A Multidimensional Szemerédi's Theorem in the Primes

by

Tatchai Titichetrakun

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Mathematics)

The University of British Columbia
(Vancouver)

July 2016

Abstract

In this thesis, we investigate topics related to the Green-Tao theorem on arithmetic progression in primes in higher dimensions. Our main tool is the pseudorandom measure majorizing primes defined in [51] concentrated on *almost primes*. In chapter 2, we combine the sieve technique used in constructing pseudorandom measure (in this case, Goldston-Yildirim sum and almost primes) with the circle method of Birch to study the number of almost prime solutions of diophantine systems (with some rank conditions). Our rank condition is similar to the integer case, due to the heuristics that almost primes are pseudorandom. In chapter 3, we investigate the generalization of Green-Tao's theorem to higher dimensions in the case of corner configuration. We apply the transference principle of Green-Tao (with hyperplane separation technique of Gowers) in this setting. This problem is also related to the *densification trick* in [16]. In chapter 4, we extend the result of Chapter 3 to obtain the full multi-dimensional analogue of the Green-Tao's theorem, using hypergraph regularity method by directly proving a version of hypergraph removal lemma in the weighted hypergraphs. The method is to run an energy increment on a parametric weight systems of measures, rather than on a single measure space, to overcome the presence of intermediate weights. Contrary to [110], [68] where the authors investigate the problem using a measure supported on primes and infinite linear form conditions, relying on the Gowers Inverse Norms Conjecture.

Preface

This thesis is a combination of three manuscripts: Chapter 2 is based on [75] which is a joint work with A. Magyar. Chapter 3 is based on [74] which is joint work with A. Magyar. Chapter 4 is based on [18] which is a joint work with B. Cook and A. Magyar.

Table of Contents

List of Figures

Notations

Index Notations

- Write $\mathbf{x} = (x_1, ..., x_d), \mathbf{y} = (y_1, ..., y_d)$ as vectors in dimension d. Write $\underline{\omega} = (\underline{\omega}(1), \ldots, \underline{\omega}(d)) \in \{0,1\}^d$, and for each such $\underline{\omega}$, let $P_{\underline{\omega}} : \mathbb{Z}_N^{2d} \to \mathbb{Z}_N^d$ be the projection defined by

$$P_{\underline{\omega}}(\mathbf{x}, \mathbf{y}) = \underline{u} = (u_1, ..., u_d), u_j = \begin{cases} x_j & \text{if} \quad \omega_j = 0 \\ y_j & \text{if} \quad \omega_j = 1 \end{cases}$$

- For each $I \subseteq [d], \mathbf{x}_I = (x_i)_{i \in I}$. We may denote $\mathbf{x}$ for $\mathbf{x}_{[d]}$.
  $\underline{\omega}_I$ means elements in $\{0,1\}^{|I|}$. Similarly we may write $\underline{\omega}$ for $\underline{\omega}_{[d]}$. We also define $P_{\underline{\omega}_I}(\mathbf{x}_I, \mathbf{y}_I)$ in the same way.

- $\underline{\omega}|_I$ denotes $\underline{\omega}$ restricted to the index set I.

- For finite sets $X_j, j \in [d], I \subseteq [d]$ then $X_I := \prod_{j \in I} X_j$ and

$$P_{\underline{\omega}_I}(X_I, Y_I) = \prod_{i \in I} Z_i,\ Z_i = \begin{cases} X_i, & \underline{\omega}_I(i) = 0 \\ Y_i, & \underline{\omega}_I(i) = 1 \end{cases}$$

- If we want to fix on some positions, we can write, for example $\underline{\omega}_{(0,[2,d])}$ means element in $\{0,1\}^d$ such that the first position is 0.

- For each $\underline{\omega}$, define $\mathbf{y}_{1(\underline{\omega})} \in \{0,1\}^d$ by

$$(\mathbf{y}_{1(\underline{\omega})})_i = \begin{cases} 0 & \text{if} \quad \omega_i = 0 \\ y_i & \text{if} \quad \omega_i = 1 \end{cases}, 1 \le i \le d.$$

  $\mathbf{y}_{0(\underline{\omega})} \in \{0,1\}^d$ is also defined similarly.

Set Notations

$[N] := \{1, 2, ..., N\}, [M, N] := \{M, M+1, ..., M+N\}$.
$\mathcal{P}$ denotes the set of primes. $\mathcal{P}_N, \mathcal{P}[N] := \mathcal{P} \cap [N]$.
For any finite set X and $f : X \to \mathbb{R}$ or $\mathbb{C}$, and for any measure μ on X,

$$\mathbb{E}_{x\in X} f(x) := \frac{1}{|X|} \sum_{x\in X} f(x). \quad \mathbb{E}_\mu f(x), \int_X f d\mu := \frac{1}{|X|} \sum_{x\in X} f(x)\mu(x).$$

Hypergraph Setting: Suppose we have a $(d+1)-$partite hypergraph with vertex sets $V_1, \ldots, V_{d+1}$. For $e \subseteq [d+1]$, we may write $V_e := \prod_{j\in e} V_j$. Let $\pi_e : V_{[d+1]} \to V_e$ be the natural projection. We write $\mathcal{A}_e = \{\pi_e^{-1}(F) : F \subseteq V_e\}$ as subsets of $V_{[d+1]}$.

Other Notations

- Linear characters: for $\theta \in \mathbb{R}/\mathbb{Z}$, $e(\theta) := e^{2\pi i\theta}$.
- $\exp(x) = e^x$.
- Unless otherwise specified, the error term $o(1)$ means a quantity that goes to 0 as $N \to \infty$ (or $N, W \to \infty$ in $W-$trick, see section 2.2).
- $f(x) \approx g(x)$ means there are absolute constants c, C such that $cf(x) < g(x) < Cf(x)$.
- $\mathbb{Z}_N, \mathbb{Z}/N\mathbb{Z}, \mathbb{Z}/N$ means additive group of integers $\pmod N$.
- $k - AP$ means arithmetic progression of length k. Is it nontrivial if the common difference is not zero.
- Multiplicative difference: $\Delta_h f(x) := f(x)\overline{f(x+h)}$. Additive difference $\Delta_h^+ f(x) := f(x+h) - f(x)$.

Acknowledgments

I would like to thank Prof. Ákos Magyar, my thesis advisor, who shares the enthusiasms of the subject and is always patient with me. For his unconditional support, trust ; without any judgments. He, with his collaboration with Prof. Neil Lyall, also introduced me to the field of additive combinatorics long time ago. Professors in Harmonic Analysis group; Prof. Izabella Łaba, Prof. Malabika Pramanik who showed me as a role model of great mathematicians and always provide kind supports. Prof. József Solymosi, who becomes my co-advisor when Ákos moved to Georgia, who taught me discrete math and always give me advice and pull me out of my comfort zone. Prof. Brian Marcus whom I have audited in his entropy and information theory class, he shared the passion in the subject and his tremendous care of students. Prof. Greg Martin whom I took his analytic number theory class and showed what does it mean to be a teacher (Yes, I have to work like a dog in his class). Also all people at harmonic Analysis group who shared passions of the subjects; Robert Fraser, Marc Carnovale, Kyle Hambrook, Kevin Henriot, Ben Krause, Dimitrios Karslidis. Having met all of these great people is like a dream. I also thank to Math Department at UBC who gives me opportunities to work here and gave some bucks for me to eat. People at math offices and UBC who are always helpful. I also have opportunities to meet many great and inspiring people at conferences, even though I am usually shy. Prof. Nikos Frantzikinakis, Prof. Terry Tao, Prof. Ben Green, Voraphan Chandee, Yufei Zhao and many more inspiring people. I also have to thank to all people in my field who did great stuff. Finally, to people who read and review this manuscript, in particular, Prof Tamar Ziegler, Prof. Malabika Pramanik and Prof Ákos Magyar. This thesis is not perfect and their comments are invaluable for my development. I would like to thank my friends I met math department, in particular, Wordsobe Mun and family, Atsushi Kanazawa, Justin Scarfy and Nishant Chandgotia. Also, Tianhan Lu, Mimi Brown, Zoe Lam, Phoebe Li. I thank my Thai friends I met in Vancouver who shared a good time as starving people here, Wisarn Yenjaichon, Patcharapa Sangsuttarat, Petcharatana Bhuanantanondh, Siriphon Siriangkanakul, Tipp & Jim Placzek, Wassamon Kunamornpong, Suvaporn Phasuk. Finally, I cannot succeed without my Thai friends and teachers: Prof. Zen Harper, Prof. Pacchara Chaisuriya, Prof. Nattapan Kittisin, Prof. Wicharn Lewkeeratiyuttakul, Prof. Aram Tangboonduangjit, Prof Vichian Laohakosol, Prof. Nittiya Pabhapote, Prof. Songkiat Sumetkijakan, Prof. Julian Paulter, Mr. Surachai Boonrueng, Mr. Suchart Somsuk, Miss Pavinee Hantanun, Smith Iampiboonvatana, Wittaya Siricheepchaiyan, Preecha Rittiwattanadech, Kiratad Neerasen, Teerayuth Chuedee, Krisda Boonthose, Anucha Sanrawang, Nattakarn Numpanviwat, Kuntalee Chaisee, Piyashat Sripratak, Preechaya Sanyatit, Raywat Thanatkithiran, Nopporn Thamrongrat, Athipat Thamrongthanyaluk and many other important people that I don't have a space to mention here.

I learned many things since I moved to Vancouver including sumo and shogi. I thank my family for their unconditional support. Diligent Mayonnaise at Fujiya. Killer Saikyo miso yaki black cod which is mind-blowing. Ten musu, chicken teriyaki, salted fish and Japanese dish at Fujiya. Thai vegetables from Asia Market. BBQ unagi at Ajisai. Korean drinks and Korean items from Kim Market. Chinese items from T & T and Henlong supermarket. Dim-sum, BBQ meat and Chinese cuisine in Vancouver. Beaty Museum where I learned to forage and do gyotaku. Fee-bee song of the black-cap chickadees. Ukiyoe, mingei and sumie at Nikkei National Museum.

Dedication

To papa and mama.

Chapter 1

Introduction: Szemerédi's and Green-Tao's Theorem

> On occasions a mathematician will have an insight that is ahead of the time in the sense that the insight is not fully expressible in the mathematical theory and language developed at the moment. For example the Poincaré Recurrence Theorem as first stated and proved by Poincaré was strictly not meaningful. What was needed was the language of Lebesgue measure which came later.-*Walter Gottschalk*

The main area of mathematics that is related to this thesis is *additive combinatorics*. Problems in additive combinatorics usually ask to count or estimate additive structures in sets. This field has some origins from additive number theory that has interested people since the ancient time. There are many classical problems from this field, for example, what is the number of solutions in a given Diophantine equation? Can every even integers $n \geq 4$ be written a sum of two primes? Various methods such as circle method first developed by Hardy-Littlewood and sieve theory are used to attack these problems. We may also ask for properties of set addition e.g. if $A \subseteq \mathbb{Z}_+$, what is the size of $A + A$? When is it small or when is it large? Usually studying problems like this involve studying two very different operations: addition and multiplication. This sometimes make problems in this area hard, even with current technology. Analogue problems may be asked in some abstract setting e.g. if we let A to be a subset of arbitrary groups. Problems in additive combinatorics nowadays can be very abstract and usually involves other area of mathematics other than purely number theory or combinatorics. Since we usually ask to estimate the size of the sets, the tools from analysis (e.g. harmonic analysis) can be handily adapted to our finite setting.

A problem posed by Erdős asking that if $A \subseteq \mathbb{Z}_+$ with $\sum_{a\in A} 1/a = \infty$ must contain a (non-trivial) $k-$term arithmetic progression? This motivates the study of additive structures in a large subset of $\mathbb{Z}$ where what we mean by large is indeed also a question. The famous result in this direction is the Szemerédi's Theorem and Green-Tao's theorem discussed in the next few sections. This direction of research extends the area of additive combinatorics in touch with other area of mathematics such as Ergodic theory and Lie Theory.

Green-Tao's theorem generalizes Szemerédi's Theorem to the case of primes. One of the main technique in proving these kinds of theorem is to decompose an arbitrary set to structural part and uniformity part (that does not correlate with the structures). What should be the notion of structures and uniformity? How to measure them ? These are already hard and interesting questions. Green-Tao managed to develop a tool to measure uniformity of primes that is sufficient to deduce results about arithmetic progressions on them. In this thesis, we prove some results motivated from their work. In chapter 2 we use the Goldston-Yildirim sieve used in the original proof of Green-Tao's theorem [50] combined with the circle method of Birch [11] to study number of solutions in almost prime of Diophantine equations (with some rank conditions). In the next two chapters we prove analogue results of Green-Tao's Theorem in higher dimensions, $\mathbb{Z}^d$.

1.1 Szemerédi's Theorem

One of the most important theorem in additive combinatorics is the Szemerédi's Theorem. Informally, Szemerédi's theorem states that the sets of integers contains so many arithmetic progressions (or affine copies of a finite set in $\mathbb{Z}^d$) that any subsets with positive density (see Definition 1.1.2) must contain many of them. The point is that there is no assumption on the set A (other than its size), it could be either purely random or supplied with some explicit structures, which we can show that they contain arithmetic progressions with different reasons for each case. In general, we try to decompose arbitrary sets into structured and random (or pseudorandom) parts where the techniques are already interesting by themselves.

There are many approaches to Szemerédi's Theorem. Basically, we are not going to find arithmetic progressions but we will count them.

1. The first combinatorial approach due to Szemerédi [99] introduces the Szemerédi's regularity lemma which is a theorem describing the structure of a large graph. The Lemma became important in combinatorics and computer science. We will talk more about regularity lemma in section 1.3.
2. There is a combinatorial *Fourier analytic approach* due to Gowers [36], [37] where he introduces the *Uniformity norms* (of various degrees determined by the complexity of structures) to measure uniformity or randomness of functions. A function which correlates with some kinds of structures would be large in uniformity norms of appropriate degrees. The harder inverse theorem asks for the converse: if a function is large with a uniformity norm, what kind of structures can the function be correlated with? An inverse theorem would make the uniformity norms an effective tool in studying structures in sets. This is a generalization of arguments of Roth [91]

or 3-term arithmetic progressions. Roth observed that if a set $A \subset \mathbb{Z}_N$ with density $\alpha > 0$ has the Fourier coefficient bound $\|\widehat{\mathbf{1}_A - \alpha}\|_\infty$ sufficiently small (i.e. it does not correlate with linear phase functions) then A contains many three terms arithmetic progressions comparable to random sets. Otherwise A has a structure in the sense that it has increased relative density in some long arithmetic progressions. Then we can do density increment (see section 1.1.4) to obtain a structured subset of A which contains arithmetic progressions.

One may want to generalize this idea the four-term arithmetic progressions using exponent of quadratic polynomials.[1] However this turns out not to be the case (see example 1.1.12), in particular, quadratic functions on multi-dimension arithmetic progressions may not correlate with quadratic polynomial phases (hence quadratic polynomial phases are not sufficient to describe structural objects in this case). Gowers exploits tools from additive number theory such as Frieman-Ruzsa Theorem (structure of sets with small sumsets in term of multi-dimension arithmetic progressions) and Balog-Gowers-Szemerédi's theorem (sets with many additive quadruples have large subsets satisfy the conditions on Frieman's theorem) to deal with objects like multidimensional quadratic phase functions. Gowers managed to obtain *a local version of the inverse U^k norm theorem* on $\mathbb{Z}_N$, meaning Gowers obtained correlations with many polynomial phase functions, each on an arithmetic progression. These works also inspired many later works such as studying global inverse theorem [59]. Full global inverse theorem as a direct generalized of Roth's argument would obtain later in [60] where obstructions are actually described in terms of *nilsequences*[2] (but not with a good bound and the proof is very long).

3. The next approach to Szemerédi's Theorem is the *ergodic theory approach* initiated by Furstenberg and Katznelson (e.g. [33], [31]) where they transfer this Szemerédi-type problems (via Furstenberg-corresponding principle, first formulated in [26]) to studying multiple recurrence in a probability measure preserving system[3]. For example,to prove Szemerédi's Theorem, one can study

$$\frac{1}{N}\sum_{n=1}^{N} f_1(T^n x) f_2(T^{2n} x) \dots f_k(T^{kn} x) \tag{1.1.1}$$

Here $f_i \in L^\infty(X)$ and $T : X \to X$ is measure preserving. This average is referred as *multiple recurrence*. To understand the limiting behavior (as $N \to \infty$) of (1.1.1) , the key idea is to understand the characteristic factor Z which is an invariant subsigma-algebra Z such that if $\mathbb{E}(f_i|Z) = 0$ for some i then the limit of (1.1.1) would be 0 in L^2 norm. Hence the explicit description of characteristic factors is a useful tool to prove results on multiple recurrence like (1.1.1). The question of finding the characteristic factor for a given multiple recurrence is a del-

[1] Indeed, we have a quadratic obstruction for four term arithmetic progressions $x^2 - (x+d)^2 + (x+2d)^2 - (x+3d)^2 = 0$. This is the only obstruction.

[2] These are technical objects and we will not define them here but we discuss a bit about them in the next paragraph.

[3] This means a set X together with a Borel $\sigma-$algebra $\mathcal{B}$ and a Borel probability measure μ and a measure preserving transformation $T : X \to X$ meaning $\mu(T^{-1}A) = \mu(A) \forall A \in \mathcal{B}$.

icate one. Host and Kra [65] and independently Ziegler [115] are able to give a nice description of the characteristic factor of (1.1.1) for any k in terms of nilrotation on nilmanifolds (this could be considered as a generalization of abelian rotation on S^1, the Kronecker's factor) and Host-Kra seminorm (an analogue of Gowers norm), see e.g. appendix A in [8] for a brief introduction to these objects. Nilsequences [7] play a role as the obstruction to uniformity similar to linear exponentials in Roth's theorem case. This motivated parallel work in additive combinatorics. The motivation of why nilpotent groups arise is that in a $k-$step nilsystem, the first $k-$term of geometric progressions will determine the rest, see [71] section 6.4. Ergodic theory is the method that can attack most general kinds of patterns in this kind of problems. Let us state a theorem of Furstenberg-Katznelson which is equivalent via corresponding principle to the multidimenisonal Szemerédi's theorem.

Theorem 1.1.1 (Furstanberg-Katznelson [31]). *Let $(X, \mathcal{B}, \mu)$ be a probability measure space and $T : X \to X$ is a measure preserving transformation on $(X, \mathcal{B}, \mu)$. Let $A \in \mathcal{B}, \mu(A) > 0, k \in \mathbb{Z}_+$ then $\exists B \subseteq A, \mu(B) > 0$ and $n \geq 1$ such that*

$$T^n B, T^{2n} B, \ldots, T^{(k-1)n} B \subseteq A.$$

i.e.

$$\mu(\bigcap_{j=0}^{k-1} T^{-jn} A) > 0.$$

Tao gave finitary ergodic argument [101] where he discretized the ergodic argument and introduce UAP-norm which is a counter part of uniformity norm.

Roth's theorem also follows from studying eigenvalues (spectra) of graph using Cheeger-type discrepancy bound, see [98]. Roughly speaking the largest second value of the adjacency matrix is an analogue of the second largest Fourier coefficients of a dense set. Generalizing this theory to spectra of hypergraphs seems challenging. Though, in ergodic setting, there is a description of characteristic factor of (1.1.1) with $k = 4$ in terms of generalized eigenfunctions, see e.g. [116].

4. Finally, there is a hypergraph regularity approaches due to Vojta Rödl, B. Nagel, M. Schacht, J. Skokan [82],[83], [85] that generalize argument in [97] and also another hypergeaph approach due to Gowers [35], [36]. These stronger hypergraph regularity lemma allow us to deduce the multidimensional Szemerédi's theorem. This exploited ideas in ergodic theory (energy increment) and conditional expectations on sigma-algebras to obtain the required decomposition. A stronger functional version of hypergraph approach due to Tao [104], which will be the version we generalize to prove the main result in Chapter 4.

Next we formally state Szemerédi's theorem.

Definition 1.1.2. *Let $A \subseteq \mathbb{N}$, we say A has* ***positive upper density*** *if $\limsup_{N\to\infty} \frac{|A\cap[N]|}{N} > 0$,i.e. there is a $\delta > 0, N_j \nearrow \infty$ such that $\frac{|A\cap[N_j]|}{N_j} \geq \delta$ for all j. We say that A has* ***positive upper Banach density*** *if there is a sequence of intervals $I_j \subseteq \mathbb{N}, |I_j| \nearrow \infty$ such that $\frac{|A\cap I_j|}{|I_j|} \geq \delta$ for all j. A similar notation may be similarly defined on $\mathbb{Z}$ or $\mathbb{Z}^d$ using the Cartesian product of intervals.*

Now we state two versions of Szemerédi's Theorem in dimension 1.

Theorem 1.1.3 (Szemerédi's Theorem; equivalent forms)**.** *1.(**Infinite Version**) If $A \subseteq \mathbb{N}$ has positive upper Banach density then for all $k \in \mathbb{Z}_+$, there exist $x, t, t \neq 0$ such that $P := \{x, x+t, ..., x+(k-1)t\} \subseteq A$*
*2.(**Finite Version**) Let $\delta > 0$ then there is an $N(k,\delta)$ such that if $N \geq N(k,\delta)$ then any $A \subseteq [1, N]$ with $|A| \geq \delta N$ has some $P = \{x, x+t, ..., x+(k-1)t, t \neq 0\} \subseteq A$.*

The point is that $N(k,\delta)$ is independent of A. It is easy to check that these two statements are equivalent. We may also replace the P with $F' := x+tF$ for any finite set F, where $F' = \{x+tf, f \in F\}$ is an affine copy of F.

Now we state the functional version of Szemerédi's Theorem. Note that it is more convenient to work in a more structured setting like in a group, e.g. $\mathbb{Z}_{N'}$ (where N' is a prime much bigger than N) and there is a standard argument that deduces the result on $[N]$ from that, as we will do.

