANDREW LOTT, ÁKOS MAGYAR, GIORGIS PETRIDIS AND JÁNOS PINTZ

ABSTRACT. Let A be a subset of the prime lattice $\mathbb{P}^d$ of positive relative upper density. We show that A contains all polynomial configurations of the form $\underline{x} + P_0(y)\underline{v}_0, \dots, \underline{x} + P_l(y)\underline{v}_l$ satisfying some non-degeneracy condition, for some $\underline{x} \in \mathbb{Z}^d$ and $y \in \mathbb{N}$. Moreover, if $A \subseteq \mathbb{P}_N^d$ of density $\delta > 0$ then one may have that $0 < y < \log^L N$ as long as N is sufficiently large, where L is a constant depending on the configuration and but is independent of N and δ .

1. INTRODUCTION

A celebrated result in analytic number theory, due to Green and Tao [8], states that the primes contain arbitrary long arithmetic progressions. In fact they have proved the following. Let $\mathbb{P}_N$ denote the set of primes up to N .

Theorem A. *Let $\delta > 0$, $V = \{v_0, \dots, v_l\} \subseteq \mathbb{Z}$. If $N \geq N(\delta, V)$ and $A \subseteq \mathbb{P}_N$ with $|A| \geq \delta|\mathbb{P}_N|$ then A contains an affine image of V i.e. a set $V' = \{x + yv_0, \dots, x + yv_l\}$.*

It is easy to see that this implies that any subset of the primes of positive relative upper density contains an affine image of any finite set of integers. In [15] Tao has obtained an analogue of Theorem A for the Gaussian primes and asked if similar results hold for subsets of the prime lattice $\mathbb{P}^d$. The first result in this direction was provided by B. Cook and the second author [3] for finite sets $V \subseteq \mathbb{Z}^d$ which are in *general position* in the sense that $|\pi_j(V)| = |V|$ for every $1 \leq j \leq d$, where $\pi_j : \mathbb{Z}^d \rightarrow \mathbb{Z}$ denotes the orthogonal projection to the j^{th} -axis.

Theorem B. *Let $d \geq 1$, $\delta > 0$ and let $V = \{v_0, \dots, v_l\} \subseteq \mathbb{Z}^d$ be a set in general position. If $N \geq N(\delta, V)$ and $A \subseteq \mathbb{P}_N^d$ with $|A| \geq \delta|\mathbb{P}_N|^d$ then A contains an affine image of V i.e. a set $V' = \{x + yv_0, \dots, x + yv_l\}$.*

The full multidimensional extension of Theorem A were obtained independently in [18, 4] and [6] where it was shown that the conclusion of Theorem B holds for any finite set $V \subseteq \mathbb{Z}^d$.

In [16] Tao and Ziegler has extended Theorem A in another direction, from arithmetic to polynomial progressions which are sets of the form $x + P_0(y), \dots, x + P_l(y)$, where $\mathcal{P} = (P_0, \dots, P_l)$ is a family of integral polynomials. In [17, 19] they have provided quantitative bounds on the size of the parameter y measuring the “size” of the polynomial progression.

Theorem C. *Let $\delta > 0$, $l \in \mathbb{N}$ and let $\mathcal{P} = (P_0, \dots, P_l)$, $P_i \in \mathbb{Z}[x]$, $P_i(0) = 0$. If $N \geq N(\delta, l, \mathcal{P})$ and $A \subseteq \mathbb{P}_N$ with $|A| \geq \delta|\mathbb{P}_N|$, then A contains a polynomial progression $\{x + P_0(y), \dots, x + P_l(y)\}$ for some $x, y \in \mathbb{N}$.*

Moreover, one may have $0 < y < \log^L N$, where $L = L(l, \mathcal{P})$ is a constant depends only the family of polynomials $\mathcal{P} = (P_0, \dots, P_l)$.

Note that the first statement in Theorem C was proved in [16] while the bound $0 < y < \log^L N$ with $L = L(l, P, \delta)$ depending also on the density δ was given in [17]. The dependence on δ was removed in [19] by estimating polynomial averages of local Gowers norms by global ones via so-called concatenation theorems.

Corresponding lower bounds were also discussed in [17], extending a construction in [5], showing that $y \gg \log^l N$ is necessary, even for linear case, when $P_i(y) = iy$ for $0 \leq i \leq l$.

Our aim in this note is twofold. On the one hand we show that Theorem C can be extended to the multi-dimensional setting for polynomial configurations that satisfy a similar non-degeneracy condition as given in Theorem B for affine linear configurations. On the other hand we'd like to present this result in an essentially self contained manner, relying only on a sieve theoretic estimate (see (2.6) below) and a quantitative version of a theorem of Bergelson-Leibman [1] on polynomial configurations in dense subsets of the integer lattice.

Definition 1.1. Let $d \geq 1$, $V = \{v_0, \dots, v_l\} \subseteq \mathbb{Z}^d$ be a finite set and let $\mathcal{P} = (P_1, \dots, P_d)$ be a family of integral polynomials. We say that the polynomial configuration $\mathcal{P}_V = (P_0(y)v_0, \dots, P_l(y)v_l)$ is in general position, if for all $0 \leq j < j' \leq l$ and $1 \leq i \leq d$,

$$(1.1) \quad \deg(\pi_i(P_j(y)v_j - P_{j'}(y)v_{j'})) = \deg(P_j(y)v_j - P_{j'}(y)v_{j'}),$$

as a polynomial in y , π_i being the natural projection to the i^{th} coordinate axis.

Our main result is the following.

Theorem 1. Let $\delta > 0$, $d, l \in \mathbb{N}$ and let $\mathcal{P}_V = (P_0(y)v_0, \dots, P_l(y)v_l)$ with $P_j \in \mathbb{Z}[y]$, $P_j(0) = 0$ be a polynomial configuration in general position. If $N \geq N(\delta, \mathcal{P}, V)$ and $A \subseteq \mathbb{P}_N^d$ with $|A| \geq \delta |\mathbb{P}_N^d|$, then there exist $x, y \in \mathbb{N}$ with $0 < y < \log^L N$, such that

$$(1.2) \quad \{x + P_0(y)v_0, \dots, x + P_l(y)v_l\} \subseteq A,$$

where $L = L(\mathcal{P}, V)$ is a constant depending only on initial data $\mathcal{P}, V$.

Remarks.

- In dimension $d = 1$, taking $v_0 = \dots = v_l = 1$, Theorem 1 reduces to Theorem C.
- If the set $V = \{v_0, \dots, v_l\}$ is in general position then for any polynomial $P \in \mathbb{Z}[y]$, $P(0) = 0$ the polynomial configuration $\mathcal{P}_V := (P(y)v_0, \dots, P(y)v_l)$ is also in general position, hence Theorem 1 extends Theorem B.
- If the polynomials $P_0, \dots, P_l$ have distinct degrees then the pattern $\mathcal{P}_V$ is in general position, as long as $\pi_i(v_j) \neq 0$ for all $1 \leq j \leq l$ and $1 \leq i \leq d$.
- If $A = \mathbb{P}_N^d$ Theorem 1 one follows from the 1-dimensional result of Tao-Ziegler [17] as the polynomial pattern $\mathcal{P}_V(y) = \{P_0(y)v_0, \dots, P_l(y)v_l\}$ is contained in the d -fold direct product of the polynomial progression $\mathcal{P}(y) := \{\pi_i(P_j(y)v_j) : 1 \leq j \leq l, 1 \leq i \leq d\}$.
- In general condition (1.1) is equivalent the following. If the leading term of $P_j(y)v_j - P_{j'}(y)v_{j'}$ is of the form $c_{jj'} y^{d_{jj'}}$ then $\pi_i(c_{jj'}) \neq 0$ for all $1 \leq i \leq d$ and $1 \leq j < j' \leq l$.

Our approach follows that of [16, 17] and roughly consists of two parts. The starting point is to control polynomial averages of the form:

$$(1.3) \quad \Lambda_{\mathcal{P}, V}(f_0, \dots, f_l) := \mathbb{E}_{x \in X} \mathbb{E}_{y \in [M]} f_0(x + P_0(y)v_0) \dots f_l(x + P_l(y)v_l),$$

where $X = (\mathbb{Z}/N\mathbb{Z})^d$, $M := \log^L N$, and we used the averaging notation $\mathbb{E}_{a \in A} f(a) := \frac{1}{|A|} \sum_{a \in A} f(a)$. The first part is to bound these averages in terms of polynomial averages of certain Gowers type

local box norms. Recall that given an s -tuple of vectors $\underline{u}_1, \dots, \underline{u}_s \in \mathbb{Z}^d$, we define the corresponding local box norm at scale $1 \leq M \leq N$, as

$$(1.4) \quad \|f\|_{\square_M(\underline{u}_1, \dots, \underline{u}_s)}^{2^s} := \mathbb{E}_{\underline{x} \in X} \mathbb{E}_{y_1^{(0)}, \dots, y_s^{(0)}, y_1^{(1)}, \dots, y_s^{(1)} \in [M]} \prod_{\omega \in \{0,1\}^s} f(\underline{x} + y_1^{(\omega_1)} \underline{u}_1 + \dots + y_s^{(\omega_s)} \underline{u}_s)$$

For bounded functions, when say $|f_i| \leq 1$, the so-called PET induction scheme of Bergelson-Leibman [1] estimates the averages in (1.3) via an polynomial averages of box norms, namely one has

$$(1.5) \quad |\Lambda_{\mathcal{P},V}(f_0, \dots, f_l)| \lesssim \min_{1 \leq j \leq s} (\mathbb{E}_{h \in [H]^t} \|f_j\|_{\square_M(\underline{Q}_1(h), \dots, \underline{Q}_s(h))}^{2^s})^{\frac{1}{2^s}} + o(1),$$

where $\underline{Q}_1, \dots, \underline{Q}_s : \mathbb{Z}^t \rightarrow \mathbb{Z}^d$ is a family of integral non-constant polynomial maps, and by $o(1)$ we will denote a quantity that goes to 0 as $N \rightarrow \infty$, that may depend on the initial configuration $\mathcal{P}, V$ and parameters d, l . In one dimension it was shown in [16] that such estimates remain true when the functions f_i are not bounded uniformly in N , but bounded by a function ν_N , referred to a pseudo-random measure [8, 16], that behaves very nicely with respect to polynomial type averages satisfying the so-called polynomial forms condition [16, Definition 3.6], described later. In our multidimensional setting it will be crucial to assume that the initial polynomial configuration $\mathcal{P}_V$ is in general position in order to verify estimate (1.5) for a family of polynomials $\underline{Q} = (\underline{Q}_1, \dots, \underline{Q}_s)$ which also remain in general position.

The second part of the argument, is a transference argument of [17] adjusted to the multi-dimensional setting to obtain a decomposition of a function $f \geq 0$ bounded by a pseudo-random measure $\nu^{(d)} = \nu_N^{(d)}$ with respect to a given small quantity $\varepsilon > 0$. More precisely, to show that given a function $f : \mathbb{Z}^n \rightarrow \mathbb{R}$ satisfying $0 \leq f \leq \nu^{(d)}$, one has the decomposition

$$(1.6) \quad f = g + h, \quad 0 \leq g \leq 2 \quad \text{and} \quad \|h\|_{\square_{H,M}(\underline{Q}_1, \dots, \underline{Q}_s)} \leq \varepsilon,$$

where we have used the notation,

$$(1.7) \quad \|h\|_{\square_{H,M}(\underline{Q}_1, \dots, \underline{Q}_s)} := (\mathbb{E}_{h \in [H]^t} \|f_j\|_{\square_M(\underline{Q}_1(h), \dots, \underline{Q}_s(h))}^{2^s})^{\frac{1}{2^s}}.$$

Note that both $\|\cdot\|_{\square_M(\underline{u}_1, \dots, \underline{u}_s)}$ and $\|\cdot\|_{\square_{H,M}(\underline{Q}_1, \dots, \underline{Q}_s)}$ are indeed norms.

The idea to use estimates like (1.5)-(1.6) to prove the existence of arithmetic progressions in the primes originated and was crucial in [8], and later also in [16, 17] for polynomial progressions. In our case starting from a set $A \subseteq [N]^d$ of size $|A| \geq \delta |\mathbb{P}_N|^d$ one constructs a function $f_A : [N]^d \rightarrow \mathbb{R}_{\geq 0}$ so that $\mathbb{E}_{\underline{x} \in [N]^d} f_A(\underline{x}) \geq c_0 \delta$ (for some absolute constant $c_0 > 0$) and $0 \leq f_A(\underline{x}) \leq \log^{(d)}(\underline{x})$ with $\log^{(d)}(\underline{x}) = \prod_{i=1}^d \log x_i$, simply by writing $f_A(\underline{x}) = \mathbf{1}_A(\underline{x}) \log^{(d)}(\underline{x})$. Then one may truncate the support of f_A to $[\sqrt{N}, N - \sqrt{N}]^d$ which changes its average only by a negligible amount, and then one may identify $[N]^d$ with $X = (\mathbb{Z}/N\mathbb{Z})^d$ in estimating the averages $\Lambda_{\mathcal{P},V}(f_A, \dots, f_A)$ as the total size of the polynomial configuration is $O(\log^{CL} N)$ which is much less than $\sqrt{N}$ thus there are no “wraparound” issues.

If one can find a pseudo-random measure $\nu_N^{(d)}(\underline{x}) \gg \log^{(d)}(\underline{x})$ for which (1.5)-(1.6) holds, then by multi-linearity one has

$$(1.8) \quad \Lambda_{\mathcal{P},V}(f_A, \dots, f_A) = \Lambda_{\mathcal{P},V}(g, \dots, g) + O(\varepsilon).$$

Invoking the multi-dimensional polynomial extension of Szemerédi’s theorem by Bergelson-Liebman [1], more precisely its more quantitative version given in [16] one has that

$$(1.9) \quad \Lambda_{\mathcal{P},V}(g, \dots, g) \geq c(\delta),$$

for some constant $c(\delta) > 0$ for sufficiently large N , that may also depend on the initial data $\mathcal{P}, V$. Choosing $\varepsilon > 0$ much smaller than $c(\delta)$ it follows that

$$(1.10) \quad \Lambda_{\mathcal{P}, V}(f_A, \dots, f_A) > 0,$$

which implies the existence of configurations $\underline{x} + P_0(y), \dots, \underline{x} + P_l(y)$ in A with $0 < y < \log^L N$.

This cannot be done due to the irregularity of distribution of primes in small residue classes, however is possible after restricting A , and hence the support of f_A , onto a residue class $\underline{x} \equiv \underline{b} \pmod{W}$, where $W = \prod_{p \leq w}$ is the product of small primes $p \leq w$, with $w = w(N)$ is a function that is a sufficiently slowly growing to infinity with N , i.e. after applying the so-called W -trick [8].

If the polynomials $P_0, \dots, P_l$ are linearly independent it is expected that there are good explicit lower bounds for the constants $c(\delta)$ in estimate (1.9), in light of the recent quantitative breakthrough results by Peluse [12], Peluse-Prendiville-Shao [13], Kuca [9] and Kravitz-Kuca-Leng [10, 11]. Estimating the various $o(1)$ error terms obtained via purely on number theoretic considerations, it seems plausible to obtain similar explicit bounds on the threshold $N(\mathcal{P}, V, \delta)$ in terms of the density δ , under the assumption that the polynomials $P_0, \dots, P_l$ are linearly independent.