Theorem 1.1.4. *Let $f : \mathbb{Z}_N \to [0,1]$ with $\mathbb{E}_{x\in\mathbb{Z}_N} f(x) \geq \delta$. Then there is a positive constant $c(k,\delta)$ such that*

$$\mathbb{E}_{x,t\in\mathbb{Z}_N} f(x)f(x+t)...f(x+(k-1)t) \geq c(\delta,k) - o_\delta(1). \tag{1.1.2}$$

Theorem 1.1.4 implies the finite version by taking f to be the characteristic function on a set A and using an average argument of Varnavides [114] stating about the conclusion in finite version of Szemerédi's theorem that we will have at least $c(\delta',k)N^2$ of such progressions. Conversely the set version also implies the functional version by considering sets of the form $\{x : f(x) \geq \delta/2\}$ with a simple average argument.

Theorem 1.1.5 (Varnavides's Theorem [114])**.** *The conclusion of Theorem 1.1.3 (finite version) may be strengthened to conclude that A contains at least $c(\alpha,k)N^2$ $k-AP$(i.e. $k-$term arithmetic progression).*

Proof [114]: Consider $N(k,\delta)$ such that $L \geq N(k,\delta)$ then any set $A \subseteq \{1,\dots,L\}$ with density $\geq \delta/2$ must contains a nontrivial $k-$AP. This follows from the following observations:

- We work in $\mathbb{Z}/N$. Consider a long $L-$ arithmetic progression $S_{a,d} = \{a+d,\dots,a+Ld\} \subseteq \mathbb{Z}/N$ where $a, d \in \mathbb{Z}_N, d \neq 0, L \leq N$. If $S_{a,d}$ intersects A with at least $\delta L/2$ elements then $S_{a,d} \cap A$ contains a nontrivial $k-$AP.

– Consider all $L-$AP with a fixed common difference $d \neq 0$. Varying a, we have $\sum_a |S_{a,d} \cap A| = L|A| \geq \delta LN$. Hence $|S_{a,d} \cap A| \geq (\delta/2)L$ for at least $(\delta/2)N$ values of a.
Now vary d then $|S_{a,d} \cap A| \geq (\delta/2)L$ for at least $(\delta/2)N(N-1)$ values of a, d. Each of these $L-$AP contains a nontrivial $k-$AP.

– We consider possible repetitions of arithmetic progressions counted. Observe that any nonrivial $k-$AP could be contained in at most $L(L-1)$ $L-$APs. (punch line: for each $k-$AP, the indices of the first two terms of this $k-$AP in the $L-$AP would determine the $L-$AP it is in). Hence the numbers of $k-$APs in A is at least

$$\frac{\delta}{2}\frac{N(N-1)}{L(L-1)} \gtrsim \frac{\delta}{2}\frac{N^2}{L^2} = c'(k,\delta)N^2$$

One may take e.g. (when $k=3$) $L = \lceil \exp(C\delta^{-1}(\log(1/\delta)^4)) \rceil$ according to Bloom [9].

□

Finally, let us remark a more general version of Szemerédi's Theorem called *Density Hales-Jewett's Theorem*. This theorem implies Szemerédi's theorem in finite group.

Theorem 1.1.6 (Density Hales-Jewett's Theorem [29])**.** *Let $F \subseteq \mathbb{Z}$ and $0 < \delta < 1$. If $N \geq N(|F|, \delta)$ then for any $A \subseteq F^d, |A| \geq \delta |F|^d$, A contains a set of the form $\{a + tr; t \in F, a \in \mathbb{Z}^d, r \in \mathbb{Z}^d, r \neq 0\}$.*

Remark 1.1.7 (Quantitative Bound in Szemerédi's Theorem)**.** *We have the following equivalent statements of Szemerédi's Theorem.*

$$0 \leq f \leq 1, \mathbb{E}_{x \in \mathbb{Z}/N} f(x) \geq \delta \Rightarrow \mathbb{E}_{x,r \in \mathbb{Z}/N} f(x) f(x+r) \ldots f(x+(k-1)r) \geq c_1(k,\delta)$$
$$(N \textit{ in term of } \delta) \quad A \subseteq [N], |A| \geq \delta N, N \geq c_2(k,\delta) \Rightarrow A \textit{ contains some } k-\textit{term arithmetic progressions.}$$
$$(\delta \textit{ in term of } N) \quad A \subseteq [N], |A| \leq r_k(N) \Rightarrow A \textit{ contains no } k-\textit{term arithmetic progressions.}$$

Example: The original bound obtained by Roth [91] is given by $c_2(3,\delta) > \exp(\exp(c\delta^{-1}))$. This is equivalent to $r_3(N) \leq C\frac{N}{\log\log N}$.

We have the following records

$$N2^{-\sqrt{8\log N}} \textit{ (O'Bryant [79])} \leq r_3(N) \leq C\frac{(\log\log N)^4}{\log N}N \textit{ (Bloom [9])}$$
$$r_4(N) \leq C\frac{N}{\exp(C\sqrt{\log\log N})} \textit{ (Green-Tao [57])}$$
$$r_4(N) \leq \frac{N}{\log^c N} \textit{ for some small constant } c > 0 \textit{ (Green-Tao [58])}.$$

For $k > 4$ we have the following bounds due to Gowers [39],

$$c_1(k,\delta) > 2^{-2^{1/\delta^{c_k}}}, c_k = 2^{2^{k+9}}$$
$$c_2(k,\delta) > \exp(\exp((1/\delta)^{c_k}))$$
$$CN\exp(-N2^{\frac{N-1}{2}}(\log N)^{1/N} + \frac{1}{2N}\log\log N)[79] \leq r_k(N) \leq C\frac{N}{(\log\log N)^{2^{-2^{k+9}}}}.$$

Improving quantitative bound of Szemerédi's Theorem, even in the case $k = 3$, is an interesting research problem. In a precedent work, Bourgain [12] obtains an upper bound of $r_3(N)$ by analyzing the structure of Bohr sets (see section 1.1.3) and doing density increment on Bohr sets. Bloom [9] obtains the bound by analyzing combinatorial properties of large spectrum set (the same to the recent work on the bound of *cap-set problem*[4] [3] on the size of subsets of $\mathbb{F}_p^n$ not containing three term arithmetic progressions, which is $C3^n/n^{1+\varepsilon}$). The lower bound of $r_3(N)$ first obtained by Behrend [5] using the idea that there is no three-term arithmetic progressions on a sphere. There is no substantial improvement and it might be the optimal shape. This turns out to be the case for Roth theorem in four variables [93]. Much less quantitative results are known in higher dimensional cases.

1.1.1 Gowers Uniformity Norms

Definition 1.1.8 (U^d-norm and inner product)**.** *Let G be a finite abelian group define[5] the U^d-(generalized) inner product of 2^d functions $f_\omega, \omega \in \{0,1\}^d$,*

$$\langle (f_\omega)_{\omega\in\{0,1\}^d}\rangle_{U^d} = \mathbb{E}_{x,h_1,\dots,h_d\in G}\prod_{\omega\in\{0,1\}^d} C^{|\omega|}f(x+\omega_1h_1+\dots+\omega_dh_d)$$

In particular when $f_\omega = f\ \forall\omega$, we have the following definition of Gowers uniformity norm

$$\|f\|_{U^d}^{2^d} = \mathbb{E}_{x,h_1,\dots,h_d\in G}\prod_{\omega\in\{0,1\}^d} C^{|\omega|}f(x+\omega_1h_1+\dots+\omega_dh_d)$$

where $Cf := \overline{f}$ is the complex conjugate and $|\omega| = \omega_1 + \dots \omega_d$.

Lemma 1.1.9 (Basic properties of Gowers uniformity norms, see e.g. [109])**.**

1. *If f is a bounded function then*[6]

$$\|f\|_{U^d} \leq \|f\|_{\frac{2^d}{d+1}}$$

2. *Gowers-Cauchy-Schwarz's inequality $|\langle (f_\omega)_{\omega\in\{0,1\}^d}\rangle| \leq \prod_{\omega\in\{0,1\}^d}\|f_\omega\|_{U^d}$. In particular,$\|\cdot\|_{U^d}$ is indeed a norm.*

[4] see [108] for a very recent improvement on this problem via polynomial method.
[5] We don't actually need conjugates as we are working on real valued functions, these conjugated may be neglected later
[6] This is sharp as mentioned in [66], an easy application of Holder inequalities can give us easy looser bound.

3. $\|f\|_{U^d} \leq \|f\|_{U^{d+1}}$.

The following recurrence relation is usually useful when working with higher uniformity norms

$$\|f\|_{U^d}^{2^d} = \mathbb{E}_h \|f T^h f\|_{U^{d-1}}^{2^{d-1}} \tag{1.1.3}$$

Example 1.1.10. *[116] Suppose $f : \mathbb{Z}_N \to \{1, -1\}$ is a random function with mean 0 then by the law of large number, $\|f\|_{U^k} = o(1)$ with high probability.*

Remark 1.1.11. *An analogue of Gowers uniformity norms called Host-Kra seminorms can in fact be defined on general measure spaces [65] where they are in general seminorms. They are norms exactly when they are defined on a nilmanifold.*

When working in the uniformity norm, we are interested in proving an inverse theorem: Finding the set of structure function $\mathcal{F}_d$ such that if f has a large uniformity norm of degree d then f correlate with some elements of $\mathcal{F}_d$, in other words, if f has small uniformity norm of degree d then f does not correlate with any structures in $\mathcal{F}_d$. From property (3) of Lemma 1.1.9, we see that functions with small higher order Gowers norm will correlate with less structures i.e. $\mathcal{F}_d \subseteq \mathcal{F}_{d+1}$. We will prove analogue properties of uniformity norm on more general weighted hypergraph.

For example as in the proof of Roth's theorem, function with large U^2 norm , due to relation $\|f\|_{U^2} = \|\hat{f}\|_4$, will correlate with some linear exponents. In higher order uniform, however, we have the following example.

Example 1.1.12. *[[39], section 4.] Consider two-dimensional arithmetic profression in $\mathbb{Z}_N, N$ is prime: $P = \{x_1 + Kx_2 : -K/10 \leq x_1, x_2 \leq K/10, K = \lfloor\sqrt{N}\rfloor\}$ and $f(x_1 + Kx_2) = e((x_1^2 + x_2^2)/N)\mathbf{1}_P$. The function $\phi(x_1 + Kx_2) = x_1^2 + x_2^2$ is a quadratic function on P in the sense that*

$$\Delta^+_{h_1}\Delta^+_{h_2}\Delta^+_{h_3}\phi(x) = 0 \; \forall x, h_1, h_2, h_3 \in P.$$

and we can show that $\|f\|_{U^3} \gg 1$. On the other hand, we can do a calculation that $|\langle f, e(\psi)\rangle| = O(N^{-c})$ for any quadratic phases $\psi(x) = rx^2/N + sx/N + t$ with $r, s \in \mathbb{Z}_N, t \in \mathbb{R}/\mathbb{Z}$. Hence the exponent of quadratic polynomial function of this form is not the only quadratic obstruction of $\|\cdot\|_{U^3}$ norm.

1.1.2 Box Norm

Box norm is a more abstract version of the Gowers uniformity norms defined on hypergraphs i.e. on a function $f : X_1 \times \cdots \times X_d \to \mathbb{R}$.

Definition 1.1.13 (Box norms)**.** *Let $X_1, \dots, X_d$ be finite sets. $f : X_1 \times \cdots \times X_d \to \mathbb{R}$. Define the box norm of order d,*

$$\|f\|_{\square^d}^{2^d} = \mathbb{E}_{\substack{\mathbf{x}\in X_1\times\cdots\times X_d,\\ \mathbf{y}\in Y_1\times\cdots\times Y_d}} \prod_{\omega\in\{0,1\}^d} f(P_\omega(\mathbf{x},\mathbf{y}))$$

For $e \subseteq \{1,\dots,d\}$, we can define $\|f\|_{\square^e}$ for $f : V_e \to \mathbb{R}$.

Example 1.1.14. $\|f\|_{\square^2}^4 = \mathbb{E}_{x\in X, x'\in X', y\in Y, y'\in Y'} f(x,y)f(x',y)f(x,y')f(x',y')$.

On a hypergraph system $(J = [d+1], (V_j)_{j\in J}, \mathcal{H})$[7], we put sigma algebras $\mathcal{B}_e$ on each $V_e = \prod_{j\in e} V_j$. We can also think of $\mathcal{B}_e$ as the $\sigma-$algebra $\mathcal{A}_e$ on V_J where $\mathcal{A}_e = \pi_e^{-1}(\mathcal{B}_e)$ for the projection $\pi_e : V_J \to V_e$.

Lemma 1.1.15 ([52], Gowers uniform functions are orthogonal to lower order sets, in other words, there are no correlations between them)**.** *Given a hypergraph system $(J = [d+1], (V_j)_{j\in J}, \mathcal{H})$. Let $e \in \mathcal{H}, |e| = d'$, $f : V_e \to \mathbb{R}$.*

1. *$\|fg\|_{\square^e} \le \|f\|_{\square^e}$ when $g : V_e \to \mathbb{R}$ is independent of x_j-variable for some $j \in e$.*
2. *$\mathbb{E}_{x_e\in V_e} f(x_e) \prod_{e'\subsetneq e} \mathbf{1}_{E_{e'}(x_{e'})} \le \|f\|_{\square^e}$*
3. *Suppose there exists sub-algebras $\mathcal{B}_{e'} \subseteq \mathcal{A}_{e'}$ with $compl(\mathcal{B}_{e'}) \le M$ for any $e' \subseteq J$ with $|e'| < d$, Then for any $e \in H$, let $E'_e \in \bigvee_{e'\subsetneq e} \mathcal{B}_{e'}$ then*

$$\mathbb{E}_{x\in V_J} \mathbf{1}_{E'_e}(x) f(\pi_e(x)) = O_M(\|f\|_{\square^e})$$

Proof. The first statement follows when we expand the LHS using the definition of box norm. The second statement follows by iterated applications of the first statement, see the proof of Theorem 3.5.5 (on weighted hypergraph) for details. The last statement follows from the second statement, triangle inequalities together with the fact that E'_e is a union of $O_M(1)$ atoms of $\bigvee_{e'\subsetneq e} \mathcal{B}_{e'}$. $\square$

Definition 1.1.16. *Given a hypergraph system $(J, V_J, \mathcal{H})$ and a $\sigma-$algebra $\mathcal{B}$ on V_J. and let $e \in \mathcal{H}_d$. Write $\partial e = \{f \subseteq e : |f| = |e| - 1\}$. Define the $e-$discrepancy $\Delta_e(E_e|\mathcal{B})$ of the set $E_e \in \mathcal{B}_e$ with respect to $\mathcal{B}$ by*[8]

$$\Delta_e(E_e|\mathcal{B}) := \sup_{\substack{E_f\in\mathcal{A}_f\\ \forall f\in\partial e}} \left| \int_{V_J} (\mathbf{1}_{E_e} - \mathbb{E}(\mathbf{1}_{E_e}|\mathcal{B})) \prod_{f\in\partial e} \mathbf{1}_{E_f} d\mu \right| \tag{1.1.4}$$

Note that the largeness of $\Delta_e(E_e|\mathcal{B})$ implies the largeness of $\|\mathbf{1}_{E_e} - \mathbb{E}(\mathbf{1}_{E_e}|\mathcal{B})\|_{\square^e}$. To see this, write $F = \mathbf{1}_{E_e} - \mathbb{E}(\mathbf{1}_{E_e}|\mathcal{B})$. Let $x'_j \in X_j$ for $1 \le j \le d$ be fixed and $y_1, \dots, y_d \in [0,1]$. For $1 \le i \le d$,

[7] See definition 4.1.4. Basically, J is a index set $V_J := (V_j)_{j\in J}$ and $\mathcal{H} \subseteq \prod_{j\in J} V_j$ is a hypergraph. See section 1.1.5 for relevant notions on $\sigma-$algebras.

[8] We could replace the product on $f \in \partial e$ to $f \subsetneq e$ or replacing $\mathbf{1}_{E_e} - \mathbb{E}(\mathbf{1}_{E_e}|\mathcal{B})$ with a bounded function with the same proof.

define

$$E_i = \{(\mathbf{x}_{[d]\setminus\{i\}}) : \prod_{\underline{\omega}_{[i,d]}} F(x_1, \ldots, x_{i-1}, x_i', P_{\underline{\omega}_{[i,d]}}(x_{[i,d]})) \geq y_i\} \in \mathcal{B}_{[d]\setminus\{i\}}$$

Then by Fubini's Theorem,

$$\begin{aligned}
\Delta_e(E_e|\mathcal{B}) &\geq \int_{V_J} F(x_1, \ldots, x_d)\mathbf{1}_{E_1} \cdot \cdots \cdot \mathbf{1}_{E_d} d\mu \\
&= \int_0^1 \cdots \int_0^1 \int_{V_J} F(x_1, \ldots, x_d)\mathbf{1}_{E_1} \cdot \cdots \cdot \mathbf{1}_{E_d} d\mu \, dy_1 \ldots dy_d \\
&= \int_{V_J} \int_0^1 \cdots \int_0^1 F(x_1, \ldots, x_d)\mathbf{1}_{E_1} \cdot \cdots \cdot \mathbf{1}_{E_d} dy_1 \ldots dy_d \, d\mu \\
&= \int_{V_J} \prod_{\underline{\omega}_d} F(\underline{\omega}_{[d]}(\mathbf{x}_{[d]}))d\mu
\end{aligned}$$

Taking average over $x_1', \ldots, x_d'$ then $\Delta_e(E_e|\mathcal{B}) \geq \|F\|_{\square^d}^{2^d}$. Combining this with Lemma 1.1.15, we conclude the relationships of the two quantities (in particular largeness of one implies largeness of the other.)

$$\|\mathbf{1}_{E_e} - \mathbb{E}(\mathbf{1}_{E_e}|\mathcal{B})\|_{\square^e} \geq \Delta_e(E_e|\mathcal{B}) \geq \|\mathbf{1}_{E_e} - \mathbb{E}(\mathbf{1}_{E_e}|\mathcal{B})\|_{\square^e}^{2^d} \tag{1.1.5}$$

1.1.3 Bohr Sets

Bohr sets can be regarded as an analogue of subspaces in integer setting where we can run density increment. Given a finite abelian group G, it is not hard to construct a non-degenerate symmetric bilinear form $(x, y) \to x \cdot y$ from $G \times G$ to $\mathbb{R}/\mathbb{Z}$ (see [109], Lem. 4.3). For example, if $G = \mathbb{Z}_N$, we can take $x \cdot y = \frac{xy}{N}$. For each $r \in \hat{G}$, the dual group of G. We can define the character on $\hat{G}$ to be the function $e_r(x) := e(r \cdot x)$. We can identify $\hat{G}$ with the set of characters on G by taking $r \mapsto e_r$. We can define the Fourier transform of f,

$$\hat{f} : \hat{G} \mapsto \mathbb{C}, \hat{f}(r) = \mathbb{E}_{x\in G} f(x)e(r \cdot x)$$

We can think of the Fourier transform as the measurement of the correlation between f and the character e_r. It is natural to put a uniform measure on $f : G \to \mathbb{C}$ and put counting measure on $\widehat{f} : \widehat{G} \to \mathbb{C}$ i.e.

$$\|f\|_p = \mathbb{E}_{x\in G}|f(x)|^p, \ \|\widehat{f}\|_p = \sum_{r\in\widehat{G}} |\widehat{f}(r)|^p.$$

We state the following basic properties which follow directly from direct calculations.

Lemma 1.1.17. *1. (Orthogonality)* $\mathbb{E}_{x\in G} f(x)e(r \cdot x) = \begin{cases} 0 & \textit{if } r \neq 0 \\ 1 & \textit{if } r = 0 \end{cases}$

2. $\langle f, g\rangle = \langle \hat{f}, \hat{g}\rangle$. *In particular* $\|\hat{f}\|_2 = \|f\|_2$
3. $\widehat{f * g} = \hat{f}\hat{g}$.
4. $\|f\|_{U^2} = \|\hat{f}\|_4$

Note that no analogue of property (4) for $U^k, k \geq 3$ is known. It would be very useful if such a formula is found.

Definition 1.1.18 (Bohr's set)**.** *Given* $\rho > 0, S \subseteq \widehat{G}$. *Let* $\|\cdot\|$ *denote the distance to the nearest integer. Define*[9]

$$B(S,\rho) := \{x \in G : \|r \cdot x\| < \rho \,\forall r \in S\}$$

S is referred as the frequency of $B(S,\rho)$. $|S|$ is called the dimension of $B(S,\rho)$, ρ is the width of $B(S,\rho)$.

The notion of Bohr sets is an analogue object of subspaces[10] in $\mathbb{Z}$ or $\mathbb{Z}_N$ or in more general groups . For example, it satisfies nearly closure properties (if it is regular, see [12].) Indeed it can be considered as approximated subgroups [48]. It can also be thought as a metric ball of radius ρ and dimension $|S|$.

An important structural theorem of Bohr sets in $\mathbb{Z}_N$ is that they look like multidimensional arithmetic progressions. This is Bogolyubov's lemma combined with geometry of numbers.

Definition 1.1.19 (Multidimensional arithmetic progression)**.** *A multidimensional arithmetic progression of dimension K with basis $x_1, \ldots x_K$ is a subset of $\mathbb{Z}_N$ or $\mathbb{Z}$ of the form*

$$\{\sum_{i=0}^{K} l_i x_i \;;\; |l_i| \leq m_i.\}$$

Lemma 1.1.20 (see [109])**.** *Let S be a nonempty subset of $\widehat{\mathbb{Z}_N}$ and $\rho > 0$. Then*

- *$B(S,\rho)$ contains an arithmetic progression of size at least $\rho N^{1/|S|}$ centered at* 0.
- *$B(S,\rho)$ contains a proper multidimensional arithmetic progression of dimension $|K|$ and has size at least $(\rho/|S|)^{|S|}N$.*

1.1.4 Density Increment Method

Here we describe the density increment method, sometimes called $L^\infty-$increment method [105]. We have to find a good notion of **Structure** (the set of structural objects in the set/space we considering) and $\|\cdot\|_{\mathbf{S}}$ (*Uniformity norm*, with respect to **Structure**). Let $S \in$ **Structure** and $\delta_S(A)$ denotes the density of A on S. We want to have the following dichotomy:

- (generalized von Neumann [11]) $\|\mathbf{1}_A - \delta_S(A)\|_{\mathbf{S}} < c(\alpha) \Rightarrow A$ contains the required configurations.

[9] equivalently, we may define $B(S,\rho) := \{x \in G : |1 - e(r\cdot x)| < \rho\}$, using $|e(t) - 1| \leq 2|\sin(\pi t)| \leq 2\pi\|t\|_{\mathbb{R}\,/\mathbb{Z}}$.
[10] If G is a vector space then Bohr set is indeed a subspace, the annihilator of S.
[11] See [31], Lemma 3.1

(structure of A can be described in term of $\|\cdot\|_{\mathbf{S}}$ and A contains the required configurations).

– (density increment) $\|\mathbf{1}_A - \delta_S(A)\|_{\mathbf{S}} > c(\alpha) \Rightarrow$ we can find $S' \in$ **Structure**, $\omega(S') \leq \omega(S) + 1$, where ω here is the notion of complexity[12] of structural sets and

$$\delta_{S'}(A) > \delta_S(A) + c(\alpha)$$

for some positive absolute constant $c(\alpha)$ depending only on α.

$\|\cdot\|_{\mathbf{S}}$ should be strong enough (i.e. not too many objects with small $\|\cdot\|_{\mathbf{S}}$ norm) to prove the generalized von Neumann but weak enough (not too many objects with large $\|\cdot\|_{\mathbf{S}}$ norm) to obtain the density increment. To run the density increment method, suppose we cannot find the configurations in the set then we can find S' where A has increased density. However this cannot continue forever as the density cannot exceed 1 So eventually, we arrive at the other case of the dichotomy and we must be able to find the required configuration in the set. We give some examples of density increment dichotomy below.

Example 1.1.21 ([46], Lemma 2.4). *(Roth's theorem in Finite field.) Let $A \subseteq \mathbb{F}_p^n$. If there is $t \neq 0$ such that $\widehat{\mathbf{1}_A}(t) \geq \alpha^2/2$ then there is a subspace of codimension 1 such that A has density on some of its translate at least $\alpha + \alpha^2/4$.*
*In this case we may take **Structure** to be the set of subspaces of $\mathbb{F}_p^n$ and $\|\cdot\|_{\mathbf{S}}$ to be $\|\hat{\cdot}\|_\infty$.*

Example 1.1.22 (Roth's Theorem.). *Suppose $\langle 1_A - \alpha, e(x\frac{r}{N})\rangle \gg \delta$, this means $\mathbf{1}_A - \alpha$ has a linear bias in some direction. We can use equidistribution property[13] of $\frac{rx}{N} \pmod 1$ to partition $[N]$ (up to small error) into long arithmetic progressions of length, say $N^t, t < 1$ for which $\frac{rx}{N}$ is almost a constant $\pmod 1$ on each of these progressions. Then our set will have increased density $\alpha + c(\alpha)$ on one of this progression. We could also do density increment in Bohr sets as in [12].*

Example 1.1.23. *(Rectangles, corners) Finding the correct notion of structures and uniformity norms can be tricky. We may expect the $\|\cdot\|_{\square^2}$ norm to control the number of rectangles to control the number of rectangles. Assume $\|\mathbf{1}_A - \alpha\|_{\square^2}$ is small and we want to show that A contains roughly $\alpha^4 N^4$ rectangles. Expand*

$$\begin{aligned}\|\mathbf{1}_A - \alpha\|_{\square^2}^4 =& \mathbb{E}_{x,x' \in X, y,y' \in Y} \mathbf{1}_A(x,y)\mathbf{1}_A(x',y)\mathbf{1}_A(x,y')\mathbf{1}_A(x',y') - \alpha A_3(x,x',y,y') + \alpha^2 A_2(x,x',y,y') \\ &- \alpha^3 A_1(x,x',y,y') + \alpha^4\end{aligned}$$

This first term is the number of rectangles (divided by N^4). A_1 is a sum of 4 terms of the form $\mathbb{E}_{x,x',y,y'}\mathbf{1}_A(x,y') = \alpha$. The second term is 6 sums of the form $\mathbb{E}_{x,y,x',y'}\mathbf{1}_A(x,y)\mathbf{1}_A(x,y') =$

[12] So S' in some sense is not too small compared to S. For example, if S are subspaces of a given vector spaces, $\omega(S)$ could be the codimension of S

[13] e.g. Dirichlet's diophantine approximation theorem

$\mathbb{E}_x(\mathbb{E}_y \mathbf{1}_A(x,y))^2 \geq (\mathbb{E}_{x,y}\mathbf{1}_A(x,y))^2 = \alpha^2$ *by Cauchy-Schwartz's inequality. However we want to estimate four terms in* A_3 *like* $\mathbb{E}_{x,y,x',y'}\mathbf{1}_A(x,y)\mathbf{1}_A(x',y')\mathbf{1}_A(x',y)$ *or* $\mathbb{E}_{x,y,x',y'}\mathbf{1}_A(x,y)\mathbf{1}_A(x',y)\mathbf{1}_A(x,y')$ *to be* $\leq \alpha^3$ *but there is no reason for this to be true for general, apart from an insufficient trivial bound of* $O(\alpha^2)$*. Indeed, in Shkredov's proof [96] of exponential bound in corner in dimension 2, he puts some uniformity conditions on the structural objects* $E_1 \times E_2$ *(with a method to uniformize general product set* $E_1 \times E_2$*) in a way that we can run the density arguments. This is still open in higher dimension or in general.*

1.1.5 Energy Increment Method

The energy increment method, first appeared in the context of graph theory in the proof of *Szemerédi's Graph Regularity Lemma* [99], gives analogue dichotomy argument as in density increment method. However, this method uses the machinery of ergodic theory (factor) to read the structure in term of $\sigma-$algebras, this is more flexible with the machinery of L^2-space. For example, it is used in showing the existence of prime arithmetic progressions [52] where there is no density argument proof of the result available. Indeed, there are many results in density Ramsey theory where only ergodic proof is known. We also use the energy increment method in chapter 4.

Definition 1.1.24 (Factor)**.** *Let* X *be a finite set, then a* $\sigma-$*algebra on* X *can be given by a (unique) partition of* X *which partitions* X *into atoms. A sigma-algebra* $\mathcal{B}$ *on* X *is sometimes called* ***a factor****. We say that a factor* $\mathcal{B}'$ *is finer than* $\mathcal{B}$ *(or* $\mathcal{B}$ *is coarser than* $\mathcal{B}'$*) if each atom of* $\mathcal{B}$ *can be written as a union of atoms in* $\mathcal{B}'$*. We say that* f *is measurable with respect to* $\mathcal{B}$ *if* f *is constant on each atoms of* $\mathcal{B}$*. Hence* $L^2(\mathcal{B}') \subseteq L^2(\mathcal{B})$*.*
Define $compl(\mathcal{B})$*, the complexity of* $\mathcal{B}$ *to be the smallest number of elements in* $\mathcal{B}$ *that can be used to generate* $\mathcal{B}$*. Note that the number of atoms is at most* $2^{compl(\mathcal{B})}$*.*
Finally define the join $\mathcal{B} \vee \mathcal{B}'$ *the least common refinement of* $\mathcal{B}$ *and* $\mathcal{B}'$*, that is the sigma-algebra whose atoms are given by the intersections* $A \cap A'$ *where* A *is an atom in* $\mathcal{B}$ *and* A' *is an atom in* $\mathcal{B}'$*.*
If we are working on $X \times Y$ *with* $\mathcal{B}_1$ *a* $\sigma-$*algebra on* X *and* $\mathcal{B}_2$ *a* $\sigma-$*algebra on* Y*. We have* $\mathcal{B}_1 \bigvee \mathcal{B}_2 = \{B_1 \times B_2; B_1 \in \mathcal{B}_1, B_2 \in \mathcal{B}_2\}$ *is a factor on* $X \times Y$*. We may also regard* $\mathcal{B}_1, \mathcal{B}_2$ *as factors on* $X \times Y$ *in the trivial way and* $\mathcal{B}_1 \vee \mathcal{B}_2 = \{B_1 \cap B_2; B_1 \in \mathcal{B}_1, B_2 \in \mathcal{B}_2\}$*.*

Definition 1.1.25 (Conditional Expectation)**.** *For a function* $f : X \to \mathbb{C}$ *we define the condition expectation to be the function* $\mathbb{E}(f|\mathcal{B}) : X \to \mathbb{C}$ *by*

$$\mathbb{E}(f|\mathcal{B})(x) := \frac{1}{|\mathcal{B}(x)|} \sum_{y \in \mathcal{B}(x)} f(y),$$

where $\mathcal{B}(x)$ *is the atom containing* x*.*

We see that $\mathbb{E}(f|\mathcal{B})$ is constant on each atom of $\mathcal{B}$. We can view $\mathbb{E}(f|\mathcal{B})$ as a version of f that could be realized by $\mathcal{B}$, in other words the information of f that is captured by $\mathcal{B}$. Indeed, the conditional expectation $\mathbb{E}(f|\mathcal{B})$ is the orthogonal projection of $L^2(X)$ to $L^2(\mathcal{B})$ where $L^2(\mathcal{B})$ is the space of functions in $L^2(X)$ which is $\mathcal{B}-$measurable. This follows from the identity

$$\langle f - \mathbb{E}(f|\mathcal{B}), \mathbb{E}(f|\mathcal{B})\rangle = 0.$$

Here $\langle f, g\rangle = \mathbb{E}_{x\in X} f(x)\overline{g(x)}$. Hence $f - \mathbb{E}(f|\mathcal{B})$ is orthogonal to all structure given by $\mathcal{B}$. Similarly, if $\mathcal{B} \subseteq \mathcal{B}'$ then $\mathbb{E}(f|\mathcal{B})$ is the orthogonal projection of $\mathbb{E}(f|\mathcal{B}')$ to $L^2(\mathcal{B})$. Indeed, we have

$$\langle \mathbb{E}(f|\mathcal{B}') - \mathbb{E}(f|\mathcal{B}), \mathbb{E}(f|\mathcal{B})\rangle = 0.$$

Next, we define the term *energy* which the name "energy increment method" (or "energy boosting argument" in computer science) comes from.

Definition 1.1.26 (Energy). *Let $f : X \to \mathbb{C}$ and $\mathcal{B}$ a factor then the energy of f with respect to $\mathcal{B}$ is given by the L^2-norm $\|\mathbb{E}(f|\mathcal{B})\|_2^2$.*

Energy increment method: Decompose $f : X \to \mathbb{C}$ as

$$f = f_1 + f_2 + f_3 \tag{1.1.6}$$

where $f_1 = \mathbb{E}(f|\mathcal{B})$ is the structural part described in terms of sigma-algebra. $f_2 = f - \mathbb{E}(f|\mathcal{B})$ is the pseudorandom part which is orthogonal to the structures given by $\mathcal{B}$. We usually refer to such decomposition as *Koopman-von Neumann decomposition.* Sometimes, we will also have f_3, the the error term which is small in size or , say, in L^2 norm.

Example 1.1.27. *Let $\mathcal{B}_{triv} = \{\emptyset, X\}$ be the trivial sigma-algebra. Then for any $x \in X, \mathbb{E}(f|\mathcal{B}_{triv})(x) = \frac{1}{X}\sum_{x\in X} f(x) := \alpha$ is a constant function. If X is a dense truly random set then we expect $\|\mathbf{1}_X - \alpha\|_{\square^d}$ to be small. We may obtain the decomposition*

$$\mathbf{1}_X = \alpha + (\mathbf{1}_X - \alpha)$$

Example 1.1.28. *[47] Let $f : \mathbb{F}_p^n \to [-1, 1], S := Spec_\eta(f) := \{r : |\hat{f}(r)| \geq \eta\}$. Hence $|S| \leq 4\delta^2$ (by Plancherel theorem). Let $H = S^\perp$ be the annihilator of S and $\mu_H := \mathbf{1}_H/\mathbb{E}1_H$ be the Haar measure on H. Let $\mathcal{B}$ be the factor generated by the linear functions $r(x) = r^T x, r \in S$. We calculate*

$$f_1(x) := f * \mu_H(x) = \frac{1}{|H|}\sum_{h\in H+x} f(h) = \mathbb{E}(f|\mathcal{B})$$
$$f_2(x) := f - f * \mu_H, |\hat{f_2}(r)| = |\hat{f}(r)||1 - \hat{\mu_H}(r)| \leq 2\eta.$$

The last inequality follows from the fact that $\hat{\mu} \in [0, 1]$ and $r \in S \Rightarrow \hat{\mu}_H(r) = 1$.

Suppose we have the decomposition as (1.1.6) then we plug in into an average like (1.1.2) to prove that the average is bounded above by a constant. With such a decomposition, we basically reduced to proving such estimate for general f to f_1. Such a statement for f_1 is sometimes called ***counting lemma***. In some context like Green-Tao's Theorem, the counting lemma is just the ordinary Szemerédi's theorem. The method that get rid of f_2, f_3 and allows us to work on only f_1 is called ***transference principle***. The term f_2 behaves randomly and are expected to cancel out in the average (1.1.2), so that any term involving f_2 giving only a small error term (this is called ***Generalized von-Neumann theorem.***). The remaining terms involving f_3 are also expected to contribute only small error term, this can be trickier to show.

Dichtomy for energy increment:

If $\|f - \mathbb{E}(f|\mathcal{B})\|_{\square} > \delta$ is large then f correlates with some structure in $\mathcal{B}$ and we can find a finer factor $\mathcal{B}'$ which incorporated some structure in $\mathcal{B}$, with increased energy,

$$\|\mathbb{E}(f|\mathcal{B}')\|_2^2 = \mathbb{E}(f|\mathcal{B})\|_2^2 + c(\delta). \tag{1.1.7}$$

Here $c(\delta)$ is a fixed constant. Since the energy cannot exceed one, this process must stop and we must arrive at a factor $\mathcal{B}$ such that $\|f - \mathbb{E}(f|\mathcal{B})\|_{\square}$ is small.

1.2 Green-Tao's Theorem

The motivation for Green-Tao's theorem is that prime should behave randomly (this is basically still open e.g. the prime tuples conjecture), hence it should contain many arithmetic progressions like a random set.

1.2.1 Green-Tao's Theorem

Theorem 1.2.1. *[52] Let $f : \mathbb{Z}_N \to \mathbb{R}_{\geq 0}$, $f(x) \leq \nu(x)$ for some pseudorandom measure ν (satisfying some pseudorandom conditions with parameter M, say) with $\mathbb{E}_{x \in \mathbb{Z}_N} f(x) \geq \delta$ then*

$$\mathbb{E}_{x,y \in \mathbb{Z}_N} f(x)f(x+y)...f(x+(k-1)y) \geq c(k,\delta) - o_{k,\delta,M}(1).$$

where $c(k,\delta)$ is the same positive constant as in the Szemerédi's Theorem (Theorem 1.1.2).

Note that if we can prove this for fixed k, y then the $k-$prime tuple conjecture would follow. We show below (as in case of Szemerédi's theorem) that this theorem implies the following

Theorem 1.2.2. *If $A \subseteq \mathcal{P}$ has relative positive upper density i.e.*

$$\limsup_{N \to \infty} \frac{|A \cap \mathcal{P}_N|}{|\mathcal{P}_N|} > 0$$

Then A contains infinitely many k-term arithmetic progressions.

Remark 1.2.3. *A way to think of Theorem 1.2.1 in its relation to Theorem 1.2.2 is that we take f to be (the characteristic functions of) a dense subset primes and ν is (a normalized function supported on) the set of almost primes for which primes has positive density. If support of ν is sparse then ν would be unbounded. The point would be that this set of almost primes has some pseudorandom property which says something like "ν and 1 behave similarly" that will allow us to prove the theorem.*

Theorem (1.2.1) $\Rightarrow$ *Theorem* (1.2.2). Suppose $A \subseteq \mathcal{P}$ with positive upper density α then there is some[14] b depending on N such that $\frac{1}{N}\sum_{n\le N} 1_A(N)\overline{\Lambda}_b(n) > c_\alpha$.

Let $[M, \delta N]$ be in the support of the Green-Tao's measure ν (see deinition 2.2.3). Consider $\delta_k = 2^{-k}$ then we have a partition

$$[M, \delta N] = [M, \delta_k \delta N] \cup \bigcup_{m=0}^{k-1} [2^m \delta_k \delta N, 2^{m+1}\delta_k \delta N]$$

so that there is some j such that $\frac{1}{N}\sum_{n\in[2^j\delta_k\delta N, 2^{j+1}\delta_k\delta N]} 1_A(n)\overline{\Lambda}_b(n) > c_{\alpha,k}$. Define

$$f(n) = \begin{cases} c_k 1_A(n)\overline{\Lambda_b}(n) & \text{if } \ 2^j\delta_k\delta N \le n \le 2^{j+1}\delta_k\delta N \\ 0 & \text{otherwise} \end{cases}$$

then there is a constant $c_{\alpha,k} > 0$ such that

$$\mathbb{E}_{n\in[1,2^{j+1}\delta_k\delta N]} f(n) > c_{\alpha,k}$$

for some $j = j(N)$ and some N arbitrarily large. We can verify that $f(n) \le \nu(n)$ (see Chapter 2 for definition of ν and the verification). Let $B = \{x \in [1, 2^{j+1}\delta_k\delta N] : f(x) \ge \frac{c_{\alpha,k}}{2}\}$ then $|B| \ge \frac{c'_{\alpha,k}}{2}N$ so by Theorem 1.2.1, B contains $c_2(\alpha,k)N^2$ arithmetic progressions $Q = \{x, x+y, ..., x+(k-1)y\}$ such that $f(x)...f(x+(k-1)y) \ge (\frac{c'_{\alpha,k}}{2})^k$ and so

$$\frac{1}{N^2}\sum_{x,y\in\mathbb{Z}_N} f(x)...f(x+(k-1)y) \ge \frac{1}{N^2}\sum_{x\in A, y\in\mathbb{Z}_N} f(x)...f(x+(k-1)y) \ge c_2(\alpha,k)\left(\frac{c'_{\alpha,k}}{2}\right)^k.$$

which is greater than 0. The contribution of trivial arithmetic progression (i.e. $y = 0$) is $O(\frac{\log^k N}{N}) = o(1)$. Also if $x, x+y, ..., x+(k-1)y \in [2^j\delta_k\delta N, 2^{j+1}\delta_k\delta N] \subseteq \mathbb{Z}_N$ is an arithmetic progression in $\mathbb{Z}_N$ then they are genuine arithmetic progressions in $\mathbb{Z}$. ($x_j + x_{j+2} - 2x_{j+1} \equiv \pmod N \Rightarrow x_j + x_{j+2} - 2x_{j+1} = 0, \forall j$.) $\square$

[14] See definition of $\overline{\Lambda}_b$ in Chapter 2, for now just think of it as the Mangoldt functions on primes

In the original proof of Green-Tao's Theorem [51], they decompose a function f majorized by ν using the dual function (also referred as generalized character) $\mathcal{D}f$ defined by

$$\langle f, \mathcal{D}f \rangle = \|f\|_{U^d}^{2^d} \tag{1.2.1}$$

For example, consider $\|\cdot\|_{\square^3}$ (where we consider box norm instead, the Gowers norm on $\mathbb{Z}_N$ could be made in a similar form by a change of variable), we have

$$\mathcal{D}f(x,y,z) = \mathbb{E}_{x',y',z'} f(x,y,z')f(x,y',z)f(x',y,z)f(x',y',z)f(x',y,z')f(x,y',z')f(x',y',z')$$

Indeed functions with large U^k norm is then correlate with $\mathcal{D}f$ just by definition. The obvious structure of $\mathcal{D}f$ we see is that they are composed of *functions of lower complexity*[15] of the form $F(x,y)G(y,z)H(z,x)$ and we will use this obvious structures to decompose functions in the regularity lemma using the machinery of sigma-algebras. It is much harder to see what these objects really are; they are actually given by nilsequences. We don't need this in soft inverse arguments as in [52] but we need it if we want to give asymptotes for linear equation in primes [54].

A function g with $\|g\|_{U^d}^* = O(1)$ is called anti uniform function which will be used to show uniform distribution property: If f is uniform then $\langle f, g \rangle$ will be small. This can be regarded of a generalization to uniformity in Roth's theorem when g is taken to be linear exponentials. Hence anti uniform functions can be used to measure the degree of structures in a function. In [52], they prove that these dual functions satisfy the dual function estimate for f_i bounded by ν

$$\|P(\mathcal{D}f_1, \ldots, \mathcal{D}f_K)\|_{U^k}^* = O_{K,d,P}(1) \tag{1.2.2}$$

where P is a polynomial of K variables, degree d. K, d can be arbitrarily large. Correlation condition is applied here in place of infinite linear forms condition which was not available at that time. See section 3.3.

This estimate allows one to prove uniform distribution of ν with respect to these dual functions (Prop. 6.2 in [52]). The dual of Gowers uniformity norms are not algebra norm in general, but they are majorized by $BAC-$norm (defined in [37]) which is a norm satisfying some algebraic properties used in proving transference principle [37] and also in Chapter 3 of this thesis to prove a transference principle. For example, the dual of $\|\cdot\|_{U^2}^*$ norm is not an algebra norm but is majorized by an algebra norm[16]. Indeed, when f is a function on $\mathbb{Z}_N$, we have

$$\|f\|_{U^2}^* = \|\hat{f}\|_{4/3} \leq \|\hat{f}\|_1$$

[15] Here, meaning they depend on less number of variables or constructed from functions which depend on less number of variables. This relates to the notion of *relatively independent joining* which is used in an explicit construction of Host-Kra factors; characteristic factors for multiple recurrence in (1.1.1). See e.g. the appendix B of [8] or section 7 in [71] for expositions.

[16] meaning $\|fg\| \leq \|f\|\|g\|$

Here $\|\hat{\cdot}\|_1$ is the Wiener norm which is an algebra norm by Young's inequality. However it would not be easy to show that $\|\widehat{\mathcal{D}f}\|_1 = O(1)$ when $0 \le f \le \nu$. A more general and systematic study of dual norms and dual functions in this direction is taken in [66].