2. THE PSEUDO-RANDOM MAJORANT AND THE POLYNOMIAL FORMS CONDITION.

We follow [17] to define the the weighted indicator function of the set $A \subseteq \mathbb{P}_{N'}$ truncated and restricted to an appropriate reduced residue class $\underline{x} \equiv \underline{b} \pmod{W}$. Let $w = w(N') = \frac{1}{10} \log \log \log N'$ and let $W = \prod_{p \leq w} (p \text{ prime})$ and note that $W \ll (\log \log N')^{1/10}$ by the prime number theorem. We will set $N := \lfloor N/W \rfloor$.

If $A \subseteq \mathbb{P}_{N'}$ such that $|A| \geq \delta |\mathbb{P}_{N'}|$ then by pigeonhole principle, we can choose $\underline{b} = (b_1, \dots, b_d)$ so that $(b_i, W) = 1$ for all $1 \leq i \leq d$ and

$$(2.1) \quad |\{\underline{x} \in [\sqrt{N}, N - \sqrt{N}]^d, W\underline{x} + \underline{b} \in A\}| \geq \frac{\delta}{2} \left(\frac{NW}{\phi(W) \log N} \right)^d,$$

for sufficiently large N' , using again the prime number theorem. Define the function $f_A : [N]^d \rightarrow \mathbb{R}_{\geq 0}$,

$$(2.2) \quad f_A(\underline{x}) := \left(c_0 \frac{\phi(W) \log N}{W} \right)^d \mathbf{1}_A(W\underline{x} + \underline{b}),$$

if $\underline{x} \in [\sqrt{N}, N - \sqrt{N}]^d$ and is equal to zero otherwise. Then by (2.1) we have that

$$(2.3) \quad \mathbb{E}_{\underline{x} \in [N]^d} f_A(\underline{x}) = N^{-d} \sum_{\underline{x} \in [N]^d} f_A(\underline{x}) \geq \frac{c_0 \delta}{4},$$

where we used the averaging notation $\mathbb{E}_{b \in B} f(b) := \frac{1}{|B|} \sum_{b \in B} f(b)$. Here $c_0 > 0$ is an absolute constant needed to be able to majorize the function $f_A(\underline{x})$ by the function $\nu^{(d)}(\underline{x}) = \prod_{i=1}^d \nu_i(x_i)$, where is the ν pseudo-random measure used originally in [16, 17], namely

$$(2.4) \quad \nu_i(x) := \frac{\phi(W) \log R}{W} \left(\sum_{d|Wx+b_i} \mu(d) \chi \left(\frac{\log d}{\log R} \right) \right),$$

for $x \in [N]$, where $R = N^{\varepsilon_0}$, μ is the Möbius function and $\chi \in C^\infty(-1, 1)$ is a smooth even function satisfying $\int_0^1 |\chi'(t)|^2 dt = 1$ and $\chi(0) \geq 1/2$. It is easy to see that if we choose c_0 small enough, say $c_0 = \frac{\varepsilon_0}{10}$, then we have

$$(2.5) \quad 0 \leq f_A(\underline{x}) \leq \nu^{(d)}(\underline{x}) \quad \text{for all } \underline{x} \in X.$$

the pseudo-randomness property of the product measure $\nu^{(d)}$ we need is a multi-dimensional analogue of the so-called polynomial forms condition of Tao-Ziegler [17, Proposition 3], based on the following correlation estimate proved for the measure $\nu = \nu_b$ (uniformly for b satisfying $(b, W) = 1$). Let $h_1, \dots, h_J$ be integers (not necessarily distinct) of size $h_j = O(\sqrt{N})$, then

$$(2.6) \quad \mathbb{E}_{x \in [N]} \prod_{i=1}^I \nu(x + h_i) = 1 + o(1) + O\left(\text{Exp}\left(O\left(\sum_{1 \leq i < i' \leq I} \sum_{\substack{w \leq p \leq R^{\log R} \\ p | h_i - h_{i'}}} \frac{1}{p}\right)\right)\right),$$

where $\text{Exp}(x) = e^x - 1$, as long as ε_0 is sufficiently small with respect to J .

Definition 2.1. Let $\underline{Q} = (\underline{Q}_1, \dots, \underline{Q}_J) : \mathbb{Z}^t \rightarrow \mathbb{Z}^d$ be a polynomial map. We say that the map $\underline{Q}$ is non-degenerate if

$$(2.7) \quad \deg(\pi_i(\underline{Q}_j - \underline{Q}_{j'})) \geq 1,$$

for all $1 \leq j < j' \leq J$, $1 \leq i \leq d$, where π_i is the natural projection the i^{th} coordinate axis.

Note in the special case $\underline{Q} = \mathcal{P}_V$ the notion of being non-degenerate is weaker than the notion of being in general position given in Definition 1.1. However this weaker notion is enough to derive the following key property of polynomial maps.

Proposition 2.1. Let t, d, D, J be fixed natural numbers, let ε_0 sufficiently small and L sufficiently large depending on t, d, D, J . Let $\underline{Q} = (\underline{Q}_1, \dots, \underline{Q}_J) : \mathbb{Z}^t \rightarrow \mathbb{Z}^d$ be a non-degenerate integral polynomial map of degree at most D , with coefficients of size $O(W^C)$. Then

$$(2.8) \quad \mathbb{E}_{\underline{h} \in [H]^t} \mathbb{E}_{\underline{x} \in X} \prod_{j=1}^J \nu^{(d)}(\underline{x} + \underline{Q}_j(\underline{h})) = 1 + o(1),$$

where $H = \log^L N$ and $X = (\mathbb{Z}/N\mathbb{Z})^d \simeq [N]^d$.

The following simple estimates involving the Exp function will be useful in proving Proposition 2.1.

Lemma 2.1. Let $C_i > 1$ and $\alpha_i > 0$ for $1 \leq i \leq d$. Then one has for sufficiently large N ,

$$(2.9) \quad \prod_{j=1}^d (1 + o(1) + C_j \text{Exp}(\alpha_j)) \leq 1 + o(1) + \prod_{j=1}^d C_j \text{Exp}(\alpha_1 + \dots + \alpha_d)$$

$$(2.10) \quad \text{Exp}(\alpha_1 + \dots + \alpha_d) \leq \text{Exp}(2^d \alpha_1) + \dots + \text{Exp}(2^d \alpha_d).$$

$$(2.11) \quad \prod_{j=1}^d (1 + o(1) + C \text{Exp}(\alpha_j)) \leq 1 + o(1) + C^d \sum_{j=1}^d \text{Exp}(2^d \alpha_j) \quad (C > 1).$$

Proof. For $d = 2$, one has

$$\begin{aligned} & (1 + o(1) + C_1 \text{Exp}(\alpha_1)) (1 + o(1) + C_2 \text{Exp}(\alpha_2)) \\ &= 1 + o(1) + C_1(1 + o(1)) \text{Exp}(\alpha_1) + C_2(1 + o(1)) \text{Exp}(\alpha_2) + C_1 C_2 \text{Exp}(\alpha_1) \text{Exp}(\alpha_2) \\ &= 1 + o(1) + C_1 C_2 \text{Exp}(\alpha_1 + \alpha_2) + C_1(1 + o(1) - C_2) \text{Exp}(\alpha_1) + C_2(1 + o(1) - C_1) \text{Exp}(\alpha_2) \\ &\leq 1 + o(1) + C_1 C_2 \text{Exp}(\alpha_1 + \alpha_2), \end{aligned}$$

using $\text{Exp}(\alpha_1) \text{Exp}(\alpha_2) = \text{Exp}(\alpha_1 + \alpha_2) - \text{Exp}(\alpha_1) - \text{Exp}(\alpha_2)$. Then (2.9) follows by induction. Writing $\text{Exp}(\alpha_i) = u_i$ for $i = 1, 2$, so $e^{\alpha_i} = 1 + u_i$, one has

$$\text{Exp}(\alpha_1 + \alpha_2) = u_1 + u_2 + u_1 u_2 \leq 2u_1 + 2u_2 + u_1^2 + u_2^2 = \text{Exp}(2\alpha_1) + \text{Exp}(2\alpha_2).$$

Then (2.10) follows by induction on d and (2.11) follows immediately from (2.9) and (2.10). $\square$

Proof of Proposition 2.1. We have by (2.6) and (2.11)

$$\begin{aligned} \mathbb{E}_{\underline{x} \in X} \prod_{j=1}^J \nu^{(d)}(\underline{x} + \underline{Q}_j(\underline{h})) &= \prod_{i=1}^d \left(\mathbb{E}_{x_i} \prod_{j=1}^J \nu_i(x_i + \pi_i(\underline{Q}_j(\underline{h}))) \right) \\ &\leq \prod_{i=1}^d (1 + o(1) + C \text{Exp}(C \alpha_i(\underline{h}))) \leq 1 + o(1) + C^d \sum_{i=1}^d \text{Exp}(C 2^d \alpha_i(\underline{h})), \end{aligned}$$

where

$$\alpha_i(\underline{h}) = \sum_{1 \leq j < j' \leq I} \alpha_{j,j',i}, \quad \text{with} \quad \alpha_{j,j',i}(\underline{h}) = \sum_{\substack{w \leq p \leq R^{\log R}, \\ p | \pi_i(\underline{Q}_j(\underline{h}) - \underline{Q}_{j'}(\underline{h}))}} \frac{1}{p}.$$

Thus to prove (2.7) it is enough to show that for fixed $1 \leq j < j' \leq I$ and $1 \leq i \leq d$, one has

$$(2.12) \quad \mathbb{E}_{\underline{h} \in [H]^t} \sum_{\substack{w \leq p \leq R^{\log R}, \\ p | \pi_j(\underline{Q}_i(\underline{h}) - \underline{Q}_{i'}(\underline{h}))}} \frac{1}{p} = o(1).$$

Note that by our assumptions the polynomial $\Delta Q(\underline{h}) := \pi_i(\underline{Q}_j(\underline{h}) - \underline{Q}_{j'}(\underline{h}))$ has degree at least 1 and at most D .

To estimate the above sum first consider set B of the primes $p \geq w$ so that the polynomial ΔQ vanishes (mod p). Since all such primes p must divide the coefficients of ΔQ which are of size $O(W^C)$ we have

$$w^{|B|} \leq \prod_{p \in B} p \leq W^C \ll e^{Cw},$$

by the prime number theorem, thus

$$\sum_{p \in B} \frac{1}{p} \leq \frac{|B|}{w} \ll \frac{1}{\log w} = o(1).$$

Next, we partition the sum in $p \notin B$ into dyadic intervals $2^k \leq p < 2^{k+1}$, and estimate the contribution of each part depending the size of the parameter k .

Assume first that $2^{k/2} \leq \log H = L \log \log N$. Since the polynomial ΔQ has degree at least 1 (mod p), and p is much smaller than H , the number of $\underline{h} \in [H]^t$ satisfying $p | \Delta Q(\underline{h})$ is at most $H^t/p + o(1)$, thus

$$(2.13) \quad \mathbb{E}_{\underline{h} \in [H]^t} \sum_{2^k \leq p < 2^{k+1}, p | \Delta Q(\underline{h})} \frac{1}{p} \leq \frac{1}{2^k} + o(1).$$

Summing this for $2^k \geq w$ is at most $O(1/w^{1/2}) = o(1)$ again.

Assume now $2^{k/2} \geq \log H$. Then left side of (2.13) is estimated by,

$$\frac{1}{2^k} \mathbb{E}_{\underline{h} \in [H]^t} |\{2^k \leq p < 2^{k+1}; p | \Delta Q(\underline{h})\}| \leq \frac{D \log H}{k 2^k} + O(H^{-1}) \ll \frac{1}{k 2^{k/2}} + o(1).$$

Here we used the well-known fact that $|\{\underline{h} \in [H]^t, \Delta Q(\underline{h}) = 0\}| = O(H^{t-1})$ and if $1 \leq |\Delta Q(\underline{h})| \leq H^D$ then $\prod_{p|\Delta Q(\underline{h})} p \leq H^D$ thus the number of primes $p \geq 2^k$, $p|\Delta Q(\underline{h})$ is $O(D \log H/k)$. This proves (2.12) and Proposition 2.1 follows. $\square$

3. PET INDUCTION AND THE GENERALIZED VON NEUMANN INEQUALITY.

We describe the PET induction scheme originally devised by Bergelson and Leibman [1] for bounded function and later extended by Tao and Ziegler to functions bounded by a pseudo-random measure ν . We'll follow the notation of [16] except that in our multi-dimensional setting we will have l shift operators in the directions of the vector $\underline{v}_1, \dots, \underline{v}_l$, namely $T_j f(\underline{x}) := f(\underline{x} + \underline{v}_j)$ for $1 \leq j \leq l$. We will estimate averages of the form

$$(3.1) \quad \Lambda_{\mathcal{P}}(\vec{f}) = \mathbb{E}_{\underline{x} \in X} \mathbb{E}_{y \in [M]} \prod_{j=1}^d T_j^{P_j(y)} f_j(\underline{x}),$$

for a family of functions $\vec{f} = (f_1, \dots, f_l)$ satisfying $|f_j| \leq \nu^{(d)}$, using repeated applications of van der Courput's lemma [16] which in our context is the following simple observation. Let x_n be a sequence of real numbers satisfying $x_n = O(\log^C N)$ uniformly in n . Let $M = \log^L N$ and $H = \log^{\sqrt{L}} N$ with L sufficiently large with respect C . Then,

$$(3.2) \quad \mathbb{E}_{n \in [M]} x_n = \mathbb{E}_{h \in H} \mathbb{E}_{n \in [M]} x_{n+h} + o(1),$$

as shifting the average by a small amount h will change it only slightly. Then by the Cauchy-Schwarz inequality,

$$(3.3) \quad |\mathbb{E}_{n \in [M]} x_n|^2 \leq \mathbb{E}_{n \in [M]} \mathbb{E}_{h, h' \in [H]} x_{n+h} x_{n+h'} + o(1),$$

After each application we will arrive at a new system of polynomials some of which will have reduced degrees in the y -variable or applied to the measure $\nu^{(d)}$ in place of a function f_j . Eventually, one arrives at a system where each polynomial corresponding to a shift applied to one of the functions f_j is linear in y but may depend on the number of additional small parameters $h_1, \dots, h_t$. Averages over such systems are then majorized by polynomial averages of local box norms, described in the introduction.