For the next step, Green-Tao employed the notions of factor and condition expectation machinery from ergodic theory. They use the dual functions to define sigma-algebras, and use the energy increment (if there is a correlation, find dual functions $\mathcal{D}F$ with correlate with f and use it to refine the factor $\mathcal{B}$ with increased energy) to prove estimate like (1.1.7) or (1.1.6) for $0 \le f \le \nu$. The sets generated by these $\sigma-$algebras of $\mathcal{D}F$ is referred in Green-Tao's paper [51] as *generalized Bohr sets*.

Now we state a transference principle which is later simplified in [37] and independently known in language of computer science as *dense model theorem* [87]. The following version is taken from [88].

Theorem 1.2.4 (Transference Principle)**.** *Let ν be a pseudorandom measure. Suppose $\|\nu - 1\|_{U^k} \le \varepsilon' := \exp(-(1/\varepsilon)^{O(1)})$ then there exists $f_1, f_2, f = f_1 + f_2, 0 \le f_1 \le 2, \|f_2\|_{U^k} \le \varepsilon$. Furthermore,*

$$\mathbb{E}_{x,r} f(x)f(x+r)\dots f(x+(k-1)r) = \mathbb{E}_{x,r} f_1(x)f_1(x+r)\dots f_1(x+(k-1)r) + O(\|f_2\|_{U^k})$$

with $O(\|f_2\|_{U^k}) = O(\varepsilon)$.

We will prove a variant of this transference principle in Chapter 3 where we adapt the method in [37]. The quantitative bound $\exp(-(1/\varepsilon)^C)$ comes from the following fact[17] (the explicit bound is not important unless we are trying to extract a quantitative bound in the application):

Fact 1.2.5. *(e.g.[16]) There is a polynomial $P(x) = p_d x^d + p_{d-1}x^{d-1} + \dots + p_0$ such that $|p(x) - x_+| \le \frac{\epsilon}{8}$ for all $x \in [-2/\varepsilon, 2/\varepsilon]$ such that*

$$|p_d|(2/\varepsilon)^d + \dots + |p_1|(2/\varepsilon) + |p_0| \sim \exp(1/\epsilon^C)$$

Now we demonstrate how to apply transference principle to give a quantitative bound for Green-Tao's Theorem, assuming the Szemerédi's Theorem as a blackbox.

Claim 1.2.6. $\|\nu - 1\|_{U^k} = O(1/\omega) + O(\log \omega/\sqrt{\log R})$.

Proof of Claim. Recall that by linear forms condition (see section 2.2.2),

$$\|\nu - 1\|_{U^k}^{2^k} = \sum_{A \subseteq \{0,1\}^k} (-1)^{|A|}(1 + o(1)) = o(1)$$

[17] If we don't need an explicit quantitative bound in the transference principle then we don't need this fact in the proof of the transference principle. See e.g. Appendix B of [] or [] for expositions.

Hence the error term $o(1)$ comes from the the error in the following linear forms estimate condition;

$$(C_\chi \frac{\phi(W)}{W \log R})^m \mathbb{E}_{x \in B} \Lambda_{\chi,R}(\theta_1(x))^2 \dots \Lambda_{\chi,R}(\theta_M(x))^2 = 1 + o(1)$$

Using the estimate obtained in section 2.2, the $o(1)$ term is given by

$$O(\frac{1}{\omega}) + O(\frac{\log \omega}{\sqrt{\log R}})$$

Here R is a small power of N. $\square$

Choose ω, N large enough so that $\|\nu - 1\|_{U^k} \le \epsilon' = \exp(-(1/\epsilon)^{O(1)})$. Choose the ϵ in the error term $O(\epsilon)$ (using the constant in the Remark 1.1.7) so that

$$\epsilon < \exp(-C \exp((1/\delta)^{C_k})) < C_2(k, \delta)$$

Hence $(1/\epsilon)^C \ge \exp(\exp(1/\delta^{C_k}))$. As we chose $\omega \ge \exp(1/\varepsilon^C)$ and $\log N \ge (\log \omega) \exp(1/\varepsilon^C)$ so we choose $N \ge \exp(\exp(\exp(1/\varepsilon^C)))$, hence we have the bound

$$N \ge \exp(\exp(\exp(\exp(\exp(1/\delta^{C_k}))))).$$

Note that there are other variants of transference principles. A natural question to ask would be the properties of the weight ν required to obtain a transference principle or what would be the natural condition of ν, as investigated[18] in [16] in the context of Green-Tao's Theorem. This can open a wider applications of the transference principles. In the case of 3 term arithmetic progressions in the primes, this question is first investigated by Green [45], to approximate f by a bounded function g. In this case, ν is required to satisfy a restriction estimate and a Fourier decay property. A variant of transference principle in this case is obtained in [63] where they approximate f by a function g which is no longer assumed to be bounded but has bounded L^2-norm. This price one pays is that ν need satisfy some more properties such as correlation estimate and some estimate of its L^2-norm. This has applications in obtaining a better quantitative bound of Roth's theorem in the primes. Naslund [78] obtains a transference principle for l^k-bounded function g with some stronger assumptions on ν than [63]. See e.g. [80] for an exposition.

[18] With correlation conditions in the definition of pseudorandom sets, we expect relative Szemerédi's Theorem to hold for pseudorandom sets of density $N^{-o(1)}$. In [16], they remove the correlation conditions and obtain the results for pseudorandom subsets of density N^{-c_k}.

1.3 Szemerédi's Regularity Lemma

The idea of regularity in graph is that the equally distribution of the edge density. Regularity lemma is a kind of structural theorem. It says that up to small error, we can describe any dense graph with some structure (partition of vertex sets which has about the same density), and apart from that information, the graph is just behaves randomly. For a survey of basic properties and applications of this lemma, see the survey [69].

1.3.1 Graph Regularity

Definition 1.3.1. *A bipartite graph $G(A,B)$ is $\varepsilon-$regular if for all $A' \subseteq A, B' \subseteq B, |A'| \geq \varepsilon|A|, |B'| \geq \varepsilon|B|$ then*

$$\left|\frac{|E(A',B')|}{|A'||B'|} - \frac{|E(A,B)|}{|A||B|}\right| \leq \epsilon \tag{1.3.1}$$

Here we don't assume regularity condition for a pair involving a small set. Equation (1.3.1) can be re written as

$$|E(A,B) \cap (A' \times B')| = \frac{|A' \times B'|}{|A \times B|}|E(A,B)| + O(\varepsilon|A \times B|) \tag{1.3.2}$$

This statement would be trivial for small A' or B'. This is used in the functional version [103].

Theorem 1.3.2 (Basic properties of regular pairs; Most degrees into a large set are large [69])**.** *Let (A,B) be an $\epsilon-$regular pair with density δ then Let $Y \subseteq B, |Y| \geq \varepsilon|B|$ then*

$$\left|\{x \in A : deg(x,Y) < (\delta - \varepsilon)|Y|\}\right| < \varepsilon|A|$$

Proof. Let $X = \{x \in A : deg(x,Y) < (\delta-\varepsilon)|Y|\}$, expect Xnot to be too large. Trivially, $|E(X,Y)| < (\delta - \varepsilon)|Y|$. Suppose $|X| \geq \varepsilon|A|$ then by regularity, this is a contradiction. $\square$

Remark 1.3.3. *Let us mention briefly a relation of the notion of graph regularity with the box norm. Suppose $\|\mathbf{1}_G - \mathbb{E}(\mathbf{1}_G|\mathcal{B}_X \vee \mathcal{B}_Y)\|_\square \leq \eta$ with $\mathcal{B}_X = X_1 \cup \cdots \cup X_m$, $\mathcal{B}_Y = Y_1 \cup \cdots \cup Y_n$. We have $\mathbb{E}(\mathbf{1}_G|\mathcal{B}_X \vee \mathcal{B}_Y)(x,y) = \frac{|G\cap(X_i \times Y_i)|}{|X_i||Y_i|} := \delta_{ij}$. By the definition of box norms we can find functions $U(x), V(y)$ such that*

$$\mathbb{E}_{x,y}(\mathbf{1}_G(x,y) - \mathbb{E}(\mathbf{1}_G|\mathcal{B}_x \vee \mathcal{B}_Y)(x,y))U(x)V(y) \leq \eta.$$

Writing $\alpha_i := |X_i|/|X|, \beta_i := |Y_i|/|Y|$ and let $U_i := \{U(x) : x \in X_i\}, V_j := \{V(y) : y \in Y_j\}$, $\gamma_i := ||G \cap (U_i \times V_j)| - \delta_{ij}|U_i||V_j|| \leq \eta$. This implies, for example,

$$\sum_{i,j:\gamma_{ij} > \sqrt{\eta}} \alpha_i\beta_j \leq \sqrt{\eta}$$

Hence we have many X_i, Y_j with $G\big|_{X_i \times Y_j}$ is $\eta-$regular.

Theorem 1.3.4 (Graph regularity lemma [99]). *$\forall \epsilon > 0, \exists K(\epsilon)$ independent of n with the following property: Any graph G_n with $n \geq n_0(\epsilon)$ vertices can be partitioned into vertex classes $V_0, V_1 \ldots, V_K$ such that $|V_0| \leq n/K, |V_i| = |V_j|$ for $1 \leq i,j \leq K$ and all but $\epsilon K^2(i,j)$ pairs give $\epsilon-$regular $G(V_i, V_j)$.*

The proof proceeds via the energy increment method on a partition of vertices, see e.g. Theorem 9.4.1 in [1]. Inspecting the proof, we would need $n_0(\varepsilon)$ to be a tower of height ε^{-5} to obtain an $\varepsilon-$regular partition with $K = \varepsilon^{-1}$. It was shown by Gowers [40] that this tower type bound is necessary. It is true for applications of regularity method that we will have terrible bound. If we want a better quantitative bound, we would need to avoid applying regularity lemma.

Now we illustrate a well known application of regularity lemma.

Theorem 1.3.5 (Triangle removal lemma; Ruzsa-Szemerédi [92]). *$\forall c > 0\ \exists \varepsilon(c) > 0$ with the following property: If G_n is the union disjoint of cn^2 edge-disjoint triangles then it must actually contains at least $\epsilon(c)n^3$ triangles where $\epsilon \to 0$ as $c \to 0$.*

Remark 1.3.6. *Trivially, the number of triangles would be at least cn^2 triangle but this theorem says that it contains much more especially when n is large. In fact, in a higher order of magnitude.*

Thm 1.3.4 $\Rightarrow$ Thm. 1.3.5. Take a random $3-$partition on the vertices to obtain vertex sets W_1, W_2, W_3. By losing a positive fraction of the cn^2 triangles, we can assume that G_n is tripartite. Choose ε, and apply the regularity lemma (Theorem 1.3.1) to our graph, we obtain the regular partition with K classes of vertices.

- Delete the edges between non $\epsilon-$ regular pairs, the number of edges deleted is less than $\varepsilon K^2(n/K)^2 = \varepsilon n^2$ edges.
- We will apply graph regularity lemma for pairs with density at least δ. Delete the edges between pairs with density $\leq \delta$ then we deleted less than $K^2\delta(n/K)^2 = \delta n^2$ edges.

Choose ε, δ much much smaller than c, hence we still have $c'n^2$ triangles in our graph. We obtain a simplified graph with the following property: If there is an edge between V_i and V_j then $G(V_i, V_j)$ is $\varepsilon-$regular and $d(V_i, V_j) \geq \delta$. Now we claim that the number of triangles is at least $c(\varepsilon)n^3$.

Consider vertex sets $V_1 \subseteq W_1, V_2 \subseteq W_2, V_3 \subseteq W_3$ where each pair (V_i, V_j) is regular with density $\geq \delta$. Apply Theorem 1.3.2 to V_2, at least $(1-2\varepsilon)|V_2|$ vertices has degree $\geq (\delta - \varepsilon)|V_1|$ to V_1 and degree $\geq (\delta-\varepsilon)|V_3|$ to V_3. Pick one of such v and assume $\delta - \varepsilon \geq \varepsilon$. One has from the definition of regular pairs,

$$|E\big(N(v)\big|_{V_1}, N(v)\big|_{V_3}\big)| \geq (\delta-\varepsilon)|N(v)\big|_{V_1}||N(v)\big|_{V_3}| \geq (\delta-\varepsilon)\big((\delta-\varepsilon)n/K\big)^2 \geq \varepsilon^3(n/K)^2.$$

This is a lower bound of number of triangles containing v. Since the number of such v is at least $(1-2\varepsilon)|V_2|$, choosing $\delta = 2\varepsilon$ then the number of triangles is at least $(1-2\varepsilon)\varepsilon^3 n^3/K(\varepsilon)^3$. □

Finally we state a functional version of graph regularity lemma [103]. We will prove analogue of this lemma in the weighted hypergraph setting in the main text.

Theorem 1.3.7 (Functional graph regularity lemma [103])**.** *Suppose $f : V_1 \times V_2 \to [0,1]$ is measurable wrt $\mathcal{B}_{1,max}, \mathcal{B}_{2,max}$ and $\varepsilon > 0$. Let $F = F_\varepsilon : \mathbb{N} \to \mathbb{N}$ be an arbitrary increasing function. Then there exists $M = O_{F,\varepsilon}(1)$ and sigma-algebras $\mathcal{B}_i \subseteq \mathcal{B}_i' \subseteq \mathcal{B}_{i,\max}$ on V_i. We obtain the following decomposition of f:*

1. $\mathbb{E}(f|\mathcal{B}_1 \vee \mathcal{B}_2), compl(\mathcal{B}_1), compl(\mathcal{B}_2) \leq M$.
2. $\|\mathbb{E}(f|\mathcal{B}_1' \vee \mathcal{B}_2') - \mathbb{E}(f|\mathcal{B}_1 \vee \mathcal{B}_2)\|_2 \leq \varepsilon$
3. $\|f - \mathbb{E}(f|\mathcal{B}_1' \vee \mathcal{B}_2')\|_\square \leq 1/F(M)$.

Functional version (Theorem 1.3.7) implies graph version (Theorem 1.3.4): Write E for the set of edges between V_1 and V_2. Apply the Theorem 1.3.7 with ε replaced by $\varepsilon^{3/2}$. By equation (1.1.5), the last condition in Theorem 1.3.7 may be translated to

$$|(f - \mathbb{E}(f|\mathcal{B}_1' \vee \mathcal{B}_2'))\mathbf{1}_{A_1 \times A_2}| \leq 1/F(M) \; \forall \, A_1 \in \mathcal{B}_1', A_2 \in \mathcal{B}_2'.$$

Let $J = \lceil 2^M/\varepsilon \rceil$ which is a large number and assume $|V_1|, |V_2| > J$ where each $\mathcal{B}_1, \mathcal{B}_2$ contains at most 2^M atoms. Subdivide each of these atoms into sets of size $\lfloor \frac{|V_i|}{(1+O(\varepsilon))J} \rfloor$ with possibly remaining sets of size (error term) $O(|V_i|/J)$ on each atom. Collect all error term into the set $V_{i,0}$, we obtain a decomposition

$$V_i = V_{i,0} \cup V_{i,1} \cup \cdots \cup V_{i,J}$$

with $|V_{i,0}| = O(\varepsilon|V_i|) + O(2^M|V_i|/J) = O(\varepsilon|V_i|)$. Our goal is to show that $(V_{1,j}, V_{2,k})$ is a regular pair for almost $j, k \geq 1$. Consider the induced bipartite graph $(V_{1,j_1}, V_{2,j_2}, E \cap (V_{1,j_1} \times V_{2,j_2}))$. Let $A_1 \in \mathcal{B}_1, A_2 \in \mathcal{B}_2$ be atoms. We want to show

$$|E \cap (A_1 \times A_2)| = \frac{|E \cap (V_{1,j_1} \times V_{2,j_2})|}{|V_{1,j_1} \times V_{2,j_2}|}|A_1 \times A_2| + O(\varepsilon|V_{1,j_1}||V_{2,j_2}|). \tag{1.3.3}$$

For this, it suffices to find d independent of A_1, A_2 such that

$$|E \cap (A_1 \times A_2)| = d|A_1 \times A_2| + O(\varepsilon|V_{1,j_1}||V_{2,j_2}|). \tag{1.3.4}$$

(To see this, take $A_1 = V_{1,j_1}, A_2 = V_{2,j_2}$ in (1.3.4) and substitute in (1.3.3)). Now $A_1 \times A_2$ is in an atom of $\mathcal{B}_1 \vee \mathcal{B}_2$, one may take $d = \mathbb{E}(\mathbf{1}_E|\mathcal{B}_1 \vee \mathcal{B}_2)$. This is equivalent to

$$\mathbb{E}_{(x,y)\in V_1 \times V_2}(\mathbf{1}_E - \mathbb{E}(\mathbf{1}_E|\mathcal{B}_1 \vee \mathcal{B}_2))\mathbf{1}_{A_1 \times A_2} = O(\frac{\varepsilon|V_{1,j_1}||V_{2,j_2}|}{|V_1||V_2|}) = O(\varepsilon/J^2)$$

Now by our assumption (the conclusion of the functional graph regularity lemma),

$$\mathbb{E}\big[\mathbf{1}_E - \mathbb{E}(\mathbf{1}_E|\mathcal{B}_1' \vee \mathcal{B}_2')\mathbf{1}_{A_1\times A_2}\big] = O(1/F(M)).$$

Take $F(M) := 2^{2M}/\varepsilon^3$ and since $J = O_\varepsilon(1)$, one has $O(1/F(M)) = O(\varepsilon^3/2^{2M}) = O(\varepsilon/J^2) = O_\varepsilon(1)$. Hence it suffices to show

$$\mathbb{E}(\Big|\mathbb{E}(\mathbf{1}_E|\mathcal{B}_1' \vee \mathcal{B}_2') - \mathbb{E}(\mathbf{1}_E|\mathcal{B}_1 \vee \mathcal{B}_2)\Big|\mathbf{1}_{V_{1,j_1}\times V_{2,j_2}}) = O(\varepsilon/J^2) \tag{1.3.5}$$

By Cauchy-Schwartz's inequality, it suffices to show

$$\mathbb{E}(\Big|\mathbb{E}(\mathbf{1}_E|\mathcal{B}_1' \vee \mathcal{B}_2') - \mathbb{E}(\mathbf{1}_E|\mathcal{B}_1 \vee \mathcal{B}_2)\Big|^2\mathbf{1}_{V_{1,j_1}\times V_{2,j_2}}) = O(\varepsilon^2/J^2) \tag{1.3.6}$$

We have from our assumption that

$$\mathbb{E}\Big|\mathbb{E}(\mathbf{1}_E|\mathcal{B}_1' \vee \mathcal{B}_2') - \mathbb{E}(\mathbf{1}_E|\mathcal{B}_1 \vee \mathcal{B}_2)\Big|^2 = O(\varepsilon^3) \tag{1.3.7}$$

So $(\varepsilon^2/J^2)|\{(j_1, j_2) : (1.3.6) \text{ fails}\}| \leq \varepsilon^3$. Hence all but $O(\varepsilon J^2)$ pairs (j_1, j_2) that (1.3.6) fails. □

We can prove the functional graph regularity lemma to prove functional triangle removal lemma stated below. This theorem this says we can clean up a graph with a small number of triangles in a *lower complexity manner* to make it triangle free.

Theorem 1.3.8 (Triangle Removal Lemma [100])**.** *Let $(X, \mu_X), (Y, \mu_Y), (Z, \mu_Z)$ be probability spaces. Suppose $f_1 : X \times Y \to [0, 1]$, $f_2 : Y \times Z \to [0, 1]$, $f_3 : X \times Z \to [0, 1]$ are measurable functions. Let $\varepsilon > 0$. Suppose*

$$\Lambda_3(f_1, f_2, f_3) := \int_X \int_Y \int_Z f_1(x, y) f_2(y, z) f_3(x, z) d\mu_X d\mu_Y d\mu_Z \leq \varepsilon$$

Then we can find measurable functions $\tilde{f}_1 : X \times Y \to [0, 1]$, $\tilde{f}_2 : Y \times Z \to [0, 1]$, $\tilde{f}_3 : X \times Z \to [0, 1]$ such that $\|f_i - \tilde{f}_i\|_1 = o_{\varepsilon\to 0}(1)$ for $i = 1, 2, 3$ such that $\tilde{f}_1\tilde{f}_2\tilde{f}_3$ vanishes entirely, in particular $\Lambda_3(\tilde{f}_1, \tilde{f}_2, \tilde{f}_3) = 0$.

Finally, let us remark that there is also an *arithmetic regularity lemma* for functions $f : [N] \to [0, 1]$ proved in [50] in terms of nilsequences. An application of this lemma in [50] is a proof of Bergelson-Host-Kra's conjecture [7]: If $A \subseteq [N]$ has density α and let $\varepsilon > 0$, then there exists $\gg_{\alpha,\varepsilon} N$ choices of h for which A contains at least $(\alpha^4 - \varepsilon)N$ 4-AP. The case of 3-AP. is proved by Green [49] and it is shown by Ruzsa in an appendix of [7] that the statement is false for 5-AP.

1.3.2 Hypergraph Removal Lemma

Generalizing graph regularity to hypergraph regularity is not trivial. A strong version of hypergraph regularity lemma due to Vojta Rödl-B.Nagel-M.Schacht-J.Skokan [82, 83, 84, 85] and Gowers [35, 36] allows one to prove hypergraph removal lemma and deduce multidimensional Szemerédi's theorem from that. The version we will use later is a stronger functional version due to Tao [104]. It turns out that by consider a projection, we can deduce the general multidimensional Szemerédi's theorem from the corner case (corresponding to $d-$regular hypergraph). This will not work for primes as we don't know if the projection of a prime point is still prime or not. Fortunately, we can use general simplices in the prime cases, allowing us to apply the *Linear forms conditions*. In graph theoretical term, as stated in [16][19], the Linear forms conditions say that our hypergraph has the asymptotically correct count for any 2-blow-up of its subgraph.

Recall a non-degenerated corner is a configuration of the form

$$\{(x_1, \ldots, x_d), (x_1 + s, x_2, \ldots, x_d), \ldots, (x_1, x_2, \ldots, x_d + s)\}$$

with $s \neq 0$, we state the corner theorem.

Theorem 1.3.9. *If $A \subseteq \mathbb{Z}^d$ has positive upper density then A contains a non-degenerate corner.*

The corner Theorem can be proved via the hypergraph removal lemma on $(d+1)-$partite $d-$ regular hypergraph. This is first observed by Solymosi in case $d = 2$ ([97]). First note that in a $(d+1)$-partite d-regular hypergraph with vertex set $X_1, \ldots, X_{d+1}$, *a simplex* is a set of size $d+1$ of $d-$hyperedges $\{(x_i)_{i \in [d+1] \setminus \{j\}}\}_{1 \leq j \leq d+1}$.

Lemma 1.3.10 (Hypergraph Removal Lemma [36]). *In a $(d+1)-$partite $d-$uniform hypergraph H, for any $\epsilon > 0$, there exists $\delta = \delta(\varepsilon) > 0$ where $\delta \to 0$ as $\varepsilon \to 0$ with the following property: Let H be a $(d+1)-$partite $d-$uniform hypergraph with vertex set $X_1, \ldots, X_{d+1}, |X_i| = N_i$ with sufficiently large N_i. Suppose H contains $\leq \delta \prod_{i=1}^{d+1} N_i$ simplices, then for each $i \leq d+1$, one can remove at most $\epsilon \prod_{j \neq i} N_i$ hyperedges of H from $\prod_{j \neq i} X_j$ in such a way that after the removals, one is left with a hypergraph which is simplex-free.*

Proof of Theorem 1.3.9 via Lemma 1.3.10. Let $A \subseteq [N]^d, |A| \geq \alpha N^d$ and consider the corresponding hypergraph $\mathcal{G}_A$ on $(\mathbb{Z}/N)^{d+1}$ (see section 3.1 for the construction with all weights are 1 in our case here.) where we put a d-hyperedge $(y_{i_1}, .., y_{i_d})$ iff all the corresponding d hyperplanes intersects at a point in A. Then we see that each simplex in $\mathcal{G}_A$ will correspond to a corner.

Each corner may be degenerated to a single point in A but this can happen for only $|A| = o(N^{d+1})$ of these simplices. Apply the hypergraph removal lemma with $\epsilon = \alpha N^{-1} d^{-1} \to 0$, as $N \to \infty$.

[19] actually a special case of Linear forms conditions

Suppose $\mathcal{G}_A$ contains $\leq \delta(\varepsilon)N^{d+1}$ corners then for a sufficiently large N, $\epsilon < \frac{\alpha}{2d}$. So if A does not contain a non-degenerated corner then by the Hypergraph Removal Lemma, we would be able to remove less than αN^d of $d-$ hyperedges to make the hypergraph simplex-free but this is impossible as $|A|$ has size $\geq \alpha N^d$. □

Remark 1.3.11. *We can ensure that the constant s in the corner can be choose to be positive by the following trick due to Ben Green [36]: If we choose random point (x, y) from $[-N, N]^2$, since A has upper density α, we have $\mathbb{P}((x, y) \in A \cap [-N, N]^d) \geq c\alpha$ for some $c > 0$ and infinitely many N. If we select a (fixed) point (a, b) at random, let $B = A \cap (a, b) - A$. Since $|A \cap ((a, b) - A)| = \mathbf{1}_A * \mathbf{1}_A(a, b)$ and*[20]

$$\frac{1}{N^2} \sum_{(a,b)} \mathbf{1}_A * \mathbf{1}_A(a, b)\mathbf{1}_{[-N,N]^d} = \frac{1}{N^2}(\sum_{(a,b)} \mathbf{1}_A(a, b))^2\mathbf{1}_{[-N,N]^d} \geq c^2\alpha^2.$$

Hence if we replace A by $B = A \cap ((a, b) - A)$ then $B = (a, b) - B$ and still has positive density for some (a, b) and B is symmetric around A.

Corner Theorem ⇒ Multidimensional Szemerédi's Theorem [36]. Suppose $A \subseteq \mathbb{Z}^r$ has positive upper density. Consider the nontrivial case $F \subseteq \mathbb{Z}^r, |F| = k + 1 \geq r + 1$. By the remark above one may consider F which is symmetric about some point. Choose a point z such that $F - z$ has one point at the origin and $A - z$ still has positive upper density. Also we may assume that $\text{span}\{F\} = \mathbb{Z}^r$: Suppose $\text{span}\{F\} \subseteq V$, a vector space of dimension $r - 1$. Let e_r be a vector outside V then $F \cup \{e_r\} \subseteq (V \cup A) \times \mathbb{Z} \subseteq \mathbb{Z}^r$. which has positive upper density. Then we may add vectors to F so that $\text{span}(F) = \mathbb{Z}^r$ without affecting the question.

Let $\{e_1, ..., e_k\}$ the standard basis of $\mathbb{R}^k$, and define a linear map Φ that maps bijectively from $\{0, e_1, ...e_k\}$ to F. Now suppose $\text{span}\{\phi(e_1), ..., \phi(e_r)\} = \mathbb{Z}^r$ and we can find infinitely many positive integers $M = M(N) \to \infty$ as $N \to \infty$ and $\Phi^{-1}(A - z) \times [M - 1]^{k-r}$ has positive upper density η for some $\eta = \eta(\alpha, F) > 0$ and is mapped into $A - z$ so we can find a large M with at least ηM^k points on $[M - 1]^k$ is mapped into $A - z$ by Φ.

Now we may apply the corner theorem to conclude that $\Phi^{-1}(A - z) \times [M - 1]^{k-r}$ contains a corner $w + c\{0, e_1, e_2, ..., e_k\}, c > 0$. So there is an affine image of $F = z + \Phi(w + c\{0, e_1, e_2, ..., e_k\}) \in A$. □

[20] if we choose $(x, y), (a, b)$ independently, we can think of this as

$$\mathbb{P}((x, y) \in B) = \mathbb{P}((x, y) \in A)\mathbb{P}((x, y) - (a, b) \in -A|(x, y) \in A) = \mathbb{P}((x, y) \in A)\mathbb{P}((a, b) \in (x, y) - A|(x, y) \in A) \geq c^2\alpha^2$$

Chapter 2

Goldston-Yildirim's Sieve and Almost Prime Solutions to Diophantine Equations

In this chapter we prove the main result , Theorem 2.4.2 in section 2.4. We develop some backgrounds, motivations and necessary tools in section 2.1-2.3.

2.1 Backgrounds and Some Classical Results

2.1.1 Basic Prime Number Estimates

Some results on sums of primes may be derived from the Prime Number Theorem using partial summation to convert sums involving an arithmetic function to an integral (see appendix A in [77]). Given $N > 1$, the set of almost prime $\mathcal{P}_{\varepsilon}[N]$ is defined to be the set of positive integers up to N which only have large prime factors, bigger that N^{ε}. Note that each integers in $\mathcal{P}_{\varepsilon}[N]$ can have at most $\lfloor 1/\varepsilon \rfloor$ prime factors.

Theorem 2.1.1 (Partial Summation Formula [77])**.** *Let* $A(x) = \sum_{1 \leq n \leq x} a_n$ *then*

$$\sum_{n=1}^{N} a_n f(n) = \int_0^N f(x) dA(x) = \int_{1^-}^N f(x) dA(x)$$

We collect some well-known facts from analytic number theory (See e.g. chapter 2 in [77])

- Standard bound on the number of divisors:

$$d(n) \leq \exp(\frac{\log n}{\log\log n}) = O(n^{\varepsilon})\ \forall\varepsilon > 0 \tag{2.1.1}$$

- The set of primes is dense in the set of almost primes[1].

$$\left|\frac{\mathcal{P}\cap[N]}{\mathcal{P}_{\varepsilon}\cap[N]}\right| \gg \varepsilon \tag{2.1.2}$$

- The Prime Number Theorem

$$\pi(x) = (1+o(1))\frac{x}{\log x} \tag{2.1.3}$$

- The Prime Number Theorem, equivalent forms:

$$\sum_{p<x}\log p = x + o(x), \prod_{p\leq\omega} p = e^{w(1+o(1))} \tag{2.1.4}$$

-

$$W = \prod_{p\leq\omega} p \Rightarrow W/\phi(W) \approx \log\omega \tag{2.1.5}$$

-

$$\sum_{p<x}\frac{1}{p} = \log\log(10+x) + O(1), x > 0 \tag{2.1.6}$$

-

$$\sum_{p<x}\frac{\log^K p}{p} \ll_K \log^K(10+x), K > 0, x > 0 \tag{2.1.7}$$

- If $\Re(s) > 1, s = 1 + o(1)$ then[2]

$$\prod_{p}(1-p^{-s}) = (1+o(1))(s-1) \tag{2.1.8}$$

2.1.2 Sieve Problems

We briefly describe what the sieve method is, however only for motivation purposes as this section is not required in the main text. We follow closely [107] and [25] in this subsection. Let $P = \prod_{p\leq z} p$ be a squarefree integer, and $\mathcal{D} = \{p|P : p \leq D\}$ be a set of divisors of P. Let a_n be a finitely supported sequence of nonnegative reals[3]. For each prime p, let E_p be a subset of integers such that for each

[1] We actually don't need this estimate, just for motivation purpose.

[2] In Tao's elementary approach with smooth compactly supported function χ ,which we employ here, we would only need (2.1.8) from the Riemann Zeta's function. In previous approach [51] or in more technical approach in [110], more information of Riemann's zeta function or the zero free region is still needed.

[3] e.g. $a_n = \mathbf{1}_{[1,x)}(n)$.

$d \in \mathbb{Z}_+$, $E_d := \cap_{p|d} E_p$ (with $E_1 = \mathbb{Z}$). Define $X_d = \sum_{n \in E_d} a_n$. We wish to estimate, that is to find upper/lower bounds of

$$\sum_n a_n \mathbf{1}_{n \notin \cup_{p|P} E_p} \tag{2.1.9}$$

By the *Inclusion-Exclusion Principle*, we can write the sum in (2.1.9) as $\sum_{d|P} \mu(d) X_d$. However, it turns out that the number of terms in the sum is too large causing the error terms to accumulate too quickly. One can do better by truncating the sum, working in a way that only X_d for $d \in \mathcal{D}$ are known or exploited. This is a linear programming problem and one can restate the problem using the so-called *linear programming duality*.

Problem 2.1.2 (Sieve Problem)**.** *Define an (normalized) upper bound sieve to be a function $\nu^+ : \mathbb{Z} \to \mathbb{R}$ of the form $\nu^+ = \sum_{d \in \mathcal{D}} \lambda_d^+ \mathbf{1}_{E_d}$ for some $\lambda_d^+ \in \mathbb{R}$, such that*

$$\nu^+(n) \geq \mathbf{1}_{n \notin \cup_{p|P} E_p}(n), \tag{2.1.10}$$

then the supremum of (2.1.9)*, subject to the condition that only $X_d, d \in \mathcal{D}$ are known, equals to the infimum of*

$$\sum_{d \in \mathcal{D}} \lambda_d^+ X_d, \tag{2.1.11}$$

where the infimum is over (λ_d^+) that constitutes an upper bound sieve. Usually X_d will be of the form $g(d)X + r_d$ where g is a multiplicative function, $0 \leq g \leq 1$, X is a quantity independent of d and r_d is negligible when d is restricted to a small range $\mathcal{D}$. Hence one is to minimize

$$\sum_{d \in \mathcal{D}} \lambda_d^+ g(d)$$

Observe that a sequence (λ_d^+) such that $\lambda_1^+ \geq 1, \sum_{d|n} \lambda_d^+ \geq 0 \; \forall n|P$ will form an upper bound sieve. Such a sequence is called *upper bound sieve coefficients.* Analogue problems for lower bound sieves may be similarly stated.

Now an important kind of upper bound sieve is the *Selberg's upper bound sieve* developed by Selberg in 1940s, see e.g. [15] . The idea of Selberg's upper bound sieve comes from observing that if P is any squarefree number and $(\rho_d)_{d|P}$ are arbitrary real numbers with $\rho_1 = 1$ then $(\sum_{d|P} \rho_d \mathbf{1}_{E_d})^2$ is an upper bound sieve as it is 1 outside $\bigcup_{p|P} E_p$. Equivalently, the sequence

$$\lambda_d^+ = \sum_{\substack{d_1, d_2 \\ lcm[d_1, d_2] = d}} \rho_{d_1} \rho_{d_2}, \text{ where } d|P \tag{2.1.12}$$

is a sequence of *upper bound sieve coefficients.* Set $D = R^2$ and we assume that ρ_d is supported on $\mathcal{D} = \{d|P : d \leq D\}$. The key advantage of Selberg's upper bound sieve is that $(\rho_d)_{1 \leq d \leq R}$ are real numbers and the sieve problem reduced to problem of optimizing quadratic forms. It turns out Selberg's sieve usually already gives good results compared to harder optimizing general upper bound sieve coefficients.

We study the following choice of ρ_d. The optimal Selberg's weight is given by, roughly, by[4] $\mu(d)\log(R/d)$. We will use the following variant obtained by Tao.

$$\rho_d := \mu(d)\chi(\frac{\log d}{\log R}) \tag{2.1.13}$$

where χ is some smooth compactly supported function.

Finally we state an important lemma in sieve theory. This lemma may be used to derive formulas for the number of solutions of various diophantine equations or counting patterns in almost primes, e.g. one can prove the analogue of the Hardy-Littlewood almost prime tuples conjecture. See [107] for details. One limitation of sieve method is that we can find asymptotic for system of linear equations where the number of equations bounded by a set of parameters, contrary to more modern results involving prime numbers[52].

Theorem 2.1.3 (The fundamental lemma of sieve theory ([107], Cor 19.)). *Let $z = D^{1/s}$ for some $D, s > 1$. Suppose g is a multiplicative function and $\kappa > 1$ (called sieve dimension) such that g obeys the bound*

$$g(p) \leq \frac{\kappa}{p} + O_\kappa(\frac{1}{p^2}), \;\; g(p) \leq 1 - c_\kappa.$$

for some small constant c_κ. In particular, for $2 < \omega < z$,

$$V(\omega) = \prod_{p<\omega}(1-g(p)) \lesssim_\kappa \left(\frac{\log z}{\log \omega}\right)^\kappa V(z)$$

Let E_p be a set of integers and $(a_n)_{n\in\mathbb{Z}}$ be a finitely supported sequence of non-negative real such that

$$\sum_{n\in E_d} a_n = Xg(d) + r_d, \;\; X > 0, \;\; r_d, \in \mathbb{R}, \;\; E_d = \cap_{p|d} E_p$$

for all squarefree $d \leq D$. Then

$$\sum_{n\notin\bigcup_{p\leq z} E_p} a_n = (1 + O_\kappa(e^{-s}))XV(z) + O(\sum_{d\leq D;\nu(d)^2=1} |r_d|).$$

[4]This is a form used by Goldston-Yildirim in their works on small gaps in the primes.

2.2 A Pseudorandom Measure Majorizing the Primes

In this subsection we construct a pseudorandom measure ν similar that used in the original proof of Green-Tao's Theorem. This subsection is mostly for expositional purpose.

2.2.1 The W-Trick

The primes has the obvious structure that they can only live in some residue classes (for example, no primes except 2 and 3 are in $2 \pmod 6, 3 \pmod 6$ i.e. they are not uniformly distributed). Let[5] $W := \prod_{p \leq \omega} p \approx e^{\omega}$, consider

$$\mathcal{P}_{W,b}[N] := \{1 \leq n \leq N : Wn + b \in \mathcal{P}\}$$

We can see from The Prime Number Theorem in arithmetic progressions that $\mathcal{P}_{W,b}$ is uniformly distributed among residue classes $\pmod p$, for all $p \leq \omega$. Using the correspondence

$$A \subseteq \mathcal{P}[N] \leftrightarrow A'' := A \cap \mathcal{P}_{W,b}[N/W]$$

One is able to get rid of the local factors arising from small primes $p \leq \omega$ while primes are much more uniformly[6]. distributed on large residue classes.

There are two viewpoints one could think of W (see Theorem 2.2.1 below). First, we can think of W as a function of x which goes to infinity sufficiently slowly. The error term in this case is of the form $o_{x\to\infty}(1)$. Another viewpoint, we could also think of W as a fixed sufficiently large constant and the error term would be of the form $o_{w\to\infty,x\to\infty}(1)$. The latter viewpoint is important in calculating explicit bounds. To make this precise, we state the following theorem stated in [106].

Theorem 2.2.1 (Overspill Principle[7][106]). *Let $F(w,x) : \mathbb{Z}_+ \times \mathbb{R} \to \mathbb{R}$ then the following are equivalent:*

1. *For a fixed $\epsilon > 0$ there exists w_ϵ such that for each fixed $w \geq w_\epsilon$*

$$|F(w,x)| \leq \epsilon + o_{w,x\to\infty}(1) = o_{\omega\to\infty,x\to\infty}(1)$$

2.

$$F(w,x) = o_{x\to\infty}(1)$$

whenever $w = w(x) \in \mathbb{Z}$ is sufficiently slowly growing to infinity. (in the sense that there is a function $w_0(x) : \mathbb{R}_+ \to \mathbb{Z}_+$ defined in the proof such that $w(x) \leq w_0(x)$.)

[5] Indeed ω cannot be bigger than $\log N$

[6] Up to error term $o_{\omega\to\infty}(1)$.

[7] The name comes from non-standard analysis. We write out explicitly the parameters in $o(1)$ in the statement.

Proof. Assume (1), then for each natural number n, and $w \leq n$ we can find x_n such that

$$|F(w,x)| \leq \frac{2}{n}$$

for all $x \geq x_n$ and $n \geq w \geq w_{1/n}$. Define the function

$$w_o : \mathbb{R}_+ \to \mathbb{N}, \quad w_0(x) := \begin{cases} n & \text{where } n \text{ is the largest integer such that } x_n \leq x \\ 1 & \text{if } x < x_1. \end{cases}$$

WLOG, we may choose x_n in a way that it is increasing in n. Define

$$w(x) = \begin{cases} w_{1/n} & \text{if } x_{n+1} < x \leq x_n \\ 1 & \text{if } x < x_1. \end{cases}$$

Hence $w(x) \leq w_0(x)$. Also $F(w,x) = o(1)$.

Conversely, assume (2) holds but (1) fails. Then there is an $\epsilon > 0$ such that for any positive integer n, there is $w_n \geq n$ such that $|F(w_n, x_n)| > \epsilon$ for arbitrarily large x_n. Letting w_n going to infinity, we can find a sequence $\{x_n\}$ going to infinity such that $|F(w_n, x_n)| \geq \epsilon$ for all n. Since $w(x)$ goes to infinity, increasing x_n as necessary, we can ensure that $w(x) \geq w_n$ for all $x \geq x_n$ and all n. We see that $|F(w,x)| \geq \epsilon$ at $x = x_n$ for all n which contradicts (2). □

2.2.2 Pseudorandomness Conditions

In this section we construct a pseudorandom measure ν majorizing the Mangoldt function Λ concentrated on primes. We prove certain correlation conditions which is a bit more general than the ones obtained in [51], [52](see also the exposition [17]), however the proof is essentially the same[8].

We defined the following modified Mangoldt function corresponding to W-trick. Let $W = \prod_{p \leq \omega} p$, $(b, W) = 1$. Define

$$\tilde{\Lambda}_b(n) = \begin{cases} \frac{\phi(W)}{W} \log(Wn + b) & \text{if } Wn + b \text{ is prime} \\ 0 & \text{otherwise.} \end{cases} \tag{2.2.1}$$

The factor $\phi(W)/W$ is for normalized purpose; we have $1/N \sum_{n \leq N} \tilde{\Lambda}_b(n) = 1 + o(1)$. We construct a pseudorandom measure, by making use of the Goldston-Yildirim division sum [42]

$$\Lambda_R(n) = \sum_{d|n, d \leq R} \mu(d) \log(R/d).$$

[8] A stronger analogue pseudorandom condition for ν is obtained in [110] in particular they need the polynomial to stay in the length about N so the range average over t has to be of the form $N^{o(1)}$. This is a nontrivial bound on the size of t.

Note that if $N \geq n \geq R$ is a prime or more generally if n has no prime factor $\leq R$, then $\Lambda_R(n) = \log R \gtrsim \tilde{\Lambda}(n)$; choosing $R = N^\eta$ for some $\eta > 0$. We state below two technical results which show that the function $\Lambda_R(n)$ is concentrated on the set of almost primes, i.e. numbers havong only large prime factors. As we don't need these facts for our main results, we omit the proofs.

Theorem 2.2.2. *[81] Let $N^{c_0} < R \leq \sqrt{N/q}(\log N)^{-C}$ and q be a prime,$q = R^\beta, \beta < c_0$ where c_0, C are suitably chosen. Then*

$$\sum_{\substack{N<n\leq 2N \\ q|n}} \Lambda_R(n)^2 \ll \frac{\beta}{q} \sum_{N<n\leq 2N} \Lambda_R(n)^2$$

In particular, if $P(N^\eta) = \prod_{p\leq N^\eta} p$ then

$$\sum_{\substack{N<n\leq 2N \\ gcd(n,P(N^\eta))>1}} \Lambda_R(n)^2 \ll \frac{\beta}{q} \sum_{N<n\leq 2N} \Lambda_R(n)^2$$

In [52], the following variant of Λ_R is introduced,

$$\Lambda_{\chi,R}(n) = \sum_{d|n} \mu(d)\chi(\frac{\log d}{\log R}) \tag{2.2.2}$$

The point is to replace $\log_+(R/d)$ with a smooth approximation $\chi(\frac{\log d}{\log R})$ where χ is a smooth compact supported, bounded function. Then instead of using the contour integral, one can apply Fourier transform. Then we can truncate the integrals over bounded interval obtaining error terms $o(1)$ due to smoothness of χ and rapid decay of its Fourier transform. We will follow this more elementary approach as opposed to that of [42] based on evaluating certain contour integrals.