Example. Let us first illustrate the procedure with an example. Let $d = 2$, $T_i f(\underline{x}) := f(\underline{x} + \underline{v}_i)$ for $i = 1, 2$.

$$\Lambda(\vec{f}) = \mathbb{E}_{\underline{x}} \mathbb{E}_y f_0 \cdot T_1^y f_1 \cdot T_2^{y^2} f_2,$$

where $\underline{x} \in X$, $y \in [M]$, and $|f_i| \leq \nu := \nu^{(2)}$. We assume that $\underline{v}_i$ has no zero coordinates so the system $\mathcal{P}(y) = (0, y\underline{v}_1, y^2\underline{v}_2)$ is in general position. Then,

$$|\Lambda(\vec{f})| \leq \mathbb{E}_{\underline{x}} \nu |\mathbb{E}_y T_1^y f_1 \cdot T_2^{y^2} f_2|,$$

hence by van der Courput's lemma

$$|\Lambda(\vec{f})|^2 \leq \mathbb{E}_{\underline{x}} \nu \mathbb{E}_{h, h'} \mathbb{E}_y T_1^{y+h} f_1 \cdot T_1^{y+h'} f_1 \cdot T_2^{(y+h)^2} f_2 \cdot T_2^{(y+h')^2} f_2 + o(1)$$

Now we shift the variables $\underline{x}$ to $\underline{x} - y\underline{v}_1$ which does not change the average but which amounts to multiplying each factor with T_1^{-y} , thus we get

$$|\Lambda(\vec{f})|^2 \leq \mathbb{E}_{\underline{x}} \mathbb{E}_{h, h'} T_1^h \nu \cdot T_1^{h'} \nu |\mathbb{E}_y T_1^{-y} \nu \cdot T_1^{-y} T_2^{(y+h)^2} f_2 \cdot T_1^{-y} T_2^{(y+h')^2} f_2| + o(1).$$

By the polynomial forms condition $\mathbb{E}_{\underline{x}} \mathbb{E}_{h,h'} T_1^h \nu \cdot T_1^{h'} \nu = 1 + o(1)$ hence after one more application of van der Courput's lemma,

$$|\Lambda(\vec{f})|^4 \leq \mathbb{E}_{\underline{x}} \mathbb{E}_{h,h'} \mathbb{E}_{k,k'} T_1^h \nu \cdot T_1^{h'} \nu \cdot \mathbb{E}_y T_1^{-y-k} \nu \cdot T_1^{-y-k'} \nu \cdot T_1^{-y-k} T_2^{(y+h+k)^2} f_2 \cdot T_1^{-y-k} T_2^{(y+h'+k)^2} f_2 \cdot T_1^{-y-k'} T_2^{(y+h+k')^2} f_2 \cdot T_1^{-y-k'} T_2^{(y+h'+k')^2} f_2$$

Note that there are four quadratic polynomials in y however each have the same main term y^2 thus after shifting the variables $\underline{x}$ to $\underline{x} - y^2 \underline{v}_2$, which amounts to multiplying each factor with $T_2^{-y^2}$, all exponents of the shift operators T_1, T_2 applied to the function f_2 will be linear.

Doing this procedure in general leads to polynomial averages of the form

$$(3.4) \quad \Lambda_{\mathcal{R}}(\vec{f}) := \mathbb{E}_{h_1, \dots, h_t} \mathbb{E}_{\underline{x}} \mathbb{E}_{y \in [M]} \prod_{\alpha \in \mathcal{A}} T^{\underline{R}_{\alpha}(y, h_1, \dots, h_t; W)} f_{\alpha}(\underline{x}),$$

where

$$(3.5) \quad \underline{R}_{\alpha}(y, h_1, \dots, h_t; W) = \sum_{j=1}^l R_{j, \alpha}(y, h_1, \dots, h_t; W) \underline{v}_j,$$

and

$$(3.6) \quad T^{\underline{R}_{\alpha}(y, h_1, \dots, h_t; W)} f_{\alpha}(\underline{x}) := \prod_{j=1}^l T_j^{R_{j, \alpha}(y, h_1, \dots, h_t; W)} f_{\alpha}(\underline{x}) \\ = f_{\alpha}(\underline{x} + \underline{R}_{\alpha}(y, h_1, \dots, h_t; W)),$$

using notation similar to [1, 16]. For simplicity of notation we will write $\underline{R}_{\alpha}(\underline{h}, y)$ suppressing the dependence on W . To describe the PET procedure in our settings we will need several definitions.

Let $\mathcal{R} := \{\underline{R}_{\alpha}(y, \underline{h}) : \alpha \in \mathcal{A}\}$ be a polynomial system with $\underline{R}_{\alpha} \in \mathbb{Z}^d[y, h_1, \dots, h_t]$ being an integral polynomial map for all $\alpha \in \mathcal{A}$. We will consider y as the primary variable of the polynomial maps $\underline{R}_{\alpha}(y, \underline{h})$ and $\underline{h} = (h_1, \dots, h_t)$ as parameters. We say that the y -degree of a polynomial map $\underline{R}(y, \underline{h})$ is equal to d and write $\deg_y \underline{R}(y, \underline{h}) = d$, if

$$\underline{R}(y, \underline{h}) = \underline{c}(\underline{h}) y^d + \underline{Q}(y, \underline{h}),$$

where $\underline{c}(\underline{h}) \neq \underline{0}$ and $\deg_y(\underline{Q}(y, \underline{h})) < d$, and $\underline{Q}(y, \underline{h}) = \underline{0}$ if $d = 0$. We extend the notion of *general position* to polynomial systems $\mathcal{R} := \{\underline{R}_{\alpha}(y, \underline{h}) : \alpha \in \mathcal{A}\}$ depending on parameters $\underline{h} = (h_1, \dots, h_t)$ as follows.

Definition 3.1. *We say that a polynomial system $\mathcal{R} := \{\underline{R}_{\alpha}(y, \underline{h}) : \alpha \in \mathcal{A}\}$ is in general position if for any two distinct nodes α, β , we have*

$$(3.7) \quad \deg_y(\pi_i(\underline{R}_{\alpha} - \underline{R}_{\beta})) = \deg_y(\underline{R}_{\alpha} - \underline{R}_{\beta}) \quad \text{for all } 1 \leq i \leq d.$$

Moreover, if $\deg_y(\underline{R}_{\alpha} - \underline{R}_{\beta}) = 0$ then $\pi_i(\underline{R}_{\alpha}) \neq \pi_i(\underline{R}_{\beta})$.

For given nodes $\alpha, \beta \in \mathcal{A}$ let $d_{\alpha\beta} := \deg_y(\underline{R}_{\alpha}(y, \underline{h}) - \underline{R}_{\beta}(y, \underline{h}))$ and write

$$(3.8) \quad \underline{R}_{\alpha}(y, \underline{h}) - \underline{R}_{\beta}(y, \underline{h}) = \underline{c}_{\alpha\beta}(\underline{h}) y^{d_{\alpha\beta}} + \underline{Q}_{\alpha\beta}(y, \underline{h}),$$

that is $\underline{c}_{\alpha\beta}(\underline{h}) \neq \underline{0}$ is the coefficient of leading term of the polynomial $\underline{R}_{\alpha}(y, \underline{h}) - \underline{R}_{\beta}(y, \underline{h})$ in the y -variable. If $d_{\alpha\beta} = 0$, then we set $\underline{Q}_{\alpha\beta} = \underline{0}$. Then (3.7) is equivalent to

$$(3.9) \quad \pi_i(\underline{c}_{\alpha\beta}(\underline{h})) \neq 0 \quad \text{for all } 1 \leq i \leq d, \quad \alpha, \beta \in \mathcal{A}, \quad \alpha \neq \beta.$$

An important note is that starting from an initial polynomial configuration $\mathcal{P}(y) = (P_1(y)\underline{v}_1, \dots, P_l(y)\underline{v}_l)$, all polynomial systems $\{\underline{R}_\alpha(\underline{h}, y) : \alpha \in \mathcal{A}\}$ will remain in general position during the PET procedure, i.e. they will satisfy (3.7) or equivalently (3.9). To formalize this observation, write $\mathcal{A} = \mathcal{A}_0 \cup \mathcal{A}_1$ where $\mathcal{A}_0 = \{\alpha : \deg_y \underline{R}_\alpha(y, \underline{h}) = 0\}$, and define the *doubling* of the system $\mathcal{R}$ as the polynomial system:

$$(3.10) \quad \mathcal{R}' = \{\underline{R}_\beta(\underline{h}) : \beta \in \mathcal{A}_0\} \cup \{\underline{R}_\beta(y + h^1, \underline{h}) : \beta \in \mathcal{A}_1\} \cup \{\underline{R}_\beta(y + h^2, \underline{h}) : \beta \in \mathcal{A}_1\}.$$

We index this system with the set of nodes $\mathcal{A}' := \mathcal{A}_0 \cup \mathcal{A}_1^1 \cup \mathcal{A}_1^2$, so that for nodes $\beta^1 \in \mathcal{A}_1^1$ and $\beta^2 \in \mathcal{A}_1^2$ there correspond the polynomials

$$(3.11) \quad \underline{R}_{\beta^1}(y, \underline{h}, h^1) = \underline{R}_\beta(y + h^1, \underline{h}), \quad \underline{R}_{\beta^2}(y, \underline{h}, h^2) = \underline{R}_\beta(y + h^2, \underline{h}).$$

Lemma 3.1. *Let $\mathcal{R} = \{\underline{R}_\alpha(\underline{h}, y) : \alpha \in \mathcal{A}\}$ be a polynomial system in general position. Then the its doubling, that is the system $\mathcal{R}'$ defined in (3.10)-(3.11) is also in general position.*

Proof. We may assume without loss of generality that $\underline{R}_{\alpha_0} = \underline{0}$ for some fixed node α_0 , as equation (3.7) is invariant under the translation $\underline{R}_\alpha \rightarrow \underline{R}_\alpha - \underline{R}_{\alpha_0}$ ($\alpha \in \mathcal{A}$). We write

$$\underline{R}_\alpha(y, \underline{h}) = \underline{c}_\alpha(\underline{h}) y^{d_\alpha} + \underline{Q}_\alpha(y, \underline{h}),$$

where $d_\alpha = \deg_y \underline{R}_\alpha(y, \underline{h})$ and $\underline{Q}_\alpha(y, \underline{h}) = 0$ if $d_\alpha = 0$, that is when $\alpha \in \mathcal{A}_0$. By (3.9), we have

$$(3.12) \quad \pi_i(\underline{c}_\alpha(\underline{h})) \neq 0, \quad \text{for all } 1 \leq i \leq d.$$

Let $\alpha, \beta \in \mathcal{A} \setminus \{0\}$ be two distinct nodes. If both nodes are in $\mathcal{A}_0$, then $\underline{R}'_\alpha = \underline{R}_\alpha$ and $\underline{R}'_\beta = \underline{R}_\beta$ so (3.9) clearly holds for $\underline{R}'_\alpha$ and $\underline{R}'_\beta$. If say $\alpha \in \mathcal{A}_1^\sigma$ with $\sigma = 1, 2$, but $\beta \in \mathcal{A}_0$ then $d_\alpha > 0$ and $d_\beta = 0$. Thus,

$$\underline{R}'_{\alpha^\sigma}(y, \underline{h}, h^\sigma) - \underline{R}'_\beta(y, \underline{h}) = \underline{R}_\alpha(y + h^\sigma, \underline{h}) - \underline{R}_\beta(\underline{h}) = \underline{c}_\alpha(\underline{h}) y^{d_\alpha} + \underline{Q}'_{\alpha\beta}(y, \underline{h}, h^\sigma),$$

with $\deg_y (\underline{Q}'_{\alpha\beta}(y, \underline{h}, h^\sigma)) < d_\alpha$ and (3.9) holds.

Assume now that α and β are both in $\mathcal{A}_1$, and consider the difference

$$(3.13) \quad \underline{R}'_{\alpha^\sigma}(y, \underline{h}, h^\sigma) - \underline{R}'_{\beta^\tau}(y, \underline{h}, h^\tau) = \underline{R}_\alpha(y + h^\sigma, \underline{h}) - \underline{R}_\beta(y + h^\tau, \underline{h}),$$

for $\sigma, \tau = 1, 2$. First we show that

$$\underline{R}'_{\alpha^\sigma}(y, \underline{h}, h^\sigma) - \underline{R}'_{\beta^\tau}(y, \underline{h}, h^\tau) = \underline{c}_{\alpha\beta}(\underline{h}) y^{d_{\alpha\beta}} + \underline{Q}'_{\alpha^\sigma\beta^\tau}(y, \underline{h}),$$

with $\deg_y (\underline{Q}'_{\alpha^\sigma\beta^\tau}(y, \underline{h})) < d_{\alpha\beta}$ in all cases, except when $d_\alpha = d_\beta$, $\underline{c}_\alpha(\underline{h}) = \underline{c}_\beta(\underline{h})$ and $\sigma \neq \tau$.

Indeed, if $d_\alpha \neq d_\beta$, say $d_\alpha > d_\beta$ then the main term in the y -variable of the expression in (3.13) is $\underline{c}_\alpha(\underline{h}) y^{d_\alpha} = \underline{c}_{\alpha\beta}(\underline{h}) y^{d_{\alpha\beta}}$, thus (3.9) holds.

If $d_\alpha = d_\beta$ and $\sigma = \tau$ then the expression in (3.13) takes the form

$$\begin{aligned} \underline{R}_\alpha(y + h^\sigma, \underline{h}) - \underline{R}_\beta(y + h^\sigma, \underline{h}) &= \underline{c}_{\alpha\beta}(\underline{h}) (y + h^\sigma)^{d_{\alpha\beta}} + \underline{Q}_{\alpha\beta}(y + h^\sigma, \underline{h}) \\ &= \underline{c}_{\alpha\beta}(\underline{h}) y^{d_{\alpha\beta}} + \underline{Q}'_{\alpha\beta}(y, \underline{h}, h^\sigma), \end{aligned}$$

where $\deg_y (\underline{Q}'_{\alpha\beta}(y, \underline{h}, h^\sigma)) < d_{\alpha\beta}$ and (3.9) holds again.

If $d_\alpha = d_\beta$, $\sigma \neq \tau$ and $\underline{c}_\alpha(\underline{h}) \neq \underline{c}_\beta(\underline{h})$ then it is easy to see from (3.8) that $\underline{c}_{\alpha\beta}(\underline{h}) = \underline{c}_\alpha(\underline{h}) - \underline{c}_\beta(\underline{h})$ and $d_{\alpha\beta} = d_\alpha = d_\beta$. In this case

$$\begin{aligned} \underline{R}_\alpha(y + h^\sigma, \underline{h}) - \underline{R}_\beta(y + h^\tau, \underline{h}) &= \underline{c}_\alpha(\underline{h}) (y + h^\sigma)^{d_\alpha} - \underline{c}_\beta(\underline{h}) (y + h^\tau)^{d_\alpha} + \underline{Q}'_{\alpha\beta}(y, \underline{h}, h^1, h^2) \\ &= (\underline{c}_\alpha(\underline{h}) - \underline{c}_\beta(\underline{h})) y^{d_\alpha} + \underline{Q}''_{\alpha\beta}(y, \underline{h}, h^1, h^2), \end{aligned}$$

where $\deg_y(\underline{Q}''_{\alpha\beta}(y, \underline{h}, h^1, h^2)) < d_\alpha$ and (3.9) holds.