Definition 2.2.3 (Green-Tao measure)**.** *Let $\chi : \mathbb{R} \to [0,1]$ supported on $[-1,1], \chi(0) = 1, C_\chi = \int_0^1 |\chi'(t)|^2 dt$ which we may assume to be 1. Let $R = N^{k^{-1}2^{-k-5}}, k \geq 3$. Let $\varepsilon > 0$ be a small constant. Define $\nu_b : \mathbb{Z}_N \to \mathbb{R}_+$*

$$\nu_b(n) = \begin{cases} \frac{\phi(W)}{W}\frac{\Lambda_{\chi,R}(Wn+b)^2}{C_\chi \log R} & \text{if } \varepsilon N \leq n \leq 2\varepsilon N \\ 1 & \text{Otherwise} \end{cases} \tag{2.2.3}$$

Remark 2.2.4. *Note that $\nu_b = \nu_b^{(N)}$ depends on N but we will not write script N for simplicity. Sometimes we also drop the subscript b as all our estimates are independent of it.*

We summarize important properties of ν:

– In general ν could become unbounded as $N \to \infty$. However, by the so-called *linear form* *conditions* (see Definition 2.2.5 below)

$$\mathbb{E}_x \nu(x) = 1 + o(1).$$

– We have
$$\tilde{\Lambda}(n) \lesssim_k \nu(n).$$
To see this[9], we may only check n for which $Wn+b$ is prime. Then $\Lambda_{\chi,R}(Wn+b) = \log R = k^{-1}2^{-k-5}\log N$. Now assume N is sufficiently large (and ω is sufficiently slowly growing or constant) then
$$k^{-1}2^{-k-5}\log N \geq k^{-1}2^{-k-6}\log(WN+b).$$
Then
$$k^{-1}2^{-k-6}\frac{\phi(W)}{W}\log(Wn+b) \leq k^{-1}2^{-k-6}\frac{\phi(W)}{W}\log(WN+b) \leq \frac{\phi(W)}{W}\log R = \nu(n).$$

– If $Wn+b$ is prime in , say, $[\varepsilon_1 N, \varepsilon_2 N]$ then $\nu(n) \approx_W \log N$. Here we may choose W to be a (large) constant.

Now we describe the pseudorandom conditions we will need later. The first one roughly says that if the linear forms L_i are not rational multiple of each other then the events that $L_i(x)+b_i$ are almost primes are independent.

Definition 2.2.5 (Linear Forms Condition)**.** *Let $m_0, t_0 \in \mathbb{N}$ be parameters then we say that ν satisfies $(m_0,t_0)-$ linear form condition if for any $m \leq m_0, t \leq t_0$, suppose $\{a_{ij}\}_{\substack{1\leq i\leq m\\ 1\leq j\leq t}}$ are subsets of integers and $b_i \in \mathbb{Z}_N$. Given m (affine) linear forms $L_i : \mathbb{Z}_N^t \to \mathbb{Z}_N$ with $L_i(\mathbf{x}) = \sum_{1\leq j\leq t} a_{ij}x_j + b_i$ for $1 \leq i \leq m$ be such that each L_i is nonzero and they are pairwise linearly independent over rational. Then*
$$\mathbb{E}_{\mathbf{x}\in\mathbb{Z}_N^t}\prod_{1\leq i\leq m}\nu(L_i(\mathbf{x})+b_i) = 1 + o_{N\to\infty,m_0,t_0}(1) \tag{2.2.4}$$

This is a very general phenomena. Two important special cases are given below.

Example 2.2.6. $\|\nu-1\|_{U^d}^{2^d} = o(1)$, *where $\|\cdot\|_{U^d}$ is the uniformity norm discussed in section 1.1.1.*

Example 2.2.7. *If $f \leq \nu$ then consider the dual function defined in (1.2.1) associated with $\|\cdot\|_{U^2}$ norm then*
$$|\mathcal{D}_2 f(x)| \leq \mathbb{E}_{a,b}\nu(x+a)\nu(x+b)\nu(x+a+b) = 1 + o(1)$$
This is referred as ***bounded dual condition***[10]*, saying that the structural component (used in [51]) in the decomposition is bounded.*

[9] W-trick is essential here, we don't expect this to holds for prime itself with pseudorandom ν as primes are not uniformly distributed among residue classes.

[10] This is the only application of linear form conditions with nonzero constant terms in the original Green-Tao's proof[51].

Now we state the so-called *correlation condition* which control some kind of mild correlation of Λ_R by functions τ. τ itself may not be bounded but it has bounded moments. The proof of linear forms condition and correlations for ν are not much different, this may be harder for primes. Roughly speaking we have for $h \neq 0$,

$$\mathbb{E}_x \nu(x)\nu(x+h) \leq \tau(h) \approx \exp(\sum_{p>\omega, p|h} 1/\sqrt{h}).$$

Note that if h has a large number of divisors then $\tau(h)$ can be arbitrarily large. As opposed, there is a strong kind of correlation that we cannot control, i.e. higher moment of ν,

$$\mathbb{E}_x \nu(x)^2 \sim \log N \to \infty.$$

Definition 2.2.8 (Correlation Condition). [11] *We say that a measure ν satisfies $(m_0, m_1, ..., m_{l_2})-$ correlation condition if there is a function $\tau : \mathbb{Z}_N \to \mathbb{R}_+$ such that*

1. $\mathbb{E}(\tau(x)^m : x \in \mathbb{Z}_N) = O_m(1)$ *for any* $m \in \mathbb{Z}_+$
2. *Suppose*
   - $\phi_i, \psi^{(k)} : \mathbb{Z}_N^t \to \mathbb{Z}_N (1 \leq i \leq l_1, 1 \leq k \leq l_2, l_1 + l_2 \leq m_0)$ *are all pairwise linearly independent linear forms over* $\mathbb{Q}$.
   - *For each* $1 \leq g \leq l_2, 1 \leq j < j' \leq m_g$ *we have* $a_j^g \neq 0$, *and* $a_j^{(g)}\psi^{(g)}(\mathbf{x})+h_j^{(g)}, a_{j'}^{(g)}\psi^{(g)}(\mathbf{x})+ h_{j'}^{(g)}$ *are different (affine) linear forms.*

   then, we have

$$\mathbb{E}_{\mathbf{x}\in\mathbb{Z}_N^d} \prod_{k=1}^{l_1} \nu(\phi_k(\mathbf{x})) \prod_{k=1}^{l_2} \prod_{j=1}^{m_k} \nu(a_j^{(k)}\psi^{(k)}(\mathbf{x})+h_j^{(k)}) \leq \prod_{k=1}^{l_2} \sum_{1\leq j<j'\leq m_k} \tau\Big(W(a_{j'}^{(k)}h_j^{(k)}-a_j^{(k)}h_{j'}^{(k)})+(a_{j'}^{(k)}-a_j^{(k)})b\Big) \tag{2.2.5}$$

where $W = \prod_{p\leq\omega} p$.

Lemma 2.2.9. *Let* $B \subseteq \mathbb{Z}_N^d$ *be a box of length* $\geq R^{10M}$ *where* $M = m_0 m_1 \ldots m_{l_2}, R = N^{k^{-1}2^{-k-5}}$ *then*

$$\mathbb{E}_{\mathbf{x}\in B} \prod_{k=1}^{m_0} \Lambda_{\chi,R}(W\phi_k(\mathbf{x})+b)^2 \prod_{k=1}^{l_2}\prod_{j=1}^{m_k} \Lambda_{\chi,R}(W\cdot(a_j^{(k)}\psi^{(k)}(\mathbf{x})+h_j^{(k)})+b)^2 \tag{2.2.6}$$
$$= (1+O(\frac{\log\omega}{\sqrt{\log R}}))\exp(O_m(1/\omega))(\frac{W\log R}{\phi(W)})^M \prod_{k=1}^{l_2}\prod_{p|\Delta_k}(1+O_M(p^{-1/2}))$$

[11]This lemma is first invented in [51] to prove dual function estimates with K arbitrarily large.

where

$$\Delta_k := \prod_{1\le j<j'\le m_k} \left(W\cdot (a_{j'}^{(k)} h_j^{(k)} - a_j^{(k)} h_{j'}^{(k)}) + (a_{j'}^{(k)} - a_j^{(k)})b \right)$$

and

$$M = m_0 + m_1 + \cdots + m_j$$

$$[M] = \bigcup_{j=1}^{m_0} I_j \cup \bigcup_{j=1}^{m_j} I_{m_j}, \quad I_j = \{j\} \ \textit{for } j \le m_0, \quad I_{m_j} = (M_{j-1}, M_j]$$

$$\psi_i^{(k)} := \begin{cases} \phi_i & \textit{if } i \in I_j, j \le m_0 \\ \psi^{(k)} & \textit{if } i \in I_k, k > m_0 \end{cases}$$

Now we verify this lemma as in [51] or [17]. Write

$$\theta_i(\mathbf{x}) = \begin{cases} W\phi_i(\mathbf{x}) + b & \text{if } i \in I_j, j \le m_0 \\ W(a_i^{(k)}\psi_i^{(k)} + h_i^{(k)}) + b & \text{if } i \in I_{m_k}, m_k > m_0. \end{cases}$$

Expand LHS of (2.2.6)

$$\mathbb{E}_{\mathbf{x}\in B} \prod_{m_0}^{k} \prod_{i\in I_m} \Lambda_{\chi,R}(\theta_i(\mathbf{x}))^2 \tag{2.2.7}$$

$$= \sum_{\mathbf{a},\mathbf{b}\in\mathbb{N}^M} \left(\prod_{i=1}^{M} \mu(a_i)\mu(b_i)\chi(\frac{\log a_i}{\log R})\chi(\frac{\log b_i}{\log R}) \right) \mathbb{E}_{\mathbf{x}\in B} \left(\prod_{i=1}^{M} \mathbf{1}_{a_i,b_i|\theta_i(\mathbf{x})}(\mathbf{x}) \right)$$

Observe that only the last term depends on $\mathbf{x}$. Recall $D = \operatorname{lcm}[a_1, b_1, \ldots, a_M, b_M] \le R^{2M}$ and B has each side of length $\ge R^{10M}$, we can approximate

$$\mathbb{E}_{\mathbf{x}\in B} \left(\prod_{i=1}^{M} \mathbf{1}_{a_i,b_i|\theta_i(\mathbf{x})}(\mathbf{x}) \right) = \mathbb{E}_{\mathbf{x}\in \mathbb{Z}_D^t} \left(\prod_{i=1}^{M} \mathbf{1}_{a_i,b_i|\theta_i(\mathbf{x})}(\mathbf{x}) \right) + O(R^{-8M})$$

To see this, consider a slightly smaller box B' whose the lengths of all its dimensions are all divisible by D and the length of each side of B' differs from the length of the corresponding side of B by $O(R^{2M})$. The total error in 2.2.7 when changing average in B to average over B' (which is the same as the average in $\mathbb{Z}_D^t$) is $O(\frac{\log^{2M} R}{R^{6M}})$.
For $X \subseteq [M]$, define the local factor

$$\omega_X(p) = \mathbb{E}_{\mathbf{x}\in\mathbb{Z}_p^t} \prod_{i\in X} \mathbf{1}_{\theta_i(\mathbf{x})\equiv 0 \pmod p},$$

$$\omega_X = \mathbb{E}_{\mathbf{x}\in\mathbb{Z}_D^t} \prod_{i\in X} \mathbf{1}_{a_i,b_i|\theta_i(\mathbf{x})},\ D = \text{lcm}[a_1, b_1, \ldots, a_M, b_M].$$

Let $D = p_1 \ldots p_h$. Using the Chinese Remainder Theorem, we rewrite the system of equations

$$\theta_i(\mathbf{x}) \equiv 0 \pmod{a_i},\quad \theta_i(\mathbf{x}) \equiv 0 \pmod{b_i},\quad 1 \le i \le m$$

as

$$\theta_i(\mathbf{x}) \equiv 0 \pmod{p_j},\ 1 \le i \le m, 1 \le j \le h.$$

Hence

$$\omega_X = \prod_p \omega_X(p).$$

We have the following *local factor estimate* that will be used later.

Lemma 2.2.10. *[Local factor estimate]*

1. $\omega_\emptyset(p) = 1$.
2. $p \le \omega, X \ne \emptyset \Rightarrow \omega_X(p) = 0$.
3. $|X| = 1 \Rightarrow \omega_X(p) = 1$.
4. *Suppose* $p > \omega$ *and* $X \subseteq I_k, |X| > 1$. *If* $p|\Delta_k$ *and* $|X| = 2$ *then* $\omega_X(p) = p^{-1}$. *If* $p|\Delta_k$ *and* $|X| > 2$ *then* $\omega_X(p) \le p^{-1}$. *If* $p \nmid \Delta_k$ *then* $\omega_X(p) = 0$.
5. *If* $p > \omega$ *and* $\exists k_1 \ne k_2$ *such that* $X \cap I_{k_1}, X \cap I_{k_2}$ *is nonempty then* $\omega_X(p) \le p^{-2}$.

Proof. (1) is trivial. (2) follows from the fact that if $p \le \omega, j \in X$ then $W \cdot (a_j^{(k)}\psi_j^{(k)} + h_j^{(k)}) + b \equiv b \ne 0 \pmod{p}$. To see (3), if $p > \omega, X \subseteq I_k, |X| = 1$,say $X = \{j\}$, then we can write

$$\omega_X(p) = \mathbb{E}_{\mathbf{x}\in\mathbb{Z}_p^t} \mathbf{1}_{W\cdot(a_j^{(k)}\psi^{(k)}a_j^{(k)}+h_j^{(k)})+b\equiv 0 \pmod{p}} = p^{-1}$$

To verify (4) and (5), assume $|X| > 1$ and $j, j' \in X$ since

$$p|W \cdot (a_j^{(k)}\psi_j^{(k)} + h_j^{(k)}) + b, p|W \cdot (a_{j'}^{(k)}\psi_{j'}^{(k)} + h_{j'}^{(k)}) + b$$

Then $p|W \cdot (a_j^{(k)}h_j^{(k)} - a_j^{(k)}h_{j'}^{(k)}) + (a_{j'}^{(k)} - a_j^{(k)})b$ so $p|\Delta_k$. Hence $p \nmid \Delta_k \Rightarrow \omega_X(p) = 0$.

Now assume $p|\Delta_k$, then the condition that $p|\theta_i(x)\forall i \in X$ and $|X| = 2$ could be reduced to $p|\theta_j(x)$ for some $j \in X$ if $p > W, p > a_i \forall i \in X$, which is true if ω is chosen sufficiently large. Hence

$$\begin{aligned}\omega_X(p) &= \mathbb{E}_{\mathbf{x}\in\mathbb{Z}_p^t} \prod_{i\in X} \mathbf{1}_{W\cdot(a_i^{(k)}\psi^{(k)}a_i^{(k)}+h_i^{(k)})+b\equiv 0 \pmod{p}} \\ &= \mathbb{E}_{\mathbf{x}\in\mathbb{Z}_p^t} \mathbf{1}_{W\cdot(a_j^{(k)}\psi^{(k)}a_j^{(k)}+h_j^{(k)})+b\equiv 0 \pmod{p}} = p^{-1}\end{aligned}$$

If $|X| > 2$ then we can crudely bound this by $\omega_{\{j,j'\}}$ for some j, j' chosen so that p divide the factor in Δ corresponding to j, j'.

Now we verify (5). Assume $j \in X \cap I_{k_1}, j' \in X \cap I_{k_2}$. For $i = 1, 2$, write

$$a_j^{(k_i)}\psi_j^{(k_i)}(\mathbf{x}) = \sum_{s=1}^{t} L_{k_i,s}x_s$$

as $p \nmid W$, our condition becomes

$$\sum_{s=1}^{t} L_{k_i,s}x_s = -W^{-1}b - W^{-1}h_{j_i}^{(k_i)} \pmod p, \quad i = 1, 2.$$

By our assumption that $(L_{k_1,s})_{1\le s\le t}, (L_{k_2,s})_{1\le s\le t}$ are not rational multiple of each other, we claim that they are also linearly independent over $\mathbb{Z}_p$. Assume indirectly that $L_{k_1,s} = rL_{k_2,s} \pmod p$ for some rational r. Then

$$a_i/b_i = rc_i/d_i \pmod p.$$

Hence for each $1 \le i \le t, \lambda = (a_i d_i)(b_i c_i)^{-1} \pmod p$, i.e. $a_1 b_1 b_i c_i \equiv b_1 c_1 a_i d_i \pmod p$. But if ω is sufficiently large so that $|a_i|, |b_i|, |c_i|, |d_i| \le \frac{\omega^{1/4}}{2}$. then $|a_1 d_1 b_i c_i - b_1 c_1 a_i d_i| \le |a_1 d_1 b_i c_i| + |b_1 c_1 a_i d_i| < \omega \le p$. Hence $a_1 d_1 b_i c_i = b_1 c_1 a_i d_i$ $\forall 1 \le i \le t$. This is a contradiction. Hence the set of solutions of $\theta_j(\mathbf{x}) = \theta_k(\mathbf{x}) \equiv 0 \pmod p$ is contained in the intersection of two skew-affine subspaces of $\mathbb{Z}_p^t$ and hence has cardinality $\le p^{t-2}$. □

Now let ψ be the inverse Fourier transform of $e^x\chi(x)$ i.e.

$$\chi(x) = \int_{\mathbb{R}} \psi(t)e^{-x(1+it)}dt$$

Since $e^x\chi(x)$ is smooth and has compact support, ψ is smooth and rapidly decays. In particular, for any $A > 0$, $|\psi(t)| = O_A((1+t)^{-A})$. Let $I = [-\sqrt{\log R}, \sqrt{\log R}]$ then

$$\chi(\frac{\log c}{\log R}) = \int_I c^{-\frac{1+it}{\log R}}\psi(t)dt + O(c^{-1/\log R}\log^{-A} R) \tag{2.2.8}$$

for any $A > 0$. Observe that $\chi(\frac{\log c}{\log R}) = O(c^{-\frac{1}{\log R}})$, hence

$$\prod_{j=1}^{M} \chi(\frac{\log a_j}{\log R})\chi(\frac{\log b_j}{\log R}) = \int_I \cdots \int_I \prod_{j=1}^{M} \frac{\psi(x_j)\psi(y_j)}{a_j^{\frac{1+ix_j}{\log R}} b_j^{\frac{1+ix_j}{\log R}}} dx_j dy_j + O_A(\log^{-A} R \prod_{j=1}^{M}(a_j b_j)^{-1/\log R})$$

Substitute this into (2.2.7). The error term can be shown to be $o(1)$ for large enough A. Indeed, using that $\mathbb{E}_{\mathbf{x}\in\mathbb{Z}_p^t}\mathbf{1}_{a_j,b_j|\theta_j(\mathbf{x})\forall j}$ is 1 if $a_j,b_j=1\forall j$ and $\leq 1/p$ otherwise. The error term is given by

$$\begin{aligned}
&\sum_{\substack{a_i,b_i\in\mathbb{N}\\ \text{squarefree}}}\mathbb{E}_{\mathbf{x}\in\mathbb{Z}_D^t}\mathbf{1}_{a_j,b_j|\theta_j(\mathbf{x})\forall j}O_A(\log^{-A}R\prod_{j=1}^M(a_jb_j)^{-1/\log R})\\
&\lesssim_A(\log R)^{-A}\prod_p\sum_{a_i,b_i\in\{1,p\}}\left[\mathbb{E}_{x\in\mathbb{Z}_p^t}\mathbf{1}_{a_j,b_j|\theta_j(x)\ \forall j}\right]\prod_{j=1}^M(a_jb_j)^{-1/\log R}\\
&\leq(\log R)^{-A}\prod_p(1+p^{-1}\sum_{a_i,b_i\in\{1,p\}}(a_1b_1...a_Mb_M)^{-1/\log R})\\
&=(\log R)^{-A}\prod_p[1+p^{-1}((p^{-1/\log R}+1)^{2M}-1)]\\
&\leq(\log R)^{-A}\prod_p(1-p^{-1-1/\log R})^{2M}\\
\text{(apply (2.1.8))}&=(\log R)^{-A}\zeta(1+1/\log R)^{2M}=O(\log R)^{-2M-A}=o(1)
\end{aligned}$$

where $A>0$ can be chosen arbitrarily large. Denote

$$x_j'=\frac{1+ix_j}{\log R},y_j'=\frac{1+iy_j}{\log R}. \tag{2.2.9}$$

The main term in (2.2.7) becomes

$$\int_I\cdots\int_I\sum_{\mathbf{a},\mathbf{b}\in\mathbb{N}^M}\prod_p\left(\mathbb{E}_{\mathbf{x}\in\mathbb{Z}_p^t}\prod_{i=1}^M\mathbf{1}_{a_i,b_i|\theta_i(\mathbf{x})}(\mathbf{x})\right)\prod_{j=1}^M\frac{\mu(x_j)\mu(y_j)}{a_j^{x_j'}b_j^{y_j'}}\prod_{j=1}^M\psi(x_j)\psi(y_j)dx_jdy_j$$

where

$$\begin{aligned}
\sum_{\mathbf{a},\mathbf{b}\in\mathbb{N}^M}\prod_p\left(\mathbb{E}_{\mathbf{x}\in\mathbb{Z}_p^t}\prod_{i=1}^M\mathbf{1}_{a_i,b_i|\theta_i(\mathbf{x})}(\mathbf{x})\right)\prod_{j=1}^M\frac{\mu(x_j)\mu(y_j)}{a_j^{x_j'}b_j^{y_j'}}&=\prod_p\sum_{\mathbf{a},\mathbf{b}\in\{1,p\}^M}\left(\mathbb{E}_{\mathbf{x}\in\mathbb{Z}_p^t}\prod_{i=1}^M\mathbf{1}_{a_i,b_i|\theta_i(\mathbf{x})}(\mathbf{x})\right)\prod_{j=1}^M\frac{\mu(x_j)\mu(y_j)}{a_j^{x_j'}b_j^{y_j'}}\\
&:=\prod_p E_p
\end{aligned}$$

is the Euler product, where the Euler factor is

$$E_p=\sum_{\mathbf{a},\mathbf{b}\in\{1,p\}^M}\left(\mathbb{E}_{\mathbf{x}\in\mathbb{Z}_p^t}\prod_{i=1}^M\mathbf{1}_{a_i,b_i|\theta_i(\mathbf{x})}(\mathbf{x})\right)\prod_{j=1}^M\frac{\mu(x_j)\mu(y_j)}{a_j^{x_j'}b_j^{y_j'}}=\sum_{I,J\subseteq[M]}\frac{(-1)^{|I|+|J|}\omega_{I\cup J}(p)}{p^{\sum_{j\in I}x_j'+\sum_{j\in J}y_j'}}$$

Define a more convenient Euler's factor

$$E'_p = \prod_{j=1}^{m} \frac{(p^{1+x'_j}-1)(p^{1+y'_j}-1)}{p(p^{1+x'_j+y'_j})} = \prod_{j=1}^{M} \frac{(1-p^{-1-x'_j})(1-p^{-1-y'_j})}{1-p^{-1-x'_j-y'_j}}$$

From (2.1.8), we have

$$\prod_p E'_p = \frac{(1+o(1))}{\log^M R} \prod_{j=1}^{M} \frac{(1+ix'_j)(1+iy'_j)}{2+ix'_j+iy'_j}.$$

Define $F_p = E_p/E'_p$ we have

Lemma 2.2.11.

$$\prod_p E_p = \prod_p F_p \prod_p E'_p = \prod_p F_p \frac{1+o(1)}{\log^M R} \prod_{j=1}^{M} \frac{(1+ix_j)(1+iy_j)}{2+i(x_j+y_j)}$$

Next, we use that[12]

$$\int_{\mathbb{R}} \int_{\mathbb{R}} \frac{(1+ix_j)(1+iy_j)}{2+i(x_j+y_j)} \psi(x_j)\psi(y_j) dx_j dy_j = \int \chi'(t)^2 dt := 1.$$

The main term becomes

$$\begin{aligned}
&\int_I \cdots \int_I \prod_p F_p \prod_{j=1}^{M} \frac{(1+ix_j)(1+iy_j)}{2+i(x_j+y_j)} \psi(x_j)\psi(y_j) dx_j dy_j (1+o(1)) \log^{-M} R \\
&= (1+o(1)) \prod_p F_p \log^{-M} R \Big(\int_I \int_I \frac{(1+ix_j)(1+iy_j)}{2+i(x_j+y_j)} \psi(x_j)\psi(y_j) dx_j dy_j \Big)^M \\
&= (1+o(1)) \prod_p F_p \log^{-M} R \Big(\int_{\mathbb{R}} \int_{\mathbb{R}} \frac{(1+ix_j)(1+iy_j)}{2+i(x_j+y_j)} \psi(x_j)\psi(y_j) dx_j dy_j + o(1) \Big)^M \\
&= (1+o(1)) \log^{-M} R \prod_p F_p
\end{aligned}$$

To find asymptotic of $\prod_p F_p$, we apply the local factor estimate.