Finally, if $d_\alpha = d_\beta$, $\underline{c}_\alpha(\underline{h}) = \underline{c}_\beta(\underline{h})$ but $\sigma \neq \tau$, then we use the expansion

$$\underline{R}_\alpha(y, \underline{h}) = \underline{c}_\alpha(\underline{h}) y^{d_\alpha} + \underline{e}_\alpha(\underline{h}) y^{d_\alpha-1} + \underline{Q}_\alpha(y, \underline{h}),$$

where $\deg_y(\underline{Q}_\alpha(y, \underline{h})) < d_\alpha - 1$ or $\underline{Q}_\alpha(y, \underline{h}) = 0$ if $d_\alpha = 1$. We have

$$\begin{aligned} \underline{R}_\alpha(y + h^\sigma, \underline{h}) - \underline{R}_\beta(y + h^\tau, \underline{h}) &= \\ &= \underline{c}_\alpha(\underline{h}) ((y + h^\sigma)^{d_\alpha} - (y + h^\tau)^{d_\alpha}) + \underline{e}_\alpha(\underline{h}) (y + h^\sigma)^{d_\alpha-1} - \underline{e}_\beta(\underline{h}) (y + h^\tau)^{d_\alpha-1} + \underline{Q}'_{\alpha\beta}(y, \underline{h}, h^1, h^2) \\ &= d_\alpha (\underline{c}_\alpha(\underline{h}) (h^\sigma - h^\tau) + \underline{e}_\alpha(\underline{h}) - \underline{e}_\beta(\underline{h})) y^{d_\alpha-1} + \underline{Q}''_{\alpha\beta}(y, \underline{h}, h^1, h^2), \end{aligned}$$

where $\deg_y(\underline{Q}''_{\alpha\beta}(y, \underline{h})) < d_\alpha - 1$ or $\underline{Q}''_{\alpha\beta}(y, \underline{h}) = 0$ if $d_\alpha = 1$. Applying the projection π_i to the main term in the above expression, we have for all $1 \leq i \leq d$,

$$(3.14) \quad d_\alpha \pi_i(\underline{c}_\alpha(\underline{h})) (h^\sigma - h^\tau) + \pi_i(\underline{e}_\alpha(\underline{h}) - \underline{e}_\beta(\underline{h})) \neq 0.$$

Indeed, by our assumption $\pi_i(\underline{c}_\alpha(\underline{h})) \neq 0$, hence the first term in (3.14) depends on the new parameters h^1 and h^2 while the second term does not. Thus (3.9) holds again for the polynomials $\underline{R}'_{\alpha\sigma}(y, \underline{h}, h^\sigma)$ and $\underline{R}'_{\beta\tau}(y, \underline{h}, h^\tau)$. $\square$

We will fix a distinguished node α_0 and we say that a node α is *non-linear* if $\deg_y(\underline{R}_\alpha - \underline{R}_{\alpha_0}) > 1$ and *linear* if $\deg_y(\underline{R}_\alpha - \underline{R}_{\alpha_0}) = 1$. We say that system $\mathcal{R}$ is in *general position w.r.t. α_0* , if it is in general position and if for any two *distinct linear nodes* α, β ,

$$(3.15) \quad \deg_y(\underline{R}_\alpha - \underline{R}_\beta) = 1.$$

Lemma 3.2. *Let $\mathcal{R}$ be a polynomial system with distinguished node α_0 and a nonlinear node α^* . Then the its doubling, that is the system $\mathcal{R}'$ defined in (3.10)-(3.11) is also in general position with respect to the distinguished node α_0^1 .*

Proof. By Lemma 3.1 it is enough to show that main terms of the linear nodes with respect to the node α_0^1 remain distinct in the new system $\mathcal{R}'$. Note that α_0^1 is in $\mathcal{A}_1^1$ and is a non-linear node. Let β, γ be linear nodes w.r.t. α_0 . Write

$$\underline{R}_\beta(y, \underline{h}) = \underline{R}_{\alpha_0}(y, \underline{h}) + \underline{c}_\beta(\underline{h})y + \underline{e}_\beta(\underline{h}) \quad \text{and similarly} \quad \underline{R}_\gamma(y, \underline{h}) = \underline{R}_{\alpha_0}(y, \underline{h}) + \underline{c}_\gamma(\underline{h})y + \underline{e}_\gamma(\underline{h}).$$

Clearly, β, γ both in $\mathcal{A}_1$. If $\beta \neq \gamma$ then $\underline{c}_\beta \neq \underline{c}_\gamma$ thus clearly $\underline{R}_\beta(y + h^\sigma, \underline{h}) - \underline{R}_\gamma(y + h^\sigma, \underline{h})$ is not constant in y for $\sigma = 1, 2$. Now consider

$$\underline{R}_\beta(y + h^1, \underline{h}) - \underline{R}_\gamma(y + h^2, \underline{h}) = \underline{R}_{\alpha_0}(y + h^1, \underline{h}) - \underline{R}_{\alpha_0}(y + h^2, \underline{h}) + (\underline{c}_\beta(\underline{h}) - \underline{c}_\gamma(\underline{h}))y + \underline{e}_{\beta\gamma}(\underline{h}).$$

Taking a derivative w.r.t the y -variable, we have

$$\partial_y(\underline{R}_\beta(y + h^1, \underline{h}) - \underline{R}_\gamma(y + h^2, \underline{h})) = \partial_y(\underline{R}_{\alpha_0}(y + h^1, \underline{h}) - \underline{R}_{\alpha_0}(y + h^2, \underline{h})) + (\underline{c}_\beta - \underline{c}_\gamma)(\underline{h}) \neq 0,$$

as the above expression must depend on both h^1 and h^2 as α_0 is a non-linear node. If $\beta = \gamma$ then we have again,

$$\partial_y(\underline{R}'_{\beta^1}(y + h^1, \underline{h}) - \underline{R}'_{\beta^2}(y + h^2, \underline{h})) = \partial_y(\underline{R}_{\alpha_0}(y + h^1, \underline{h}) - \underline{R}_{\alpha_0}(y + h^2, \underline{h})) \neq 0.$$

□

Following [1] we will assign a *weight matrix* $\mathfrak{W}_{\mathcal{R}} = (\mathfrak{w}_{j,k})$ to a polynomial system $\mathcal{R} = \{\underline{R}_{\alpha}(h, y) : \alpha \in \mathcal{A}\}$ as follows. Let $\mathcal{R}_j := (R_{j,\alpha} : \alpha \in \mathcal{A})$ and define $\overline{\mathcal{R}}_j$ to be the set of those polynomials $R_{j,\alpha}(h, y)$ which satisfy

$$(3.16) \quad \deg_y(R_{\alpha,j}) \geq 1, \text{ but } \deg_y(R_{j',\alpha}) = 0, \text{ for all } j < j' \leq l.$$

Let D denote the maximum degree in y of the polynomials $R_{j,\alpha}(y, h)$. For given $1 \leq j \leq l$ and $1 \leq k \leq D$ define the entry $\mathfrak{w}_{j,k}$ as the number of equivalence classes of the polynomials $R_{j,\alpha} \in \overline{\mathcal{R}}_j$ of y -degree k , where two polynomials $R_{j,\alpha}$ and $R_{j,\alpha'}$ are equivalent if they have identical leading terms in the y -variable. The *weight matrix* $\mathfrak{W}_{\mathcal{R}}(\alpha^*) = (\mathfrak{w}_{j,k}(\alpha^*))$ of the system $(\underline{R}_{\alpha} : \alpha \in \mathcal{A})$ with respect to a specific node α^* is defined as the weight matrix of the shifted system $\{\underline{R}_{\alpha} - \underline{R}_{\alpha^*} : \alpha \in \mathcal{A}\}$. Note that in the scalar case $l = 1$ (and $v_1 = 1$) this agrees with definition of the weight vector $(w_k(\alpha^*))_{1 \leq k \leq D}$ given in [16].

We will estimate averages of the form (3.4) corresponding to systems $(\mathcal{R}, \vec{f}) := \{(\underline{R}_{\alpha}, f_{\alpha}) : \alpha \in \mathcal{A}\}$ where $|f_{\alpha}| \leq \nu^{(d)}$ for all $\alpha \in \mathcal{A}$. We define a node α *inactive* if $f_{\alpha} = \nu^{(d)}$ and *active* otherwise. The *weight matrix* $\mathfrak{W}_{\mathcal{R}, \vec{f}}(\alpha^*)$ of the system $(\mathcal{R}, \vec{f})$ is defined as the weight matrix of the restricted polynomial system $(\underline{R}_{\alpha} - \underline{R}_{\alpha^*} : \alpha \in \mathcal{A}_1)$, where $\mathcal{A}_1$ denotes the set of active nodes.

We define the ordering of $l \times D$ weight matrices by reversed lexicographic ordering; we write $\mathfrak{W}' \prec \mathfrak{W}$ if there exist $1 \leq j \leq l \leq l$, $1 \leq k \leq D$ so that $\mathfrak{w}'_{j,k} < \mathfrak{w}_{j,k}$ but $\mathfrak{w}'_{j',k'} = \mathfrak{w}_{j',k'}$ for $j < j'$ or $j = j'$ and $k < k'$. We remark that any ordered chain of weight matrices $\mathfrak{W}_1 \succ \mathfrak{W}_2 \succ \dots$ terminates in finitely many steps.

The key proposition of the PET procedure, which is a straightforward multidimensional extension of Proposition 5.14 in [16], is the following.

Proposition 3.1. *Let $(\mathcal{R}, \vec{f}) = \{(\underline{R}_{\alpha}, f_{\alpha}) : \alpha \in \mathcal{A}\}$ be a polynomial system in general position with distinguished node α_0 . Assume the system $\mathcal{R}$ is in general position with respect to α_0 , and there is an active non-linear node α^* .*

Then there exists a polynomial system $(\mathcal{R}', \vec{f}') = \{(\underline{R}'_{\alpha'}, f'_{\alpha'}) : \alpha' \in \mathcal{A}'\}$ in general position with respect to a distinguished node α'_0 , and an active node α' such that,

$$(3.17) \quad \mathfrak{W}_{\mathcal{R}'}(\alpha') \prec \mathfrak{W}_{\mathcal{R}}(\alpha^*).$$

and

$$(3.18) \quad |\Lambda_{\mathcal{R}}(\vec{f})|^2 \ll |\Lambda_{\mathcal{R}'}(\vec{f}')| + o_{N \rightarrow \infty}(1),$$

Moreover $f'_{\alpha_0} = f_{\alpha_0}$, and for each $\alpha' \in \mathcal{A}'$ one has that $f'_{\alpha'} = f_{\alpha}$ for some $\alpha \in \mathcal{A}$, or $f'_{\alpha'} = \nu^{(d)}$.

Proof. Shifting the system $\mathcal{R}$ by $\underline{R}_{\alpha^*}$, which amounts to changing the polynomials $\underline{R}_{\alpha} \rightarrow \underline{R}_{\alpha} - \underline{R}_{\alpha^*}$, we may assume that $\underline{R}_{\alpha^*} = 0$, where α^* is a fixed non-linear node with respect to α_0 . Clearly the shifted system will remain in general position.

Write $\mathcal{A} = \mathcal{A}_0 \cup \mathcal{A}_1$, where $\mathcal{A}_0 = \{\beta \in \mathcal{A} : \deg_y(\underline{R}_\beta) = 0\}$ and $\mathcal{A}_1 := \mathcal{A} \setminus \mathcal{A}_0$. Accordingly, write

$$G(x, \underline{h}) := \prod_{\beta \in \mathcal{A}_0} T^{\underline{R}_\beta(\underline{h})} f_\beta(\underline{x})$$

and note that

$$|G(\underline{x}, \underline{h})| \leq H(\underline{x}, \underline{h}) := \prod_{\beta \in \mathcal{A}_0} T^{\underline{R}_\beta(\underline{h})} \nu_\beta^{(d)}(\underline{x}).$$

Then by the polynomial forms condition we have that,

$$(3.19) \quad \mathbb{E}_{\underline{h} \in [H]^t} \mathbb{E}_{\underline{x} \in X} H(\underline{x}, \underline{h}) = 1 + o(1).$$

Let

$$(3.20) \quad F(\underline{x}, y, \underline{h}) := \prod_{\beta \in \mathcal{A}_1} T^{\underline{R}_\beta(y, \underline{h})} f_\beta(\underline{x}),$$

then

$$(3.21) \quad |\Lambda_{\mathcal{P}}(\vec{f})| \leq \mathbb{E}_{\underline{h} \in [H]^t} \mathbb{E}_{\underline{x} \in X} H(\underline{x}, \underline{h}) |\mathbb{E}_{y \in [M]} F(\underline{x}, y, \underline{h})| \\ \leq \mathbb{E}_{\underline{h} \in [H]^t} \mathbb{E}_{\underline{x} \in X} H(\underline{x}, \underline{h}) \mathbb{E}_{y \in [M]} |\mathbb{E}_{h \in [H]} F(\underline{x}, y + h, \underline{h})| + o(1).$$

Here the $o(1)$ term is coming from shifting the variables $y \rightarrow y + h$ causing an error term $O((\log N)^{C_{\mathcal{R}}} H/M) = O((\log N)^{-\sqrt{L}/2}) = o(1)$ if L is chosen sufficiently large with respect to $\mathcal{R}$. Applying the Cauchy-Schwarz inequality in the h -variable, using (3.19), we get

$$(3.22) \quad |\Lambda_{\mathcal{P}}(\vec{f})|^2 \leq \mathbb{E}_{(\underline{h}, h^1, h^2) \in [H]^{t+2}} \mathbb{E}_{\underline{x} \in X} \mathbb{E}_{y \in [M]} H(\underline{x}, \underline{h}) F(\underline{x}, y + h^1, \underline{h}) F(\underline{x}, y + h^2, \underline{h}) + o(1) \\ := \Lambda_{\mathcal{R}'}(\vec{f}') + o(1),$$

where

$$(3.23) \quad \mathcal{R}' = \{\underline{R}_\beta(\underline{h}) : \beta \in \mathcal{A}_0\} \cup \{\underline{R}_\beta(y + h^1, \underline{h}) : \beta \in \mathcal{A}_1\} \cup \{\underline{R}_\beta(y + h^2, \underline{h}) : \beta \in \mathcal{A}_1\}.$$

We index this system with the set of nodes $\mathcal{A}' := \mathcal{A}_0 \cup \mathcal{A}_1^1 \cup \mathcal{A}_1^2$, so that for the nodes $\beta^1 \in \mathcal{A}_1^1$ and $\beta^2 \in \mathcal{A}_1^2$ we have the polynomials

$$\underline{R}_{\beta^1}(y, \underline{h}, h^1, h^2) = \underline{R}_\beta(y + h^1, \underline{h}), \quad \underline{R}_{\beta^2}(y, \underline{h}, h^1, h^2) = \underline{R}_\beta(y + h^2, \underline{h})$$

and functions

$$f'_{\beta^1}(\underline{x}) = f'_{\beta^2}(\underline{x}) = f_\beta(\underline{x}).$$

Note that $\mathcal{R}'$ is the doubling of the system $\mathcal{R}$ and hence by Lemma 3.1 and Lemma 3.2 it is in general position with respect to the distinguished node α_0^1 . The doubling $R_{j,\alpha}(y, \underline{h}) \rightarrow R_{j,\alpha}(y + h^\sigma, \underline{h})$ ($\sigma = 1, 2$) does not change the main terms of the polynomials $R_{j,\alpha}(y, \underline{h})$ in the y -variable, hence $\mathfrak{W}_{\mathcal{R}'} = \mathfrak{W}_{\mathcal{R}}$. In the above doubling procedure we do not activate inactive nodes but we may deactivate some active nodes, in fact those the active nodes α such that $\deg_y(\underline{R}_\alpha) = 0$. Thus $\mathfrak{W}_{\mathcal{R}', \vec{f}'} \prec \mathfrak{W}_{\mathcal{R}, \vec{f}}$ or $\mathfrak{W}_{\mathcal{R}', \vec{f}'} = \mathfrak{W}_{\mathcal{R}, \vec{f}}$.