Lemma 2.2.12 (Euler factor estimate for Linear forms condition)**.** *We have*

$$E_p = (1+O(p^{-2}))E'_p$$

[12]This follows from the identity

$$\frac{1}{2+ix+iy} = \int_0^\infty e^{-(1+ix)t} e^{-(1+iy)t} dt.$$

and moreover[13],

$$\prod_{p\le\omega} F_p = \exp(O_m(1/\omega))(1+O(\frac{\log\omega}{\sqrt{\log R}}))(W/\phi(W))^M = (W/\phi(W))^M(1+O_M(1/\omega)1+O(\frac{\log\omega}{\sqrt{\log R}})).$$

Proof. Recall notation x_j', y_j' defined in (2.2.9).

$$\begin{aligned}
E_p &= \sum_{I,J\subseteq[M]} \frac{(-1)^{|I|+|J|}\omega_{I\cup J}(p)}{p^{\sum_{j\in I}x_j'+\sum_{j\in J}y_j'}} \\
&= \omega_\emptyset(p) - \sum_{j=1}^{M}\left(\frac{1}{p^{x_j'}}+\frac{1}{p^{y_j'}}-\frac{1}{p^{1+x_j'+y_j'}}\right) + \sum_{I,J\subseteq[M],|I\cup J|\ge 2}\frac{O_M(p^{-2})}{p^{\sum_{j\in I}x_j'+\sum_{j\in J}y_j'}} \\
&= 1-\sum_{j=1}^{M}\left(\frac{p^{x_j'}+p^{y_j'}-1}{p^{1+x_j'+y_j'}}\right)+O_M(p^{-2})
\end{aligned}$$

Now recall $|x_j'|, |y_j'| = O((\log R)^{-1/2})$ and $E_p' = \prod_{j=1}^M \frac{(p^{1+x_j'}-1)(p^{1+y_j'}-1)}{p^{1+x_j'+y_j'}-1}$. We compute

$$\begin{aligned}
F_p = E_p/E_p' &= \big(1-\sum_{j=1}^{M}\frac{p^{x_j'}+p^{y_j'}-1}{p^{1+x_j'+y_j'}}\big)\prod_{j=1}^{M}(1-p^{-1-x_j'-y_j'})p^{-1}(1-p^{-1-x_j'})^{-1}(1-p^{1-y_j'})^{-1}+O(p^{-2}) \\
&= 1+O_M(p^{-2}).
\end{aligned}$$

Now
$$\prod_{p>\omega}(1+O_M(p^{-2})) = \exp(O_M(\sum_{p>\omega}p^{-2})) = \exp(O_M(1/\omega)).$$

For $p<\omega$, $E_p = 1$, whereas if $|z_j| = O(\log^{-1/2}R)$ then $1-p^{-1-z_j} = 1-p^{-1}\exp(-z_j\log p) = 1-p^{-1}(1+O(|z_j|\log p)) = (1-p^{-1})(1+O(\frac{|z_j|\log p}{p}))$, applying this with $z_j = x_j, y_j, x_j+y_j$, we have

$$\begin{aligned}
\prod_{p\le\omega}E_p'^{-1} = \prod_{p\le\omega}\prod_{j=1}^{M}\frac{1-p^{-1-x_j'-y_j'}}{(1-p^{-1-x_j'})(1-p^{-1-y_j'})} &= \prod_{p\le\omega}\prod_{j=1}^{M}\frac{1}{(1-1/p)(1+O(|z_j|\frac{\log p}{p}))} \\
&= \prod_{p\le\omega}\left[(\frac{p}{p-1})^M(1+O(\frac{\log p}{p\log^{1/2}R}))\right]
\end{aligned}$$

The Lemma follows by recalling that $\prod_{p\le\omega}(p/p-1)^M = (W/\phi(W))^M$ and

$$\prod_{p\le\omega}(1+O(\frac{\log p}{p\log^{1/2}R})) = \exp(\frac{1}{\sqrt{\log R}}O(\sum_{p\le\omega}\frac{\log p}{p})) = \exp(O(\frac{\log\omega}{\sqrt{\log R}})).$$

[13] Here the term $(W/\phi(W))^M$ comes from primes $p\le\omega$.

$\square$

Lemma 2.2.13 (Euler factor estimate for the correlation conditions)**.**

$$E_p = (1 + O(p^{-2}))E'_p \quad \textit{when } \; p \nmid \Delta_k \textit{ for all } k.$$
$$E_p = (1 + O(p^{-1/2}))E'_p \quad \textit{when } \; p|\Delta_k \textit{ for some } k.$$
$$\prod_p F_p = \exp(O_M(1/\omega))(W/\phi(W))^M(1 + O(\frac{\log \omega}{\sqrt{\log R}})) \prod_{p|\Delta_1\ldots\Delta_k}(1 + O(p^{-1/2}))$$

Proof. The first statement is similar to the first part of Lemma 2.2.12. Assume $p|\Delta_k$, using $x'_j, y'_j = o(1)$, and applying the local factor estimate (Lemma 2.2.10), we have

$$E_p = 1 + O(1/p) \sum_{I,J\subseteq[M], I\cup J\neq\emptyset} \frac{(-1)^{|I|+|J|}}{p^{\sum_{i\in I} x'_j + \sum_{j\in J} y'_j}} = 1 + O(p^{-1/2})$$

$$\begin{aligned} E'_p = \prod_{j=1}^{M} \frac{(p^{1+x'_j} - 1)(p^{1+y'_j} - 1)}{p(p^{1+x'_j+y'_j} - 1)} &= \prod_{j=1}^{M} \frac{(1 - p^{-1-x'_j})(1 - p^{-1-y'_j})}{1 - p^{-1-x'_j-y'_j}} \\ &= \prod_{j=1}^{M} (1 - p^{-1-x'_j})(1 - p^{-1-y'_j})(1 + p^{-1-x'_j-y'_j} + O(p^{-3/2})) \\ &= 1 + O(p^{-1/2}) \end{aligned}$$

$\square$

Now define $\tau = \tau_M : \mathbb{Z}_N \to \mathbb{R}_{\geq 0}, \tau(0) := \exp(\frac{CM \log N}{\log\log N})$ so $\|\nu\|_\infty^M \leq \tau(0)$. Define $\tau(n) = O_M(n) \prod_{p|n}(1 + O(p^{-1/2}))^{O_M(1)}$. One estimates

$$\begin{aligned} \prod_{p|\Delta_k} (1 + O_M(p^{-1/2})) &= \prod_{1\leq i<j\leq m_k} (\prod_{p|W\cdot[(a^{(k)}_{j'} h^k_j - a^{(k)}_j h^k_{j'}) + (a^k_{j'} - a^{(k)}_j)b]} (1 + O_M(p^{-1/2})) \\ &\leq \prod_{1\leq i<j\leq m_k} \prod_{p|W\cdot[(a^{(k)}_{j'} h^k_j - a^{(k)}_j h^k_{j'}) + (a^k_{j'} - a^{(k)}_j)b]} (1 + p^{-1/2})^{O_M(1)} \\ &\leq O_M(1) \sum_{1\leq i<j\leq M} \prod_{p|W\cdot[(a^{(k)}_{j'} h^k_j - a^{(k)}_j h^k_{j'}) + (a^k_{j'} - a^{(k)}_j)b]} (1 + p^{-1/2})^{O_M(1)} \\ &\leq \sum_{1\leq i<j\leq M} \tau(W \cdot ((a^{(k)}_{j'} h^k_j - a^{(k)}_j h^k_{j'}) + (a^k_{j'} - a^{(k)}_j)b). \end{aligned}$$

Now we verify

$$\mathbb{E}_{x\in\mathbb{Z}_N} \tau(x)^q = O_q(1)$$

Since $1/\sqrt{p} \to 0, p \to \infty$ then $(1+O(p^{-1/2}))^{O_M(q)} \le 1 + p^{-1/4}$. for all but finitely many p.

$$\mathbb{E}_{0<|n|\le N}(\prod_{p|m}(1+O(p^{-1/2})))^{O_M(q)} \le O_{M,q}(1)\mathbb{E}_{0\le n\le N}\prod_{p|n}(1+p^{-1/4}) \le O_{M,q}(1)\mathbb{E}_{0\le n\le N}(\sum_{d|n} d^{-1/4})$$
$$= O_M, q(1)\sum_{d=1}^{N} d^{-5/4} = O_{M,q}(1).$$

Lemma 2.2.14. *ν satisfies the pseudorandomness conditions (2.2.4), (2.2.5).*

Proof. We follow the argument in [51]. First by clearing the denominator (in $\mathbb{Z}_N$, N is prime), we may assume the linear forms have integer coefficients with coefficient bounds from k to $k(k!) < (k+1)!$. Choose ω sufficiently large so that $(k+1)! < \sqrt{\frac{\omega}{2}}$. To prove (2.2.5), we just use the trivial bound $\nu(x) \le 1 + \frac{\phi(W)}{W}\Lambda_R(\theta_i(\mathbf{x}))$ and apply Lemma 2.2.9.

To deal with the two-part definition of ν and to get the asymptotic of the form $1+o(1)$, let $Q = Q(N)$ chosen to be a small power of N such that $N/Q \ge R^{10M}$, we subdivide $\mathbb{Z}_N^t$ into Q^t roughly equal sized boxes

$$B_{u_1,\dots,u_t} = \{\mathbf{x} \in \mathbb{Z}_N^t : x_j \in [\lfloor\frac{u_jN}{Q}\rfloor, \lfloor\frac{(u_{j+1})N}{Q}\rfloor]\}, \text{ where } u_1,\dots,u_t \in \mathbb{Z}_Q.$$

Then $|B_{u_1,\dots,u_t}| = (N/Q + O(1))^t = (N/Q)^t(1 + O(Q/N))$. Then

$$\mathbb{E}_{\mathbf{x}\in\mathbb{Z}_N^t}\nu(\phi_1(\mathbf{x}))\dots\nu(\phi_m(\mathbf{x})) = \frac{1}{Q^t(N/Q)^t}\sum_{(u_1,\dots,u_t)\in\mathbb{Z}_Q^t}\sum_{\mathbf{x}\in B_{u_1,\dots,u_t}}\nu(\phi_1(\mathbf{x}))\dots\nu(\phi_m(\mathbf{x}))$$
$$= (1+o(1))\mathbb{E}_{(u_1,\dots,u_t)\in\mathbb{Z}_Q^t}\mathbb{E}_{\mathbf{x}\in B_{u_1,\dots,u_t}}\nu(\phi_1(\mathbf{x}))\dots\nu(\phi_m(\mathbf{x}))$$

We say that a box $B_{u_1,\dots,u_t}$ is **nice** if $\phi_i(B_{u_1,\dots,u_t}) \subseteq [\varepsilon_k N, 2\varepsilon_k N]\ \forall i$. Since $N/Q > R^{10M}$, applying Lemma 2.2.12 (in particular, Lemma 2.2.9), we obtain

$$\mathbb{E}_{\mathbf{x}\in B_{u_1,\dots,u_t}\ \text{nice}}\nu(\psi_1(\mathbf{x}))\dots\nu(\psi_m(\mathbf{x})) = 1 + O_M(1/\omega)$$

(we could replace each ν with either 1 or $\frac{\phi(W)\log R}{W}\Lambda_R^2$).

Now we claim that the proportion of the number of boxes that are not nice is $O(1/Q)$. Suppose $B_{u_1,\dots,u_t}$ is not nice then there is a linear form ψ and $\mathbf{x},\mathbf{y} \in B_{u_1,\dots,u_t}$ such that $\psi(\mathbf{x}) \in [\varepsilon N, 2\varepsilon N], \psi(\mathbf{y}) \notin [\varepsilon N, 2\varepsilon N]$. Hence by continuity, either $a = 1$ or 2, one has

$$a\varepsilon N = \sum_{j=1}^{t} L_j\lfloor\frac{Nu_j}{Q}\rfloor + b + O(N/Q).$$

Hence
$$\sum_{j=1}^{t} L_j u_j = a\varepsilon Q + \frac{bQ}{N} + O(1) \pmod Q.$$

For any choice of $u_1, \dots, u_{t-1}$, the number of choices of u_t is $O(1)$. Hence the number of non-nice boxes is $O(Q^{t-1})$. If $B_{u_1,\dots,u_t}$ is not nice, then using the trivial bound $\nu(x) \le 1 + \frac{\phi(W)}{W}\Lambda_R(\theta_i(\mathbf{x}))^2$, we obtain
$$\mathbb{E}_{\mathbf{x}\in B_{u_1,\dots,u_t} \text{ not nice}} = \exp(1/o(\omega))(O(1)+o(1)).$$

Hence

$$\begin{aligned}\mathbb{E}_{\mathbf{x}\in\mathbb{Z}_N^t}\nu(\theta_1(\mathbf{x}))\dots\nu(\theta_M(\mathbf{x})) &= \mathbb{E}_{\mathbf{u}\in\mathbb{Z}_Q^t}\Big[\mathbb{E}_{\mathbf{x}\in B_{\mathbf{u}} \text{ nice}}\nu(\theta_1(\mathbf{x}))\dots\nu(\theta_M(\mathbf{x})) + \mathbb{E}_{\mathbf{x}\in B_{\mathbf{u}} \text{ not nice}}\nu(\theta_1(\mathbf{x}))\dots\nu(\theta_M(\mathbf{x}))\Big] \\ &= \frac{1}{Q^t}\Big[Q^t(1+O(1/\omega)+O(\log\omega/\sqrt{\log R})) + O(Q^{t-1})\exp(O_M(1/\omega))(O(1)+o(1 \\ &= 1 + O(1/\omega) + O(\log\omega/\sqrt{\log R}) + O(1/Q(N)).\end{aligned}$$

Here R is small power of N, $Q(N)$ is a power of N. $O(1/Q)$ may be neglected.

The correlation condition for ν can be verified in the same way from Lemma 2.2.13 (in particular Lemma 2.2.9) and the definition and properties of τ.

□

2.3 Birch-Davenport's Circle Method

The main objective of this section is to prove the Theorem 2.3.4 below whose proof is a simple adaptation of arguments of Birch [11]. We may skip some details that are the same as in [11].

2.3.1 Set Up

Denote $\|x\|$ the distance of x to the closest integer.

Let $\mathcal{F} = (F_1, \dots, F_r)$ be a system of r homogeneous forms of degree k on $\mathbb{Z}^d$. We are particularly interested in the case $k \ge 2$. Let $V_{\mathcal{F}} = \{\mathbf{x} : \mathcal{F}(\mathbf{x}) = 0\} \subseteq \mathbb{C}^d$, we define $V_{\mathcal{F}}^* \subseteq V_{\mathcal{F}}$, the *singular variety* of $\mathcal{F}$, to be the set of points such that $Jac_{\mathcal{F}}$, the Jacobian of $\mathcal{F}$, drops rank.

Let us introduce the notation

$$\mathcal{R}_N(\mathbf{v}) = \#\{\mathbf{x}\in[N]^d : \mathcal{F}(\mathbf{x}) = \mathbf{v}\}.$$

$$\mathcal{R}_N(M, \mathbf{s}; \mathbf{v}) := |\{\mathbf{x}\in[N]^d;\ \mathbf{x}\equiv\mathbf{s}\ (mod\ M),\ \mathcal{F}(\mathbf{x}) = \mathbf{v}\}|.$$

For a family of integral forms $\mathcal{F} = (F_1, \dots, F_r)$ in variables $\mathbf{x}^i \in \mathbb{Z}^d$, write $\mathbf{x}^i = (x_1^i, \dots, x_d^i)$ and

$$F_i(x_1^i,\dots,x_d^i)=\sum_{0\le j_1+\dots+j_d\le k} c^i_{j_1,\dots,j_d}(x_1^i)^{j_1}\dots(x_d^i)^{j_d} \tag{2.3.1}$$

We may write it as a symmetric form:

$$F_i(x_1^i,\dots,x_d^i)=\sum_{0\le i_1,\dots,i_l\le d} a^i_{i_1,\dots,i_k}x^i_{i_1}\dots x^i_{i_k}$$

with $k!a^i_{i_1,\dots,i_k}\in\mathbb{Z}$. Define a symmetric integral $d-$linear form on $\mathbf{x}^i,\dots,\mathbf{x}^k$; $\mathbf{x}^i=(x_1^i,\dots x_d^i)$:

$$\Phi^i(\mathbf{x}^1,\dots,\mathbf{x}^k)=k!\sum_{0\le i_1,\dots,i_l\le d} a^i_{i_1,\dots,i_k}x^1_{i_1}\dots x^k_{i_k} \tag{2.3.2}$$

Then

$$k!F_i(\mathbf{x})=\Phi^i(\mathbf{x},\dots,\mathbf{x}). \tag{2.3.3}$$

For $\underline{\alpha}\in[0,1]^r$, define the exponential sum

$$S_N(M,\mathbf{s},\underline{\alpha}):=\sum_{\mathbf{x}\in\mathbb{Z}^d} e^{2\pi i\underline{\alpha}\cdot\mathcal{F}((M\mathbf{x}+\mathbf{s}))}\phi_N(M\mathbf{x}+\mathbf{s}) \tag{2.3.4}$$

where ϕ_N is the characteristic function of $[0,N]^d$. Recall the notation $\Delta^+_{\mathbf{h}}f(\mathbf{x})=f(\mathbf{x}+\mathbf{h})-f(\mathbf{x})$ then

$$k!\Delta^+_{\mathbf{h}}F_i(\mathbf{x})=k\Phi^i(\mathbf{x},\dots,\mathbf{x},\mathbf{h})+R_{k-2}(\mathbf{x},\mathbf{h}) \tag{2.3.5}$$

where $deg(R_{k-2})$ (in $\mathbf{x}$) $\le k-2$. Hence

$$k!\Delta^+_{\mathbf{h}^{k-1}}\dots\Delta^+_{\mathbf{h}^1}F_i(\mathbf{x})=k!\Phi^i(\mathbf{x},\mathbf{h}^{k-1}\dots,\mathbf{h}^1)+R_0(\mathbf{h}) \tag{2.3.6}$$

Here

$$\Phi^i(\mathbf{x},\mathbf{h}^1,\dots,\mathbf{h}^{k-1})=\sum_{1\le j,i_1,\dots,i_{k-1}\le d} a^i_{j,i_1,\dots,i_{d-1}}x_j h^1_{i_1}\dots h^{k-1}_{i_{k-1}}:=\sum_{1\le j\le d}x_j\Psi^i_j(\mathbf{h}^1,\dots,\mathbf{h}^{k-1}). \tag{2.3.7}$$

Definition 2.3.1 (Rank of $\mathcal{F}$; defined by Birch [11]). [14] *Let $\mathcal{F}=(F_1,\dots,F_r)$ be a system of r homogeneous form of degree k. The* ***Rank of*** $\mathcal{F}$ *is the codimension of the singular variety $V^*_{\mathcal{F}}$, the set of points $\mathbf{z}\in\mathbb{C}^d$ where the Jacobian $\partial\mathcal{F}/\partial\mathbf{z}$ drops rank.*

We will write

$$K:=\frac{codim(V^*_{\mathcal{F}})}{2^{k-1}}$$

[14] There is also a notion of Schmidt rank [94]. Let F be a single homogeneous polynomial, the Schmidt rank is defined to be the smallest integers h such that we can find homogeneous forms $T_1,\dots,T_h,R_1,\dots,R_h$ of positive degree such that $Q=T_1R_1+\dots+T_hR_h$. Note that if $Q=\sum_{i=1}^h T_iR_i$ then $V_{\nabla Q}\le 2h$.