Let $\mathfrak{w}_{j_0 k_0}$ be the first non-zero entry of the weight matrix $\mathfrak{W}_{\mathcal{R}, \vec{f}}$, such an entry must exist as α_0 is a nonlinear node. This means there is a polynomial $R_{j_0, \alpha} \in \mathcal{R}_{j_0}$ such that $\deg_y(R_{j_0, \alpha}) = k_0$ but $\overline{\mathcal{R}}_j = \emptyset$ for $j < j_0$ and $\deg_y(R_{j_0, \beta}) \geq k_0$ for all $\overline{\mathcal{R}}_{j_0, \beta} \in \mathcal{R}_{j_0}$.

Also $\deg_y(\mathcal{R}_{j,\alpha}) = 0$ for $j > j_0$ by the definition of the reduced class of polynomials $\overline{\mathcal{R}}_{j_0}$.

We claim that $\mathfrak{w}_{\mathcal{R}', \vec{f}, (\alpha^1)} \prec \mathfrak{w}_{\mathcal{R}', \vec{f}}$, and hence $\mathfrak{W}_{\mathcal{R}', \vec{f}, (\alpha^1)} \prec \mathfrak{W}_{\mathcal{R}', \vec{f}}$, where $\mathfrak{w}_{\mathcal{R}', \vec{f}, (\alpha^1)}$ is the weight matrix of the shifted system $\{\underline{R}'_{\alpha'} - \underline{R}'_{\alpha^1} : \alpha' \in \mathcal{A}'\}$. Indeed if $j > j_0$ then $\deg_y(\mathcal{R}'_{j,\alpha^1}) = 0$ thus we do not change the entries $\mathfrak{w}'_{j,k}$. If $k > k_0$ then again we do not change the entries $\mathfrak{w}'_{j_0,k}$ as they depend only on the main terms of polynomials in $\mathcal{R}'_{j_0}$ of degree k . However we reduce the entry $\mathfrak{w}'_{j_0,k_0}$ by 1 when we subtract the polynomial $R_{j_0,\alpha^1}(y, \underline{h})$ from the polynomials $R_{j_0,\alpha'}(y, \underline{h}) \in \overline{\mathcal{R}}_{j_0}$ as the equivalence class of $R_{j_0,\alpha^1}(y, \underline{h})$ vanishes. We may get new equivalence classes of smaller degrees than k_0 but that does not affect the ordering of the matrices. This proves (3.17) with and (3.18) follows from (3.22). $\square$

We have shown, after shifting the system so that $\underline{R}_{\alpha_0} = \underline{0}$,

$$(3.24) \quad |\Lambda(f)|^{2^s} \leq \mathbb{E}_{\underline{h}, y, \underline{x}} f_{\alpha_0}(\underline{x}) \prod_{\alpha \in A_{nl}} T^{\underline{R}_{\alpha}(\underline{h}, y)} \nu^{(d)}(\underline{x}) \times \prod_{\alpha \in A_l} T^{\underline{b}_{\alpha}(\underline{h})y + \underline{c}_{\alpha}(\underline{h})} f_{\alpha}(\underline{x}) + o(1),$$

where $\underline{b}_{\alpha}(\underline{h}), \underline{c}_{\alpha}(\underline{h})$ are integral polynomial maps dependent on shift parameters $\underline{h} = (h_1, \dots, h_t)$. We use A_l to denote the active nodes (l stands for linear), and we let A_{nl} denote the inactive nodes. Note that polynomial system $\mathcal{R} = \{\underline{R}_{\alpha}(y, \underline{h}), \alpha \in A_{nl}\} \cup \{\underline{R}_{\alpha}(y, \underline{h}) = \underline{b}_{\alpha}(\underline{h})y + \underline{c}_{\alpha}(\underline{h})\}$ is in general position in the sense of (3.7). In particular, $\pi_i(\underline{b}_{\alpha}) \neq 0$ for all $i \in [d]$ and $\alpha \in \mathcal{A}$.

We will show that the expression in (3.24) is bounded by

$$(3.25) \quad \leq \left(\mathbb{E}_{\underline{h}} \mathbb{E}_{\underline{x}, y^{(0)}, \underline{y}^{(1)}} \prod_{\omega \in \{0,1\}^{|A_l|}} T^{\sum_{\gamma \in A_l} y_{\gamma}^{(\omega_{\gamma})} \underline{b}_{\gamma}(\underline{h})} f_{\alpha_0}(\underline{x}) \right)^{2^{-|A_l|}} + o(1) \\ = \|f_{\alpha_0}\|_{\square_{H,M}(\underline{b}_{\gamma}: \gamma \in A_l)}$$

i.e. the average of the local Gowers box norms $\|f_{\alpha_0}\|_{\square(\underline{b}_{\gamma}(\underline{h}): \gamma \in A_l)} + o(1)$.

The main tool in this step is the weighted generalized von Neumann inequality [16, Appendix A]. To this end, we introduce new shift parameters $\underline{y} = (y_{\alpha})_{\alpha \in A_l}$ and define $\underline{Q}_0(\underline{y}) := \sum_{\alpha \in A_l} \underline{b}_{\alpha}(\underline{h}) y_{\alpha}$. Then, after shifting y by $-\sum_{\alpha \in A_l} y_{\alpha}$ and shifting the polynomial system by $\underline{Q}_0(\underline{y})$, the right-hand side of (3.24) becomes

$$(3.26) \quad \mathbb{E}_{\underline{h}} \mathbb{E}_{\underline{y}, y, \underline{x}} T^{\underline{Q}_0(\underline{y})} f_{\alpha_0}(\underline{x}) \prod_{\alpha \in A_{nl}} T^{\underline{Q}_0(\underline{y}) + \underline{R}_{\alpha}(\underline{h}, y - \sum_{\gamma \in A_l} y_{\gamma})} \nu^{(d)}(\underline{x}) \\ \times \prod_{\alpha \in A_l} T^{\underline{b}_{\alpha}(\underline{h})y + \sum_{\gamma \in A_l} (\underline{b}_{\gamma}(\underline{h}) - \underline{b}_{\alpha}(\underline{h}))y_{\gamma} + \underline{c}_{\alpha}(\underline{h})} f_{\alpha}(\underline{x}) + o(1).$$

Notice that since we simply shifted the system, it is still in general position as a system in the variables $\underline{h}, \underline{y}$ and hence will satisfy (2.7) and the polynomial forms condition (2.8) applies.

Let

$$f_{0, \underline{x}, \underline{h}, \underline{y}}(\underline{y}) = T^{\underline{Q}_0(\underline{y})} f_{\alpha_0}(\underline{x}) \prod_{\alpha \in A_{nl}} T^{\underline{Q}_0(\underline{y}) + \underline{R}_{\alpha}(\underline{h}, y - \sum_{\gamma \in A_l} y_{\gamma})} \nu^{(d)}(\underline{x}), \\ f_{\alpha, \underline{x}, \underline{h}, \underline{y}}(\underline{y}) = T^{\underline{b}_{\alpha}(\underline{h})y + \sum_{\gamma \in A_l} (\underline{b}_{\gamma}(\underline{h}) - \underline{b}_{\alpha}(\underline{h}))y_{\gamma} + \underline{c}_{\alpha}(\underline{h})} f_{\alpha}(\underline{x}), \\ \nu_{\alpha, \underline{x}, \underline{h}, \underline{y}}^{(d)}(\underline{y}) = T^{\underline{b}_{\alpha}(\underline{h})y + \sum_{\gamma \in A_l} (\underline{b}_{\gamma}(\underline{h}) - \underline{b}_{\alpha}(\underline{h}))y_{\gamma} + \underline{c}_{\alpha}(\underline{h})} \nu^{(d)}(\underline{x}).$$

With this in mind, (3.26) becomes

$$\mathbb{E}_{\underline{h}, \underline{y}, \underline{x}} \mathbb{E}_{\underline{y}} f_{0, \underline{x}, \underline{h}}(\underline{y}) \prod_{\alpha \in A_l} f_{\alpha, \underline{x}, \underline{h}}(\underline{y}).$$

By design, $0 \leq f_{\alpha, \underline{x}, \underline{h}, \underline{y}} \leq \nu_{\alpha, \underline{x}, \underline{h}, \underline{y}}^{(d)}$ pointwise and $f_{\alpha, \underline{x}, \underline{h}}(\underline{y})$ is independent of y_α for every $\alpha \in A_l$. Thus, we may apply the generalized weighted von Neumann theorem in the $\underline{y}$ -variable [16, Appendix A]:

$$(3.27) \quad \mathbb{E}_{\underline{h}, \underline{y}, \underline{x}} \mathbb{E}_{\underline{y}} f_{0, \underline{x}, \underline{h}, \underline{y}}(\underline{y}) \prod_{\alpha \in A_l} f_{\alpha, \underline{x}, \underline{h}}(\underline{y}) \leq \mathbb{E}_{\underline{h}, \underline{y}, \underline{x}} \|f_{0, \underline{x}, \underline{h}, \underline{y}}\|_{\square^{A_l}(\nu^{(d)})} \prod_{\alpha \in A_l} \|\nu_{\alpha, \underline{x}, \underline{h}, \underline{y}}^{(d)}\|_{\square^{A_l \setminus \{\alpha\}}}^{1/2}.$$

Thus by Hölder's inequality,

$$(3.28) \quad \mathbb{E}_{\underline{h}, \underline{y}, \underline{x}} \mathbb{E}_{\underline{y}} f_{0, \underline{x}, \underline{h}, \underline{y}}(\underline{y}) \prod_{\alpha \in A_l} f_{\alpha, \underline{x}, \underline{h}}(\underline{y}) \leq \left(\mathbb{E}_{\underline{h}, \underline{y}, \underline{x}} \|f_{0, \underline{x}, \underline{h}, \underline{y}}\|_{\square^{A_l}(\nu^{(d)})}^{2|A_l|} \right)^{2^{-|A_l|}},$$

as by the polynomial forms condition, we have for all $\alpha \in A_l$

$$\begin{aligned} \mathbb{E}_{\underline{h}, \underline{x}} \|\nu_{\alpha, \underline{x}, \underline{h}, \underline{y}}^{(d)}\|_{\square^{A_l \setminus \{\alpha\}}}^{2|A_l| - 1} &= \mathbb{E}_{\underline{h}, \underline{x}, \underline{y}} \mathbb{E}_{\underline{y}^{(0)}, \underline{y}^{(1)}} \prod_{\omega \in \{0,1\}^{A_l \setminus \{\alpha\}}} T^{\underline{b}_\alpha(\underline{h})\underline{y} + \sum_{\gamma \in A_l} (\underline{b}_\gamma(\underline{h}) - \underline{b}_\alpha(\underline{h})) y_\gamma^{(\omega_\gamma)} + \underline{c}_\alpha(\underline{h})} \nu^{(d)}(\underline{x}) \\ &= 1 + o(1). \end{aligned}$$

Note that,

$$(3.29) \quad \mathbb{E}_{\underline{h}, \underline{y}, \underline{x}} \|f_{0, \underline{x}, \underline{h}, \underline{y}}\|_{\square^{A_l}(\nu^{(d)})}^{2|A_l|} = \mathbb{E}_{\underline{h}, \underline{x}} \mathbb{E}_{\underline{y}^{(0)}, \underline{y}^{(1)}} \prod_{\omega \in \{0,1\}^{A_l}} T^{\underline{Q}_0(\underline{y}^{(\omega)})} f(\underline{x}) w(\underline{h}, \underline{x}, \underline{y}^{(0)}, \underline{y}^{(1)}),$$

where

$$\begin{aligned} w(\underline{h}, \underline{x}, \underline{y}^{(0)}, \underline{y}^{(1)}) &:= \mathbb{E}_{\underline{y}} \prod_{\alpha \in A_{nl}} \prod_{\omega \in \{0,1\}^{A_l}} T^{\underline{Q}_0(\underline{y}^{(\omega)}) + \underline{R}_\alpha(\underline{h}, \underline{y} - \sum_{\gamma \in A_l} y_\gamma^{(\omega_\gamma)})} \nu^{(d)}(\underline{x}) \\ (3.30) \quad &\times \prod_{\alpha \in A_l} \prod_{\omega^{(\alpha)} \in \{0,1\}^{A_l \setminus \{\alpha\}}} T^{\underline{b}_\alpha(\underline{h})\underline{y} + \sum_{\gamma \in A_l} (\underline{b}_\gamma(\underline{h}) - \underline{b}_\alpha(\underline{h})) y_\gamma^{(\omega_\gamma^{(\alpha)})} + \underline{c}_\alpha(\underline{h})} \nu^{(d)}(\underline{x}). \end{aligned}$$

Notice that the right-hand side of (3.29) is a weighted parallelogram average, and, ignoring the weight w , we end up with an average of local Gowers norms:

$$\begin{aligned} \mathbb{E}_{\underline{h}, \underline{x}} \mathbb{E}_{\underline{y}^{(0)}, \underline{y}^{(1)}} \mathbb{E}_{\omega \in \{0,1\}^{A_l}} \prod_{\omega \in \{0,1\}^{A_l}} T^{\underline{Q}_0(\underline{y}^{(\omega)})} f_{\alpha_0}(\underline{x}) &= \mathbb{E}_{\underline{h}, \underline{x}} \mathbb{E}_{\underline{y}^{(0)}, \underline{y}^{(1)}} \prod_{\omega \in \{0,1\}^{A_l}} T^{\sum_{\gamma \in A_l} y_\gamma^{(\omega_\gamma)} \underline{b}_\gamma(\underline{h})} f_{\alpha_0}(\underline{x}) \\ &= \mathbb{E}_{\underline{h}} \|f_{\alpha_0}\|_{\square(\underline{b}_\gamma(\underline{h}))}^{2|A_l|}. \end{aligned}$$

Thus, it suffices to prove

$$(3.31) \quad \mathbb{E}_{\underline{h}, \underline{x}, \underline{y}^{(0)}, \underline{y}^{(1)}} \prod_{\omega \in \{0,1\}^{A_l}} T^{\underline{Q}_0(\underline{y}^{(\omega)})} f_{\alpha_0}(\underline{x}) (w(\underline{h}, \underline{x}, \underline{y}^{(0)}, \underline{y}^{(1)}) - 1) = o(1)$$

By the Cauchy-Schwarz inequality and the bound $|f_{\alpha_0}| \leq \nu^{(d)}$, the square of the left-hand side of (3.31) is at most

$$\begin{aligned} & \mathbb{E}_{\underline{h}, \underline{x}, \underline{y}^{(0)}, \underline{y}^{(1)}} \prod_{\omega \in \{0,1\}^{A_l}} T_{\underline{Q}_0(\underline{y}^{(\omega)})}^{\nu^{(d)}}(\underline{x}) (w(\underline{h}, \underline{x}, \underline{y}^{(0)}, \underline{y}^{(1)}) - 1)^2 \\ &= \mathbb{E}_{\underline{h}, \underline{x}, \underline{y}^{(0)}, \underline{y}^{(1)}} \prod_{\omega \in \{0,1\}^{A_l}} T_{\underline{Q}_0(\underline{y}^{(\omega)})}^{\nu^{(d)}}(\underline{x}) (w(\underline{h}, \underline{x}, \underline{y}^{(0)}, \underline{y}^{(1)})^2 - 2w(\underline{h}, \underline{x}, \underline{y}^{(0)}, \underline{y}^{(1)}) + 1) \\ &= (1 + o(1)) - 2(1 + o(1)) + (1 + o(1)) = o(1), \end{aligned}$$

where we have repeatedly applied the polynomial forms condition to the coordinate projections of the polynomials appearing in (3.29)-(3.30), in the $\underline{h}, \underline{x}, \underline{y}^{(0)}, \underline{y}^{(1)}$ and y variables. To justify this, let $i \in [d]$ be fixed and consider the polynomials

$$(3.32) \quad x_i + \pi_i(\underline{Q}_0(\underline{y}^{(\omega)}, \underline{h})) = \sum_{\alpha \in \mathcal{A}_l} \pi_i(\underline{b}_\alpha(\underline{h})) y_\alpha^{(\omega_\alpha)},$$

for $\underline{\omega} \in \{0,1\}^{A_l}$. The polynomials,

$$(3.33) \quad x_i + \sum_{\alpha \in \mathcal{A}_l} \pi_i(\underline{b}_\alpha(\underline{h})) y_\alpha^{(\omega_\alpha)} + \pi_i(\underline{R}_{\alpha'}(\underline{h}, y - \sum_{\gamma \in A_l} y_\gamma^{(\omega_\gamma)})),$$

for $\alpha' \in \mathcal{A}_{nl}$ and $\underline{\omega} \in \{0,1\}^{A_l}$, and the polynomials

$$(3.34) \quad x_i + \pi_i(\underline{b}_\alpha(\underline{h})) y + \sum_{\gamma \in A_l} \pi_i((\underline{b}_\gamma - \underline{b}_\alpha)(\underline{h})) y_\gamma^{(\omega_\gamma^{(\alpha)})} + \pi_i(\underline{c}_\alpha(\underline{h})),$$

for $\alpha \in \mathcal{A}_l$, $\underline{\omega}^{(\alpha)} \in \{0,1\}^{A_l \setminus \{\alpha\}}$.

Since $\pi_i(\underline{b}_\alpha(\underline{h})) \neq 0$, the polynomials in (3.32) are distinct and satisfy (2.7). The polynomials in (3.33) are distinct for the same reason, and they are also distinct from the polynomials in (3.32) as either $\deg_y \underline{R}_{\alpha'}(\underline{h}, y) \geq 1$ hence they depend also on the y -variable, or $\pi_i(\underline{R}_{\alpha'}) \neq 0$ hence contain a term depending only on the variables $\underline{h}$. Finally the polynomials in (3.34) are distinct as $\pi_i((\underline{b}_\gamma - \underline{b}_\alpha)(\underline{h})) \neq 0$ for $\gamma \neq \alpha$, and they are distinct from the polynomials in (3.32)-(3.33), as they are independent of the $y_\alpha^{(0)}, y_\alpha^{(1)}$ variables but depend on the y -variable.

This shows the validity of (3.25). Since our initial family of functions is $f_1, \dots, f_l$ we have that $\alpha_0 = k$ and $f_{\alpha_0} = f_k$, and after re-indexing and renaming the variables, we may write

$$(\underline{b}_\gamma(\underline{h}) : \gamma \in \mathcal{A}_l) = (\underline{Q}_j^k(\underline{h}^k) : j \in [s_k])$$

thus we have for all $1 \leq k \leq l$,

$$(3.35) \quad |\Lambda_{\mathcal{P}, V}(\vec{f})| \leq \|f_k\|_{\square_{H, M}(\underline{Q}_j^k(\underline{h}^k) : j \in [s_k])}^{2^{-S_k}} + o(1).$$

Note that the norms on right side of (3.35) depend on k , however they can be majorized by a single local polynomial box norm using a simple concatenation property. Indeed one can define the concatenation of two polynomial systems $\underline{Q} = (\underline{Q}_1(\underline{h}), \dots, \underline{Q}_s(\underline{h}) : \underline{h} \in [H]^t)$ and $\underline{Q}' = (\underline{Q}'_1(\underline{h}'), \dots, \underline{Q}'_{s'}(\underline{h}') : \underline{h}' \in [H]^{t'})$ as $\underline{Q} \oplus \underline{Q}' = (\underline{Q}_1(\underline{h}), \dots, \underline{Q}_s(\underline{h}), \underline{Q}'_1(\underline{h}'), \dots, \underline{Q}'_{s'}(\underline{h}') : (h, h') \in [H]^{t+t'})$. Inductively one defines the concatenation $\underline{Q} = \underline{Q}^1 \oplus \dots \oplus \underline{Q}^l$ of more than two polynomial systems $\underline{Q}^1, \dots, \underline{Q}^l$. The following ‘‘concatenation lemma’’ follows from simple application of Hölder’s and the Cauchy-Schwarz inequality, see [16, Lemma A.3]

Lemma 3.3. *Let $\underline{Q} = \underline{Q}^1 \oplus \dots \oplus \underline{Q}^l$. Then one has for any $f : X \rightarrow \mathbb{R}$*

$$(3.36) \quad \|f\|_{\square_{H,M}(\underline{Q})} \geq \|f\|_{\square_{H,M}(\underline{Q}^k)},$$

for all $1 \leq k \leq l$.

Note that if $\underline{Q}^k$ is in general position and moreover if $\pi_i(Q_j^k) \neq 0$ for all $i \in [d]$, $k \in [l]$ and $j \in [s_k]$ then $\underline{Q}$ is also in general position. This holds in our case $\pi_i(b_\alpha) \neq 0$ for all $\alpha \in \mathcal{A}_l$. Since $|f|_k \leq \nu^{(d)}$ and $\|\nu^{(d)}\|_{\square_{H,M}(\underline{Q}^k)} = 1 + o(1)$ by the polynomial forms condition as the system $\underline{Q}^k$ is in general position. We have our key estimate by Lemma 3.3 and inequality (3.35).

Corollary 3.1. (generalized von-Neumann inequality)

Let $\mathcal{P} = (P_1(y)v_1, \dots, P_l(y)v_l)$ be a polynomial system in general position. Then there exist $t, s, S \geq 1$ and a polynomial system $\underline{Q} = (\underline{Q}_1, \dots, \underline{Q}_s)$ in general position, satisfying $\pi(\underline{Q}_j) \neq 0$, such that

$$(3.37) \quad |\Lambda_{\mathcal{P},V}(\vec{f})| \leq \min_{1 \leq k \leq l} \|f_k\|_{\square_{H,M}(\underline{Q}_1, \dots, \underline{Q}_s)}^{2^{-S}} + o(1).$$

4. DUAL FUNCTIONS AND THE MAIN DECOMPOSITION.

In this section we prove the crucial decomposition (1.6) following [17] which combines Gowers approach to decompositions theorems based on the Hahn-Banach theorem [7]. The starting point is to show the orthogonality of $\nu^{(d)} - 1$ to products of so-called *dual functions* associated to polynomial averages of local box norms.

Definition 4.1. Let $\underline{Q} = (\underline{Q}_1, \dots, \underline{Q}_s) : [H]^t \rightarrow \mathbb{Z}^{sd}$ be a polynomial map where $\underline{Q}_j \in \mathbb{Z}^d[h_1, \dots, h_t]$. Let $f_{\underline{\omega}} : X \rightarrow \mathbb{R}$, $\underline{\omega} \in \{0, 1\}^s \setminus \{0\}$ be a family of functions. We define the dual function of the family of functions $(f_{\underline{\omega}})_{\underline{\omega} \in \{0, 1\}^s \setminus \{0\}}$, as

$$(4.1) \quad D_{\underline{Q}}(f_{\underline{\omega}})(\underline{x}) := \mathbb{E}_{\underline{h} \in [H]^t} \mathbb{E}_{\underline{y}^{(0)}, \underline{y}^{(1)} \in [M]^s} \prod_{\underline{\omega} \in \{0, 1\}^s, \underline{\omega} \neq 0} f_{\underline{\omega}}(\underline{x} + y_1^{(\omega_1)} \underline{Q}_1(\underline{h}) + \dots + y_s^{(\omega_s)} \underline{Q}_s(\underline{h})).$$

If $f_{\underline{\omega}} = f$ for all $\underline{\omega} \in \{0, 1\}^s \setminus \{0\}$ then we write $D_{\underline{Q}}(f)$ and refer to $D_{\underline{Q}}(f)$ as the dual function of the function f .

This means that for fixed $\underline{h}$ and fixed $\underline{y}^{(0)} = (y_1^{(0)}, \dots, y_s^{(0)})$, $\underline{y}^{(1)} = (y_1^{(1)}, \dots, y_s^{(1)})$ we take the product of the functions $f_{\underline{\omega}}$ evaluated at the vertices $\underline{x} + y_1^{(\omega_1)} \underline{Q}_1(\underline{h}) + \dots + y_s^{(\omega_s)} \underline{Q}_s(\underline{h})$ forming a parallelepiped as $\underline{\omega} = (\omega_1, \dots, \omega_s)$ runs through all $\{0, 1\}^s \setminus \{0\}$, where the vectors $\underline{Q}_1(\underline{h}), \dots, \underline{Q}_s(\underline{h})$ give the directions of the side vectors of the parallelepiped, and then we sum the products for all such parallelepipeds as $\underline{y}^{(0)}, \underline{y}^{(1)}$ are running through $[M]^s$ and $\underline{h}$ through $[H]^t$. The terminology *dual function* is originated in [8] for global Gowers norms, and is due to the fact that the inner product,

$$(4.2) \quad \langle f, D_{\underline{Q}}(f) \rangle := \mathbb{E}_{\underline{x} \in X} f(\underline{x}) D_{\underline{Q}}(f)(\underline{x}) = \mathbb{E}_{\underline{h} \in [H]^t} \|f\|_{\square_M(\underline{Q}_1(\underline{h}), \dots, \underline{Q}_s(\underline{h}))}^{2^s}.$$

The Gowers type local box-norm inner product of a family of functions $(f_{\underline{\omega}})_{\underline{\omega} \in \{0, 1\}^s}$ of scale M and directions $\underline{u}_1, \dots, \underline{u}_s$, is defined as

$$(4.3) \quad \langle f_{\underline{\omega}} \rangle_{\square_M(\underline{u}_1, \dots, \underline{u}_s)} := \mathbb{E}_{\underline{x} \in X} \mathbb{E}_{\underline{y}^{(0)}, \underline{y}^{(1)} \in [M]^s} \prod_{\underline{\omega} \in \{0, 1\}^s} f_{\underline{\omega}}(\underline{x} + y_1^{(\omega_1)} \underline{u}_1 + \dots + y_s^{(\omega_s)} \underline{u}_s).$$

Let us recall a basic inequality referred to as the Gowers-Cauchy-Schwarz inequality for box norms that is well-known in various forms see e.g. [16, Appendix B].

Lemma 4.1. *One has*

$$(4.4) \quad |\langle f_{\underline{\omega}} \rangle_{\square_M(\underline{u}_1, \dots, \underline{u}_s)}| \leq \prod_{\underline{\omega} \in \{0, 1\}^s} \|f_{\underline{\omega}}\|_{\square_M(\underline{u}_1, \dots, \underline{u}_s)}$$

Proof. Let us write.

$$f_{\underline{\omega}, \underline{x}}(y_1, \dots, y_s) := f_{\underline{\omega}}(\underline{x} + y_1 \underline{u}_1 + \dots + y_s \underline{u}_s).$$

For a fixed $\underline{x} \in X$, we show

$$(4.5) \quad \left| \mathbb{E}_{\underline{y}^{(0)}, \underline{y}^{(1)} \in [M]^s} \prod_{\underline{\omega} \in \{0,1\}^s} f_{\underline{\omega}, \underline{x}}(y_1^{(\omega_1)}, \dots, y_s^{(\omega_s)}) \right| \leq \prod_{\underline{\omega} \in \{0,1\}^s} \|f_{\underline{\omega}, \underline{x}}\|_{\square_M},$$

where for a function $f : X \rightarrow \mathbb{R}$,

$$\|f\|_{\square_M}^{2^s} = \mathbb{E}_{\underline{y}^{(0)}, \underline{y}^{(1)} \in [M]^s} \prod_{\underline{\omega} \in \{0,1\}^s} f(y_1^{(\omega_1)}, \dots, y_s^{(\omega_s)}).$$

This is easy to see by repeated application of the Cauchy-Schwarz inequality. Indeed, separating the last variables and writing $\underline{y}^{(0)} = (\underline{y}^{(0)'}, y_s^{(0)})$ $\underline{y}^{(1)} = (\underline{y}^{(1)'}, y_s^{(1)})$, one may write

$$\langle f_{\omega} \rangle_{\square_M(\underline{u}_1, \dots, \underline{u}_s)} = \mathbb{E}_{\underline{y}^{(0)'}, \underline{y}^{(1)'}} \left(\mathbb{E}_{\underline{y}_s^{(0)}} \prod_{\underline{\omega}' \in \{0,1\}^{s-1}} f_{(\underline{\omega}', 0)}(y_1^{(\omega_1')}, \dots, y_{s-1}^{(\omega_{s-1}')}, y_s^{(0)}) \right) \left(\mathbb{E}_{\underline{y}_s^{(1)}} \prod_{\underline{\omega}' \in \{0,1\}^{s-1}} f_{(\underline{\omega}', 1)}(y_1^{(\omega_1')}, \dots, y_{s-1}^{(\omega_{s-1}')}, y_s^{(1)}) \right)$$

Then by the Cauchy-Schwarz inequality

$$\begin{aligned} \langle f_{\omega} \rangle_{\square_M(\underline{u}_1, \dots, \underline{u}_s)}^2 &\leq \\ &\left(\mathbb{E}_{\underline{y}^{(0)'}, \underline{y}^{(1)'}} \mathbb{E}_{\underline{y}_s^{(0)}, \underline{y}_s^{(1)}} \prod_{\underline{\omega} \in \{0,1\}^s} f_{(\underline{\omega}', 0)}(y_1^{(\omega_1')}, \dots, y_{s-1}^{(\omega_{s-1}')}, y_s^{(0)}) f_{(\underline{\omega}', 0)}(y_1^{(\omega_1')}, \dots, y_{s-1}^{(\omega_{s-1}')}, y_s^{(1)}) \right) \\ &\left(\mathbb{E}_{\underline{y}^{(0)'}, \underline{y}^{(1)'}} \mathbb{E}_{\underline{y}_s^{(0)}, \underline{y}_s^{(1)}} \prod_{\underline{\omega} \in \{0,1\}^s} f_{(\underline{\omega}', 1)}(y_1^{(\omega_1')}, \dots, y_{s-1}^{(\omega_{s-1}')}, y_s^{(0)}) f_{(\underline{\omega}', 1)}(y_1^{(\omega_1')}, \dots, y_{s-1}^{(\omega_{s-1}')}, y_s^{(1)}) \right) \end{aligned}$$

Notice that in each factors the functions $f_{\underline{\omega}}$ depend only on the first $s-1$ components of $\underline{\omega}$. Repeating the procedure in the other $s-1$ variables give (4.5). Then by Hölder's inequality left side of (4.3) is estimated by

$$\begin{aligned} \langle f_{\omega} \rangle_{\square_M(\underline{u}_1, \dots, \underline{u}_s)}^{2^s} &= \left(\mathbb{E}_{\underline{x}} \mathbb{E}_{\underline{y}^{(0)}, \underline{y}^{(1)} \in [M]^s} \prod_{\underline{\omega} \in \{0,1\}^s} f_{\underline{\omega}, \underline{x}}(y_1^{(\omega_1)}, \dots, y_s^{(\omega_s)}) \right)^{2^s} \\ &\leq \left(\mathbb{E}_{\underline{x}} \prod_{\underline{\omega} \in \{0,1\}^s} \|f_{\underline{\omega}, \underline{x}}\|_{\square_M} \right)^{2^s} \leq \prod_{\underline{\omega} \in \{0,1\}^s} \left(\mathbb{E}_{\underline{x}} \|f_{\underline{\omega}, \underline{x}}\|_{\square_M} \right)^{2^s} = \prod_{\underline{\omega} \in \{0,1\}^s} \|f_{\omega}\|_{\square_M(\underline{u}_1, \dots, \underline{u}_s)}^{2^s}. \end{aligned}$$

□

The following orthogonality property is crucial for the decomposition (1.6).

Proposition 4.1. *Let $K, s \geq 1$ and for $1 \leq k \leq K$ let $f_{\underline{\omega}}^k : X \rightarrow [-1, 1]$ be given for $\underline{\omega} \in \{0, 1\}^s \setminus \{\underline{0}\}$. Let $\underline{\mathcal{Q}} = (\underline{Q}_0, \underline{Q}_1, \dots, \underline{Q}_s)$, $\underline{Q}_0 = \underline{0}$, $\underline{Q}_j \in \mathbb{Z}^d[h]$ be a family of polynomial maps in general position, satisfying condition (2.7) for all $1 \leq i \leq d$ and $1 \leq j < j' \leq s$. Then one has*

$$(4.6) \quad \langle (\nu^{(d)} - 1), \prod_{k=1}^K D_{\underline{\mathcal{Q}}}(f_{\underline{\omega}}^k) \rangle := \mathbb{E}_{\underline{x} \in X} (\nu^{(d)}(\underline{x}) - 1) \cdot \prod_{k=1}^K D_{\underline{\mathcal{Q}}}(f_{\underline{\omega}}^k)(\underline{x}) = o(1),$$

where the implicit constant may depend on K, s and the polynomial map $\underline{\mathcal{Q}}$,

Proof. By expanding the product of the dual functions, we have

$$(4.7) \quad \prod_{k=1}^K D_{\underline{\mathcal{Q}}}(f_{\underline{\omega}}^k)(\underline{x}) = \mathbb{E}_{\underline{h}_1, \dots, \underline{h}_K} \mathbb{E}_{\substack{\underline{y}_1^{(0)}, \dots, \underline{y}_K^{(0)} \\ \underline{y}_1^{(1)}, \dots, \underline{y}_K^{(1)}}} \prod_{k=1}^K \prod_{\underline{\omega} \in \{0,1\}^s}^* f_{\underline{\omega}}^k(\underline{x} + y_{k1}^{(\omega_1)} \underline{Q}_1(\underline{h}_k) + \dots + y_{ks}^{(\omega_s)} \underline{Q}_s(\underline{h}_k)),$$

where we have written $\prod_{\omega}^*$ for the restricted product $\prod_{\omega \in \{0,1\}^s, \omega \neq \underline{0}}$ for simplicity of notations.

Let $B_{K,t}(H)$ be the set of those k -tuples of vectors $(\underline{h}_1, \dots, \underline{h}_K) \in [H]^{kt}$ such that $\pi_i(Q_j(\underline{h}_k)) = 0$ for some $i \in [d]$, $j \in [s]$ and $k \in [K]$. Then $|B_{K,t}(H)| \ll H^{kt-1}$ by the Schwarz-Zippel lemma, see [16, Lemma D.3], since $\pi_i(Q_j(\underline{h}))$ is not zero by our assumption. Thus the contribution of the K -tuples that are in $B_{K,t}(H)$ to the right side of (4.4) is $O(H^{-1}) = o(1)$.

Let $(\underline{h}_1, \dots, \underline{h}_K) \notin B_{K,t}(H)$ be a fixed K -tuple. We estimate

$$(4.8) \quad \mathbb{E}_{\underline{x}}(\nu^{(d)}(\underline{x}) - 1) \mathbb{E}_{\substack{\underline{y}_1^{(0)}, \dots, \underline{y}_K^{(0)} \\ \underline{y}_1^{(1)}, \dots, \underline{y}_K^{(1)}}} \prod_{k=1}^K \prod_{\omega \in \{0,1\}^s}^* f_{\omega}^k(\underline{x} + y_{k1}^{(\omega_1)} \underline{Q}_1(\underline{h}_k) + \dots + y_{ks}^{(\omega_s)} \underline{Q}_s(\underline{h}_k)).$$

Note that the product is taken over the union of K -distinct parallelepipeds with vertices $\underline{x} + y_{k1}^{(\omega_1)} \underline{Q}_1(\underline{h}_k) + \dots + y_{ks}^{(\omega_s)} \underline{Q}_s(\underline{h}_k)$. We may embed these K parallelepipeds into a Ks -dimensional parallelepiped with vertices of the form

$$\underline{x} + \sum_{j=1}^s \sum_{k=1}^K y_{k1}^{(\omega_{k1})} \underline{Q}_1(\underline{h}_k) + \dots + y_{ks}^{(\omega_{ks})} \underline{Q}_s(\underline{h}_k),$$

corresponding to $\underline{\Omega} = (\omega_1, \dots, \omega_K)$, $\omega_k = (\omega_{k1}, \dots, \omega_{ks}) \in \{0,1\}^s$. Define $f_{\underline{\Omega}} = 1$ unless $\underline{\Omega} = (0, \dots, 0, \underline{\omega}, 0, \dots, 0)$, i.e. when $\omega_j = \underline{0}$ for all $j \neq k$ and $\omega_k = \underline{\omega}$; in which case let $f_{\underline{\Omega}} = f_{\underline{\omega}}^k$. In other words, to all the remaining vertices of Ks -dimensional parallelepiped we attach weight 1. Then the expression in (4.8) may be written as

$$(4.9) \quad \mathbb{E}_{\underline{x}}(\nu^{(d)}(\underline{x}) - 1) \mathbb{E}_{\substack{\underline{y}_1^{(0)}, \dots, \underline{y}_K^{(0)} \\ \underline{y}_1^{(1)}, \dots, \underline{y}_K^{(1)}}} \prod_{\underline{\Omega} \in \{0,1\}^{Ks}} f_{\underline{\Omega}}(\underline{x} + \sum_{k=1}^K \sum_{j=1}^s y_{kj}^{(\omega_{kj})} \underline{Q}_j(\underline{h}_k)) \\ = \langle \nu^{(d)} - 1, f_{\underline{\Omega}} \rangle_{\square_M(\underline{Q}_j(\underline{h}_k): j \in [s], k \in [K])}.$$

By the Gowers-Cauchy-Schwarz inequality for box norms we have that

$$(4.10) \quad |\langle \nu^{(d)} - 1, f_{\underline{\Omega}} \rangle_{\square_M(\underline{Q}_j(\underline{h}_k): j \in [s], k \in [K])}| \leq \|\nu^{(d)} - 1\|_{\square_M(\underline{Q}_j(\underline{h}_k): j \in [s], k \in [K])},$$

as $|f_{\underline{\Omega}}| \leq 1$ and hence $\|f_{\underline{\Omega}}\|_{\square_M(\underline{Q}_j(\underline{h}_k): j \in [s], k \in [K])} \leq 1$ for all $\underline{\Omega} \in \{0,1\}^{Ks}$. Thus by Hölder's inequality the left side of (4.7) is estimated by

$$\mathbb{E}_{\underline{h}_1, \dots, \underline{h}_K} \|\nu^{(d)} - 1\|_{\square_M(\underline{Q}_j(\underline{h}_k): j \in [s], k \in [K])} \leq \left(\mathbb{E}_{\underline{h}_1, \dots, \underline{h}_K} \|\nu^{(d)} - 1\|_{\square_M(\underline{Q}_j(\underline{h}_k): j \in [s], k \in [K])}^{2Ks} \right)^{\frac{1}{2Ks}} = o(1),$$

by the polynomial forms condition (2.8) applied in the variables $\underline{x}_i, y_{kj}^{(\sigma)}, \underline{h}_k$ ($i \in [d], j \in [s], k \in [K]$) to the family of polynomials $x_i + \sum_{k=1}^K \sum_{j=1}^s y_{kj}^{(\omega_{kj})} \pi_i(Q_j(\underline{h}_k))$. $\square$

Using the orthogonality property (4.6) one can obtain the crucial decomposition (1.6) of an unbounded function $0 \leq f \leq \nu^{(d)}$ via an abstract decomposition theorem which appears implicitly in [8] and is stated in various essentially explicit forms in [14, 7, 17]. We recall the statement from [17] and for the sake of completeness provide a proof, which is a slight modification of an argument of Gowers [7, Theorem 4.8].

Theorem D. (*Dense model theorem* [17])

Let $\varepsilon > 0$ and let $\mathcal{F}$ be a set of bounded functions $F : X \rightarrow [-1, 1]$. Then there exists a constant $C > 0$ and a constant $\eta > 0$ depending only on ε such that the following holds; If a function $\nu : X \rightarrow \mathbb{R}_{\geq 0}$ obeys the bound

$$(4.11) \quad \mathbb{E}_{\underline{x} \in X} (\nu(\underline{x}) - 1) F_1 \cdots F_K(x) \leq \eta,$$

for all $K \leq \varepsilon^{-C}$ and $F_j \in \mathcal{F}$, then to every function $0 \leq f \leq \nu$ there exists a function $0 \leq g \leq 2$ such that

$$(4.12) \quad |\mathbb{E}_{\underline{x} \in X} (f(\underline{x}) - g(\underline{x})) F(\underline{x})| \leq \varepsilon,$$

for all $F \in \mathcal{F}$.

This means that if the function $\nu - 1$ is approximately orthogonal to long products of functions $F_i \in \mathcal{F}$ then every function $0 \leq f \leq \nu$ can be approximated by a bounded function g so that $f - g$ is approximately orthogonal to all function $F \in \mathcal{F}$. If $1 \in \mathcal{F}$ (as in most cases of interest) then f and g has approximately the same average, and g is referred to as a dense model of f for the family of functions $\mathcal{F}$.

Proof. Let $\epsilon > 0$, and define $J : \mathbb{R} \rightarrow \mathbb{R}$ by $J(x) = \frac{x+|x|}{2}$. By the Weierstrass approximation theorem, there exists a polynomial $P(x) = \sum_{i=0}^j a_i x^i$ such that

$$(4.13) \quad |P(x) - J(x)| < \frac{1}{8}$$

for all $x \in [-2\varepsilon^{-1}, 2\varepsilon^{-1}]$. Let C be large enough that $\varepsilon^{-C} > j$, and let $\eta = \frac{1}{4} \left(\sum_{i=0}^j |a_i| (2\varepsilon^{-1})^i \right)^{-1}$. Let $\nu : X \rightarrow \mathbb{R}_{\geq 0}$ be a function satisfying (4.11), and suppose for the sake of contradiction there exists a function $0 \leq f \leq \nu$ which cannot be written as $g + h$ where $0 \leq g \leq 2$ and

$$\sup_{F \in \mathcal{F}} |\mathbb{E}_{\underline{x} \in X} h(\underline{x}) F(\underline{x})| \leq \varepsilon.$$

By the Hahn-Banach theorem (specifically [7, Corollary 3.2]), there exists $\phi : X \rightarrow \mathbb{R}$ such that $\langle f, \phi \rangle > 1$, $\langle g, \phi \rangle < 1/2$ for every $0 \leq g \leq 1$, and $\langle h, \phi \rangle \leq \varepsilon^{-1}$ for every $h : X \rightarrow \mathbb{R}$ such that

$$\sup_{F \in \mathcal{F}} |\mathbb{E}_{\underline{x} \in X} h(\underline{x}) F(\underline{x})| \leq 1.$$

The above condition forces ϕ to be in the subspace spanned by the functions $F \in \mathcal{F}$, and then by [7, Corollary 3.5], we may write $\phi = \sum_{i=1}^s \lambda_i F_i$ where $F_1, \dots, F_s \in \mathcal{F}$ and

$$(4.14) \quad \sum_{i=1}^s |\lambda_i| < 2\varepsilon^{-1}.$$

By (4.11) and some manipulation, we have

$$(4.15) \quad \begin{aligned} |\langle \nu - 1, P\phi \rangle| &\leq \sum_{i=0}^j |a_i| |\langle \nu - 1, \left(\sum_{k=1}^s \lambda_k F_k \right)^i \rangle| \\ &\leq \eta \sum_{i=0}^j |a_i| \left(\sum_{k=1}^s |\lambda_k| \right)^i \\ &\leq \eta \sum_{i=0}^j |a_i| (2\varepsilon^{-1})^i \\ &= 1/4. \end{aligned}$$

Since $\langle g, \phi \rangle \leq 1/2$ for every $0 \leq g \leq 1$, we have $\langle 1, J\phi \rangle \leq 1/2$. By (4.14), it is clear that $|\phi| < 2\varepsilon^{-1}$ pointwise, and so by (4.13), $\|P\phi - J\phi\|_\infty \leq 1/8$. Combining these facts, we have $\langle 1, P\phi \rangle \leq 5/8$. Hence, by (??), $\langle \nu, P\phi \rangle \leq 7/8$. Finally, once again using the fact that $\|P\phi - J\phi\|_\infty \leq 1/8$, we have $\langle \nu, J\phi \rangle \leq 1$. But then

$$\langle f, \phi \rangle \leq \langle f, J\phi \rangle \leq \langle \nu, J\phi \rangle \leq 1,$$

contradicting the fact that $\langle f, \phi \rangle > 1$. □

Applying Proposition 4.1 and Theorem D to the family $\mathcal{F} := \{D_{\underline{Q}}(f_{\underline{\omega}})_{\underline{\omega} \in \{0,1\}^s \setminus \{0\}}, f_{\underline{\omega}} : X \rightarrow [-1, 1]\}$ where $\underline{Q} = (Q_0, Q_1, \dots, Q_s)$ is a family of integral polynomials, we have

Corollary 4.1. *Let $s, t \geq 1$ and let $\underline{Q} = (Q_0, Q_1, \dots, Q_s)$, $Q_0 = 0$, $Q_j \in \mathbb{Z}^d[h_1, \dots, h_t]$ be a family of polynomials in general position. Given $\varepsilon > 0$, if $N \geq N(\underline{Q}, t, s, \varepsilon)$ are sufficiently large then to every function $0 \leq f \leq \nu^{(d)}$ there exists a function $0 \leq g \leq 2$ s.t.*

$$(4.16) \quad |\mathbb{E}_{x \in X} (f(x) - g(x)) D_{\underline{Q}}(f_{\underline{\omega}})(x)| \leq \varepsilon,$$

for all families of functions $(f_{\underline{\omega}})_{\underline{\omega} \in \{0,1\}^s \setminus \{0\}}$, $f_{\underline{\omega}} : X \rightarrow [-1, 1]$.

At this point, for the function $0 \leq f_A \leq \nu^{(d)}$ we may obtain a decomposition $f_A = g + \mathfrak{h}$ with $0 \leq g \leq 2$ and the function $\mathfrak{h}$ satisfying

$$(4.17) \quad |\mathbb{E}_{x \in X} \mathfrak{h}(x) D_{\underline{Q}}(f_{\underline{\omega}})(x)| \leq \varepsilon,$$

for all dual functions $D_{\underline{Q}}(f_{\underline{\omega}})$ corresponding to bounded functions $f_{\underline{\omega}} : X \rightarrow [-1, 1]$. Note that $\mathfrak{h}$ may not be bounded independently of N but $|\mathfrak{h}| \leq \nu^{(d)} + 2$. If (4.12) would hold for $D_{\underline{Q}}(\mathfrak{h})$ i.e. when $f_{\underline{\omega}} = \mathfrak{h}$ for all $\underline{\omega} \in \{0, 1\}^s \setminus \{0\}$, then we would have

$$(4.18) \quad \|\mathfrak{h}\|_{\square_{H,M}(\underline{Q}_1, \dots, \underline{Q}_s)} \leq \varepsilon.$$

Hence by (1.5) and multi-linearity

$$(4.19) \quad \Lambda_{\mathcal{P},V}(f_A, \dots, f_A) = \Lambda_{\mathcal{P},V}(g, \dots, g) + O(\varepsilon^c) + o(1),$$

where $c > 0$ is a constant depending on the initial data $\mathcal{P}$ and V . The Bergelson-Liebman theorem [1] implies (see next section),

$$(4.20) \quad \Lambda_{\mathcal{P},V}(g, \dots, g) \geq c(\delta) + o(1),$$

where $c(\delta) > 0$ is a constant depending on the initial data but is independent of N . Thus choosing $\varepsilon > 0$ sufficiently small we have that $\Lambda_{\mathcal{P},V}(f_A, \dots, f_A) > 0$ which implies our main result.

The idea that the bounded functions $f_{\underline{\omega}}$ can be replaced by the unbounded function $\mathfrak{h}$ appeared first in [2]. The proof is essentially the same as that of [17, Theorem 11] and include it here only for the sake of completeness. Under the same conditions as in Proposition 4.1, we have

Proposition 4.2. *Let $\varepsilon > 0$ and let $\mathfrak{h} : X \rightarrow \mathbb{R}$ s.t. $|\mathfrak{h}| \leq \nu^{(d)} + 2$. If*

$$(4.21) \quad |\mathbb{E}_{x \in X} \mathfrak{h}(x) D_{\underline{Q}}(f_{\underline{\omega}})(x)| \leq \varepsilon + o(1),$$

for all bounded families of functions $(f_{\underline{\omega}})_{\underline{\omega} \in \{0,1\}^s \setminus \{0\}}$, $f_{\underline{\omega}} : X \rightarrow [-1, 1]$, then

$$(4.22) \quad \|\mathfrak{h}\|_{\square_{H,M}(\underline{Q}_1, \dots, \underline{Q}_s)} \leq \varepsilon^C + o(1)$$

for some constant $C > 0$ that is independent of ε .

Proof. By multi-linearity it is enough to show that

$$(4.23) \quad \langle f_{\underline{\omega}} \rangle_{\square_{H,M}(\underline{Q}_1, \dots, \underline{Q}_s)} \ll \varepsilon^C + o(1)$$

whenever $|f_{\underline{\omega}}| \leq 1$ or $|f_{\underline{\omega}}| \leq \nu^{(d)}$, assuming that at least one of the functions $f_{\underline{\omega}} = \mathfrak{h}$. We prove (4.23) on the number n unbounded functions $f_{\underline{\omega}}$.

If $n = 1$ we may assume $f_{\underline{0}} = \mathfrak{h}$ and (4.23) follows from (4.20). If $n \geq 2$ we may assume again that $|f_{\underline{0}}| \leq \nu^{(d)}$ is unbounded and write $|\langle f_{\underline{\omega}} \rangle_{\square(\underline{Q})}| \leq \mathbb{E}_{\underline{x}} \nu^{(d)}(\underline{x}) |D_{\underline{Q}}(f_{\underline{\omega}})|$. By the Cauchy-Schwarz inequality, we have

$$|\langle f_{\underline{\omega}} \rangle_{\square_{H,M}(\underline{Q}_1, \dots, \underline{Q}_s)}|^2 \leq \mathbb{E}_{\underline{x}} D_{\underline{Q}}^2(f_{\underline{\omega}})(\underline{x}) + \mathbb{E}_{\underline{x}} (\nu^{(d)}(\underline{x}) - 1) D_{\underline{Q}}^2(f_{\underline{\omega}})(\underline{x}) := I + II.$$

To estimate the first term note that

$$\mathbb{E}_{\underline{x}} D_{\underline{Q}}^4(f_{\underline{\omega}})(\underline{x}) \leq \mathbb{E}_{\underline{x}} D_{\underline{Q}}^4(\nu^{(d)})(\underline{x}) \ll 1,$$

by the polynomial forms condition and the assumption that the family $\underline{Q} = (\underline{Q}_1, \dots, \underline{Q}_s)$ is in general position. On the other hand

$$\mathbb{E}_{\underline{x}} |D_{\underline{Q}}(f_{\underline{\omega}})(\underline{x})| \leq \varepsilon^C + o(1),$$

as it is equivalent to the fact that

$$\mathbb{E}_{\underline{x}} D_{\underline{Q}}(f_{\underline{\omega}})(\underline{x}) g(\underline{x}) \leq \varepsilon^C + o(1),$$

uniformly for all functions $|g| \leq 1$ which follows from the inductive hypothesis on the number of unbounded factors. Then $I \leq \varepsilon^C + o(1)$ by interpolation.

For the second term we may write again,

$$\begin{aligned} D_{\underline{Q}}^2(f_{\underline{\omega}})(\underline{x}) &= \mathbb{E}_{\underline{h}, \underline{h}'} \mathbb{E}_{\substack{\underline{y}_1^{(0)}, \underline{y}_1^{(1)} \\ \underline{y}_2^{(0)}, \underline{y}_2^{(1)}}} \prod_{\substack{* \\ (\underline{\omega}, \underline{\omega}') \in \{0,1\}^{2s}}} f_{\underline{\omega}}(\underline{x} + \sum_{j=1}^s y_{1j}^{(\omega_j)} \underline{Q}_j(\underline{h}) \text{big}) f_{\underline{\omega}}(\underline{x} + \sum_{j=1}^s y_{2j}^{(\omega_j)} \underline{Q}_j(\underline{h}')) \\ &= \mathbb{E}_{\underline{h}, \underline{h}'} \mathbb{E}_{\substack{\underline{y}_1^{(0)}, \underline{y}_1^{(1)} \\ \underline{y}_2^{(0)}, \underline{y}_2^{(1)}}} \prod_{\substack{* \\ (\underline{\omega}, \underline{\omega}') \in \{0,1\}^{2s}}} f_{(\underline{\omega}, \underline{\omega}')}(\underline{x} + \sum_{j=1}^s y_{1j}^{(\omega_j)} \underline{Q}_j(\underline{h}) + \sum_{j=1}^s y_{1j}^{(\omega'_j)} \underline{Q}_j(\underline{h}')), \end{aligned}$$

where $f_{(\underline{\omega}, \underline{\omega}')} = 1$, unless $(\underline{\omega}, \underline{\omega}') = (\underline{\omega}, \underline{0})$ or $(\underline{\omega}, \underline{\omega}') = (\underline{0}, \underline{\omega})$ (with $\underline{\omega} \neq \underline{0}$), in which case $f_{(\underline{\omega}, \underline{\omega}')} = f_{\underline{\omega}}$. Writing $\underline{Q}'(\underline{h}, \underline{h}') = (\underline{Q}_1(\underline{h}), \dots, \underline{Q}_s(\underline{h}), \underline{Q}_1(\underline{h}'), \dots, \underline{Q}_s(\underline{h}'))$ for the extended polynomial system, we have that by box-norm Cauchy-Schwarz inequality

$$\begin{aligned} |\mathbb{E}_{\underline{x}} (\nu^{(d)}(\underline{x}) - 1) D_{\underline{Q}}^2(f_{\underline{\omega}})(\underline{x})| &= |\langle \nu^{(d)} - 1, D_{\underline{Q}'}(f_{\underline{\omega}, \underline{\omega}'}) \rangle| \\ &\leq \|\nu^{(d)} - 1\|_{\square_{H,M}(\underline{Q}')} \prod_{(\underline{\omega}, \underline{\omega}') \neq (\underline{0}, \underline{0})} \|f_{\underline{\omega}, \underline{\omega}'}\|_{\square_{H,M}(\underline{Q}')} = o(1). \end{aligned}$$

Indeed the extended system $\underline{Q}'$ is also in general position (as $\pi_i(\underline{Q}_j(\underline{h})) \neq 0$ for $i \in [d]$, $j \in [s]$), hence $\|\nu^{(d)} - 1\|_{\square_{H,M}(\underline{Q}')} = o(1)$, while

$$\|f_{\underline{\omega}, \underline{\omega}'}\|_{\square_{H,M}(\underline{Q}')} \leq \|\nu^{(d)}\|_{\square_{H,M}(\underline{Q}')} + 1 \ll 1,$$

by the polynomial forms condition. This proves the Proposition. $\square$

Corollary 4.2. *Let $f : X \rightarrow \mathbb{R}$ satisfying $0 \leq f \leq \nu^{(d)}$ and let $\varepsilon > 0$. Then there exists functions $0 \leq g \leq 2$ and $\mathfrak{h}$, such that*

$$f = g + \mathfrak{h}, \quad \text{and} \quad \|\mathfrak{h}\|_{\square_{H,M}(\underline{\mathcal{Q}})} \leq \varepsilon.$$

Proof. Let $0 \leq g \leq 2$ be given as Corollary ?? with $\varepsilon' = \varepsilon^{1/C}$. Then by (4.17) and Proposition 4.2 we have that

$$\|(f - g)\|_{\square_{H,M}(\underline{\mathcal{Q}})} \leq \varepsilon + o(1).$$

The Corollary follows by writing $\mathfrak{h} = f - g$. □

5. PROOF OF THEOREM 1.

The other main ingredient of obtaining polynomial patterns in relative dense subsets of the primes is using an appropriate version of the polynomial extension of Szemerédi's theorem due to Bergelson-Leibman [1]. The version we use, given [16, 17], is as follows.

Theorem E. *Let $\delta > 0$, $\mathcal{P} = (P_1, \dots, P_l)$, $P_j \in \mathbb{Z}[y]$, $P_j(0) = 0$ be a polynomial map, and let $V = \{v_1, \dots, v_l\} \subseteq \mathbb{Z}^d$.*

If $g : X \rightarrow [0, 1]$ is a function satisfying $\mathbb{E}_{x \in X} g(x) \geq \delta$, then one has

$$(5.1) \quad \Lambda_{\mathcal{P}, V}(g, \dots, g) \geq c(\delta) - o(1),$$

where $c(\delta) > 0$ is a constant, that depends only on δ , V and polynomial map $\mathcal{P}$.

Now it is easy to prove our main result.

Proof of Theorem 1. Given $A \subseteq \mathbb{P}_{N'}^d$ with $|A| \geq \delta |\mathbb{P}_{N'}|^d$ we choose W , $N = [N/W], \underline{b}$ and the function f_A as in (2.1)-2.2, then by (2.3)-(2.5) we have

$$\mathbb{E}_{\underline{x} \in X} f_A(\underline{x}) \geq \delta', \quad 0 \leq f_A(\underline{x}) \leq \nu^{(d)}(\underline{x}),$$

with $\delta' = c_0 \delta / 4$. Then by Corollary 4.2 with $\varepsilon > 0$ sufficiently small with respect to $c(\delta')$, we have $f_A = g + \mathfrak{h}$ with $\|\mathfrak{h}\|_{\square_{H,M}(\underline{\mathcal{Q}})} \leq \varepsilon$, thus Corollary 3.1:

$$\Lambda_{\mathcal{P}, V}(f_A, \dots, f_A) \geq \Lambda_{\mathcal{P}, V}(g, \dots, g) - \varepsilon^c - o(1) \geq c(\delta') - \varepsilon^c - o(1) \geq c(\delta')/2 - o(1) > 0,$$

as long as $L \geq L(\mathcal{P}, V)$ and $N \geq N(\mathcal{P}, V, \delta)$ is sufficiently large. □

REFERENCES

- [1] V. BERGELSON, A. LEIBMAN. Polynomial extensions of van der Waerden and Szemerédi’s theorems. *J. Amer. Math. Soc.* **9** (1996), 725–753.
- [2] D. CONLON, J. FOX, Y. ZHAO, A relative Szemerédi theorem *Geometric and Functional Analysis* **25.3** (2015): 733–762.
- [3] B. COOK, Á. MAGYAR. Cook, Brian, and Akos Magyar, Constellations in $\mathbb{P}^d$ *International Mathematics Research Notices* **2012.12** (2012): 2794–2816.
- [4] B. COOK, Á. MAGYAR, T. TITICHETRAKUN A multidimensional Szemerédi theorem in the primes via combinatorics. *Annals of Combinatorics* **22** (2018): 711–768.
- [5] D. GOLDSTON, J. PINTZ, YILDIRIM Primes in tuples I *Ann. of Math.* **170** (2009): 819–862
- [6] J. FOX, Y. ZHAO A short proof of the multidimensional Szemerédi theorem in the primes. *American Journal of Mathematics* **137**, no. 4 (2015): 1139–1145.
- [7] W. T. GOWERS Decompositions, approximate structure, transference, and the Hahn–Banach theorem. *Bulletin of the London Mathematical Society* **42**, no. 4 (2010): 573–606.
- [8] B. GREEN, T. TAO The primes contain arbitrarily long arithmetic progressions. *Annals of mathematics* (2008): 481–547.
- [9] B. KUČA Multidimensional polynomial Szemerédi theorem in finite fields for polynomials of distinct degrees. *Israel Journal of Mathematics* **259**, no. 2 (2024): 589–620.
- [10] N. KRAVITZ, B. KUČA, J. LENG” Quantitative concatenation for polynomial box norms. *arXiv preprint arXiv:2407.08636* (2024).
- [11] N. KRAVITZ, B. KUČA, J. LENG” Corners with polynomial side length. *arXiv preprint arXiv:2407.08637* (2024).
- [12] S. PELUSE Bounds for sets with no polynomial progressions. *Forum of Mathematics, Pi*, vol.8 pp. 1–53 Cambridge University Press (2020)
- [13] S. PELUSE, S. PRENDIVILLE, X. SHAO Bounds in a popular multidimensional nonlinear Roth theorem. *Journal of the London Mathematical Society* **110**, no. 5 (2024): e70019.
- [14] O. REINGOLD, L. TREVISAN, M. TULSIANI AND S. VADHAN Dense subsets of pseudorandom sets. *Electronic Colloquium on Computational Complexity*, Report TR08-045 (2008).
- [15] T. TAO The Gaussian primes contain arbitrarily shaped constellations. *Journal d’Analyse Mathématique* **99** (2006): 109–176.
- [16] T. TAO, T. ZIEGLER The primes contain arbitrarily long polynomial progressions. *Acta Mathematica*, **201**(2) (2008): 213–305.
- [17] T. TAO, T. ZIEGLER Narrow progressions in the primes. *Analytic Number Theory: In Honor of Helmut Maier’s 60th Birthday* (2015): 357–379.
- [18] T. TAO, T. ZIEGLER A multi-dimensional Szemerédi theorem for the primes via a correspondence principle. *Israel Journal of mathematics* **207**, no. 1 (2015): 203–228.
- [19] T. TAO, T. ZIEGLER Polynomial patterns in the primes. In *Forum of Mathematics, Pi*, v.6/1 (2018). Cambridge University Press