Example 2.3.2. *If $F(\mathbf{x}) = A\mathbf{x} \cdot \mathbf{x}$ is a quadratic form i.e. A is an integral symmetric matrix, then $\nabla F(\mathbf{x}) = A\mathbf{x}, V_F^* = Ker(A), codim V_F^* = rank(A)$.*

2.3.2 The Circle Method

Theorem 2.3.3 (Birch's Theorem)**.** *Let $\mathcal{F}$ be a system of r homogeneous integral forms of degree k in d variables. Suppose $K > (k-1)r(r+1), N > 1$, then*

$$\mathcal{R}_N(\mathbf{v}) = N^{d-kr}\sigma(\mathbf{v})J(N^{-k}\mathbf{v}) + O(N^{d-kr-\varepsilon})$$

for some $\varepsilon > 0$, where

$$\sigma(\mathbf{v}) = \prod_{p \text{ prime}} \sigma_p(\mathbf{v})$$
$$\sigma_p(\mathbf{v}) = \lim_{r\to\infty} p^{-r(d-1)}\#\{\mathbf{x} \in (\mathbb{Z}/p)^d : \mathcal{F}(\mathbf{x}) = \mathbf{v} \pmod{p^r}\}$$

$J(\mathbf{u}) = J_{\mathcal{F}}(\mathbf{u})$ *is the singular integral defined in* (2.3.31).

Recall from [11] that the singular series has a positive lower bound independent of $\mathbf{u}$ if we can find nonsingular solutions $\pmod p$ for every p. The singular integral $J(\mathbf{u}) \geq c(\delta) > 0$ independently of N, provided that the equation $\mathcal{F}(\mathbf{x}) = \mathbf{u}$ has a nonsingular real point in the cube $[\delta, 1-\delta]^d$, see [94] Section 9 and [11] Section 6.

Theorem 2.3.4. *Let $\mathcal{F} = (F_1, \ldots, F_r)$ be a family of integral forms of degree $k \geq 2$ satisfying the rank condition*

$$Rank(\mathcal{F}) > r(r+1)(k-1)2^{k-1} \tag{2.3.8}$$

and for given $M \in \mathbb{N}$ and $\mathbf{s} \in \mathbb{Z}^d$, recall that

$$\mathcal{R}_N(M, \mathbf{s}; \mathbf{v}) := |\{\mathbf{x} \in [N]^d;\ \mathbf{x} \equiv \mathbf{s}\ (mod\ M),\ \mathcal{F}(\mathbf{x}) = \mathbf{v}\}|. \tag{2.3.9}$$

Then there exists a constant $\delta' = \delta'(k,r) > 0$ such that the following holds.

(i) If $0 < \eta \leq \frac{1}{4r^2(r+1)(r+2)k^2}$ then for every $1 \leq M \leq N^{\frac{\eta}{1+\eta}}$ and $\mathbf{s} \in \mathbb{Z}^d$ one has the asymptotic

$$\mathcal{R}_N(M, \mathbf{s}; \mathbf{v}) = N^{d-rk}M^{-d}\, J(N^{-k}\mathbf{v}) \prod_p \sigma_p(M, \mathbf{s}, \mathbf{v}) + O(N^{d-rk-\delta'}M^{-d}). \tag{2.3.10}$$

(ii) Moreover if

$$Rank\,(\mathcal{F}) > \big(r(r+1)(k-1) + rk\big)2^k \tag{2.3.11}$$

then the asymptotic formula (2.3.10) holds for $\eta \leq \frac{1}{4r(r+2)k}$.

In the remaining if this subsecton, we describe the proof of Theorem 2.3.4.

Proof of Theorem 2.3.4

Recall that $\|x\|$ is the distance of x to the closet integer. The first lemma is an exponential sum estimate analogous to Lemma 2.1 in [11].

Lemma 2.3.5. *Let* $1 \leq M < N$ *and* $\mathbf{s} \in \mathbb{Z}^d$. *Then*

$$|(N/M)^{-d} S_N(M, \mathbf{s}, \underline{\alpha})|^{2^{k-1}} \lesssim (N/M)^{-kd} \sum_{\mathbf{h}^1,\ldots,\mathbf{h}^{k-1} \in [-N/M, N/M]^d} \prod_{j=1}^{d} \min\left(N/M, \|M^k x_j \underline{\alpha} \cdot \Psi_j(\mathbf{h}^1, \ldots, \mathbf{h}^{k-1})\|\right)$$

where the i^{th} *component of the multi-linear form* $\Psi_j = (\Psi_j^i)_{i=1}^r$ *is given by*

$$\Psi_j^i(\mathbf{x}, \mathbf{h}^1, \ldots, \mathbf{h}^{k-1}) = k! \sum_{1 \leq j_1, \ldots, j_{k-1} \leq d} a^i_{j, j_1, \ldots, j_{k-1}} \mathbf{h}^1_{j_1} \ldots \mathbf{h}^{k-1}_{j_{k-1}}.$$

Proof. We will invoke the following simple inequality: Let I be an interval of length at most N/M and $\beta \in \mathbb{R}$ then

$$|\sum_{x \in I} e^{2\pi i \beta x}| \leq \min\{N/M, \|\beta\|^{-1}\} \tag{2.3.12}$$

Write

$$F(M\mathbf{x} + \mathbf{s}) = M^k F(\mathbf{x}) + G_{M,\mathbf{s}}(\mathbf{x}),\ deg(G_{d,\mathbf{s}}) < k \tag{2.3.13}$$

and note that

$$|N^{-d} \sum_{x \in [N]^d} f(x)|^2 = N^{-d} |N^{-d} \sum_{x,h} f(x) \overline{f(x+h)}|$$

Applying this $k-1$ times, we have

$$|(N/M)^{-d} S_N(M, \mathbf{s}, \underline{\alpha})|^{2^{k-1}} = \left|(N/M)^{-d} \sum_{\mathbf{x} \in \mathbb{Z}^d} e^{2\pi i \underline{\alpha} \cdot \mathcal{F}((M\mathbf{x}+\mathbf{s}))} \phi_N(M\mathbf{x} + \mathbf{s})\right|^{2^{k-1}}$$

$$\leq (N/M)^{-(k-1)d} \sum_{\mathbf{h}^1, \ldots \mathbf{h}^{k-1}} \left|(N/M)^{-d} \sum_{\mathbf{x}} e^{2\pi i \underline{\alpha} \cdot \Delta^+_{\mathbf{h}^{k-1}} \ldots \Delta^+_{\mathbf{h}^1} \mathcal{F}(M\mathbf{x}+\mathbf{s})} \Delta_{\mathbf{h}^{k-1}} \ldots \Delta_{\mathbf{h}^1} \phi(M\mathbf{x} + \mathbf{s})\right|$$

By (2.3.13) and (2.3.6), we calculate

$$\Delta^+_{\mathbf{h}^{k-1}} \ldots \Delta^+_{\mathbf{h}^1} \mathcal{F}(M\mathbf{x} + \mathbf{s}) = M^k \Delta^+_{\mathbf{h}^{k-1}} \ldots \Delta^+_{\mathbf{h}^1} \mathcal{F}(\mathbf{x}) = M^k \sum_{j=1}^{d} x_j \Psi_j(\mathbf{h}^1, \ldots, \mathbf{h}^{k-1})$$

and one may verify that $\Delta_{\mathbf{h}^{k-1}} \ldots \Delta_{\mathbf{h}^1}\phi_N(Mx+s) = 0$ unless $\mathbf{h}^1, \ldots, \mathbf{h}^{k-1} \in [-N/M, N/M]^d$. Hence

$$|(N/M)^d S_N(M,\mathbf{s},\underline{\alpha})|^{2^{k-1}} \leq (N/M)^{-kd} \sum_{\mathbf{h}^1,\ldots\mathbf{h}^{k-1}\in[-N/M,N/M]^d} \left| \prod_{j=1}^{d} \left(\sum_{x_j} e^{2\pi i M^k \sum_{i=1}^{r} \alpha_i x_j \Psi_j^i(\mathbf{h}^1,\ldots,\mathbf{h}^{k-1})} \right) \right|$$

The result then follows from (2.3.12). $\square$

In the next step we will use the above lemma to divide S^1 into ***major arc*** and ***minor arc***. The argument follows directly as in [11]. We sketch the argument below.

Given $\eta, \gamma > 0$, define the following sets

$$R((N/M)^\eta, (N/M)^{-\gamma}; \underline{\alpha})$$
$$:= \left|\{(\mathbf{h}^1,\ldots,\mathbf{h}^{k-1}) : \mathbf{h}^i \in [-(N/M)^\eta, (N/M)^\eta]^d; \|M^k\underline{\alpha}\cdot\Psi_j(\mathbf{h}^1,\ldots,\mathbf{h}^{k-1})\| \leq (N/M)^{-\gamma}\ \forall 1 \leq j \leq d\}\right|.$$

Now for fixed $\mathbf{h}^2, \ldots, \mathbf{h}^{k-1}$, consider the following map from $\mathbb{R}^d$ to $\mathbb{R}^d$

$$\mathbf{h} \to (M^k\underline{\alpha}\cdot\Psi_1(\mathbf{h},\mathbf{h}^2,\ldots,\mathbf{h}^{k-1}),\ \ldots\ , M^k\underline{\alpha}\cdot\Psi_d(\mathbf{h},\mathbf{h}^2,\ldots,\mathbf{h}^{k-1})).$$

Define the following symmetric convex body

$$B_{Q,K} = B_{Q,K,\mathbf{h}^2,\ldots,\mathbf{h}^{k-1}} = \{(\mathbf{x},\mathbf{y}) : \mathbf{x}\in[-Q,Q]^d, |y_j - M^k\underline{\alpha}\cdot\Psi_j(\mathbf{x},\mathbf{h}^2,\ldots,\mathbf{h}^{k-1})| \leq K^{-1}\ \forall 1\leq j\leq d\}$$

We have $R((N/M)^\eta, (N/M)^{-\gamma}; \underline{\alpha}) = \left|\{(\mathbf{x},\mathbf{y})\in\mathbb{Z}^{2d} : (\mathbf{x},\mathbf{y})\in B_{(N/M)^\eta,(N/M)^{-\gamma}}\}\right|$. Now we state the following fact which says that this set is essentially $d-$dimensional object.

Lemma 2.3.6 (Davenport [21] Lemma 3.3, [22] Chapter 12)**.** *Let $L > 1$ then*

$$|\mathbb{Z}^{2d}\cap B_{L^{-1}N/M, L^{-1}(N/M)^{-1}}| \gtrsim L^{-d}|\mathbb{Z}^{2d}\cap B_{N/M,(N/M)^{-1}}|$$

Applying this lemma repeatedly in $\mathbf{h}^{k-1},\ldots,\mathbf{h}^1$ respectively (with other $\mathbf{h}^i$ fixed at a time) with $L = (N/M)^{1-\theta}, 0<\theta<1$, one obtains

$$R((N/M)^\theta, (N/M)^{-k+(k-1)\theta}, \underline{\alpha}) \gtrsim (N/M)^{-(k-1)d(1-\theta)} R(N/M, (N/M)^{-1}, \underline{\alpha}) \tag{2.3.14}$$

Now subdividing $[-\frac{1}{2},\frac{1}{2}]^d$ into small cubes $\prod_{j=1}^d [\frac{i_j}{N/M}, \frac{i_{j+1}}{N/M}]$ of size $1/(N/M)$. Observe that if two points

$$\big(M^k\underline{\alpha}\cdot\Psi_1(\mathbf{h}^1,\ldots,\mathbf{h}^{k-1}),\ldots,M^k\underline{\alpha}\cdot\Psi_d(\mathbf{h}^1,\ldots,\mathbf{h}^{k-1})\big),$$
$$\big(M^k\underline{\alpha}\cdot\Psi_1(\mathbf{h},\mathbf{h}^2,\ldots,\mathbf{h}^{k-1}),\ldots,M^k\underline{\alpha}\cdot\Psi_d(\mathbf{g},\mathbf{h}^2,\ldots,\mathbf{h}^{k-1})\big)$$

are in the same cube, then one has

$$\|M^k\underline{\alpha}\cdot\Psi_j(\mathbf{h}-\mathbf{g},\mathbf{h}^2,\dots,\mathbf{h}^{k-1})\| \le \frac{1}{N/M}, 1\le j\le d$$

Hence the number of $\mathbf{h}^1,\dots,\mathbf{h}^{k-1}$ such that $(M^k\underline{\alpha}\cdot\Psi_1(\mathbf{h}^1,\dots,\mathbf{h}^{k-1}),\dots,M^k\underline{\alpha}\cdot\Psi_d(\mathbf{h}^1,\dots,\mathbf{h}^{k-1}))$ are in a given cube is bounded above by $R(N/M,(N/M)^{-1};\underline{\alpha})$ for every cube. Hence

$$\sum_{\mathbf{h}^1,\dots\mathbf{h}^k}\left|\prod_{j=1}^d\left(e^{2\pi i M^k x_j\underline{\alpha}\cdot\Phi_j(\mathbf{h}^1,\dots,\mathbf{h}^{k-1})}\right)\right| \lesssim \sum_{\mathbf{h}^1,\dots\mathbf{h}^k}\left|\prod_{j=1}^d \min\{N/M,\frac{1}{\|\underline{\alpha}\cdot\Psi_j(\mathbf{h}^1,\dots,\mathbf{h}^{k-1})\|}\}\right| \tag{2.3.15}$$

$$\lesssim R(N/M,(N/M)^{-1};\underline{\alpha})(\sum_{1\le i\le N/2M}(1/i))^d$$

$$\lesssim (\log(N/M))^d R((N/M),(N/M)^{-1};\underline{\alpha}) \tag{2.3.16}$$

Combining Lemma 2.3.5, (2.3.14) and (2.3.15), one obtains that for $0<\theta<1$,

$$\left((N/M)^{-d}|S_N(M,\mathbf{s},\underline{\alpha})|\right)^{2^{k-1}} \lesssim (N/M)^{-(k-1)d\theta}\log^d(N/M)\times \tag{2.3.17}$$
$$\times\left|\{\mathbf{h}^1,\dots,\mathbf{h}^{k-1}\in[-(N/M)^\theta,(N/M)^\theta]^d : \|M^k\underline{\alpha}\cdot\Psi_j(\mathbf{h}^1,\dots,\mathbf{h}^{k-1})\|\le (N/M)^{-k+(k-1)\theta}\,\forall 1\le j\le d\}\right|$$

Now consider the inequality (2.3.17), suppose $\exists \mathbf{h}^1,\dots,\mathbf{h}^{k-1}$ such that the matrix $(\Psi^i_j)_{1\le i\le r,1\le j\le d}$ has rank r, i.e. there is a non vanishing $r\times r$ minor, which we may assume to be $(\Psi^i_j)_{\substack{1\le i\le r\\1\le j\le r}}$. Let q denote the absolute value of the determinant of $(\Psi^i_j)_{\substack{1\le i\le r\\1\le j\le r}}$, since the degree of Ψ^i_j is $k-1$, we have $1\le q\le (N/M)^{r(k-1)\theta}$. Then

$$\|qM^k\underline{\alpha}\cdot\Psi_j(\mathbf{h}^1,\dots,\mathbf{h}^{k-1})\|\le q(N/M)^{-k+(k-1)\theta}\le (N/M)^{-k+(r+1)(k-1)\theta}$$

Then as in [11], we can find integers $a_1,\dots,a_r$ such that for $i=1,\dots,r$

$$|qM^k\alpha_i-a_i|\le (N/M)^{-k+(k-1)r\theta}$$

Now we divide the torus $\mathbb{T}^r=(\mathbb{R}/\mathbb{Z})^r$ into "*Major arcs*" and "*Minor arcs*". The major arcs is defined as

$$\mathcal{M}(\theta)=\bigcup_{1\le q\le (N/M)^{(k-1)r\theta}}\ \bigcup_{(\mathbf{a},q)=1}\mathcal{M}_{\mathbf{a},q}(\theta).$$

Here $(\mathbf{a},q)=gcd(a_1,\dots,a_r,q)$ and

$$\mathcal{M}_{\mathbf{a},q}(\theta):=\{\underline{\alpha}\in[0,1]^r;|M^k\alpha_i-a_i/q|\le q^{-1}(N/M)^{-k+(k-1)r\theta},\forall 1\le i\le r\}.$$

The minor arcs $\mathfrak{m}(\theta)$ is defined to be $\mathbb{T}^r \backslash \mathcal{M}(\theta)$. The name comes from the fact that even though minor arcs contributes most of the arcs on the "circle", the contribution to the integral is small from minor arc which we will verify now.

It is easy to see that $\Psi^i_j(\mathbf{z}, \dots, \mathbf{z}) = (k-1)!\partial_j F_i(\mathbf{z})$. Define $\Delta := \{(\mathbf{z}, \cdots, \mathbf{z}) : \mathbf{z} \in \mathbb{C}^d\} \subseteq \mathbb{C}^{d(k-1)}$ which is isomorphic to $\mathbb{C}^d$. Assume the first $r-1$ columns of $(\Psi^i_j)_{\substack{1\le i\le r\\ 1\le j\le d}}$ are linearly independent, let $W_\Phi \subseteq \mathbb{C}^{(k-1)d}$ be the locus of points satisfying the equations saying that the remaining $d-r+1$ containing these r column is zero. Hence $\Delta \cap W_\Phi = V^*_{\mathcal{F}}$ and hence $codim(\Delta) + codim W_\Phi \geq (k-1)d - dim(V^*_{\mathcal{F}})$ where $codim(\Delta) = (k-2)d$, so

$$codim(W_\Phi) \geq codim(V^*_{\mathcal{F}}).$$

Now if $\underline{\alpha} \notin \mathcal{M}(\theta)$ then we estimate the size of the set on the RHS of (2.3.17) by

$$|\mathbb{Z}^{d(k-1)} \cap [-(N/M)^\theta, (N/M)^\theta]^{d(k-1)} \cap W_\Phi|,$$

which is $|(N/M)^{-\theta}\mathbb{Z})^{d(k-1)} \cap [-1,1]^{d(k-1)} \cap W_\Phi|$ by homogeneousity. We estimate this by the number of radius $\rho = c(N/M)^{-\theta}$ needed to cover $[-1,1]^{d(k-1)} \cap W_\Phi$. We use the following lemma.

Lemma 2.3.7 ([34], Chapter 7). *Let $W \subseteq \mathbb{C}^m$ be a homogeneous algebraic set of topological dimension l and $0 < \rho < 1$. Then $W \cap [-1,1]^m$ can be covered by $c\rho^{-l}$ balls of radius ρ.*

We obtain the minor arc estimate

Lemma 2.3.8 ([11], Lemma 3.3). *If $\{M^k\underline{\alpha}\} \notin \mathcal{M}(\theta)$ then for every $\tau > 0$,*

$$|S_N(M, \mathbf{s}, \underline{\alpha})| \ll_\tau (N/M)^{d-K\theta+\tau} \tag{2.3.18}$$

Proof. This is similar to [11]. Suppose $M^k\underline{\alpha} \notin \mathcal{M}(\theta)$ then by (2.3.17), we obtain

$$\begin{aligned}|(N/M)^{-d}S_N(M,\mathbf{s},\underline{\alpha})|^{2^{k-1}} &\lesssim (N/M)^{-d(k-1)\theta}\left|[(-N/M)^\theta, (N/M)^\theta]^{d(k-1)} \cap W_\Phi \cap \mathbb{Z}^{d(k-1)}\right| \log^d(N/M)\\ &\lesssim (N/M)^{-d(k-1)\theta}(N/M)^{\theta dim(W_\Phi)+\tau}\\ &= (N/M)^{-\theta codim(V^*_{\mathcal{F}})+\tau}\end{aligned}$$

as required. □

Lemma 2.3.9. *Let $0 < \theta, \epsilon < 1$ and let $0 < \eta \leq \epsilon r(1-k^{-1})\theta$. Suppose $M \leq N^{\frac{\eta}{1+\eta}}$ then for $\alpha \notin \mathcal{M}(\theta)$ one has uniformly in $\mathbf{s} \in \mathbb{Z}^d$*

$$|S_N(M,\mathbf{s},\alpha)| \lesssim_\tau (N/M)^{d-\frac{K}{1+\epsilon}\theta+\tau}, \forall \tau > 0 \tag{2.3.19}$$

Proof. If $M^k \underline{\alpha} \in \mathcal{M}_{\mathbf{a},q}(\theta) \pmod 1$ then there is $q \leq (N/M)^{r(k-1)\theta}$ and $a_i \in \mathbb{Z}$ such that $(a_i, q) = 1$ and $|M^k\alpha_i - a_i/q| \leq q^{-1}(N/M)^{-k+(k-1)r\theta}$. Hence

$$|\alpha_i - a_i'/q_1| \leq q_1^{-1}(N/M)^{-k+(k-1)r\theta}$$

for some $q_1 \leq M^k(N/M)^{(k-1)r\theta}$ and $(a_i', q) = 1$.

Now since $M \leq N^{\frac{\eta}{1+\eta}}$,we have $M \leq (N/M)^\eta$ hence $q_1 \leq (N/M)^{k\eta + r(k-1)\theta} \leq (N/M)^{(1+\epsilon)r(k-1)\theta}$. This implies $\underline{\alpha} \notin \mathcal{M}((1+\epsilon)\theta)$. By contrapositive and Lemma 2.3.8, one has (2.3.19). $\square$

The first application of Lemma 2.3.9 is the following estimate for Gauss sums.

Lemma 2.3.10 (Gauss sum estimate)**.** *Let $q \in \mathbb{N}$ and $\mathbf{a} \in \mathbb{Z}^r$ and $(\mathbf{a}, q) = 1$ and $\mathbf{s} \in \mathbb{Z}^d$. Define the **Gauss Sum***

$$S_{\mathbf{a},q}(M, \mathbf{s}) := \sum_{\mathbf{x} \in \mathbb{Z}_q^d} e^{2\pi i \frac{\mathbf{a} \cdot \mathcal{F}(M\mathbf{x}+\mathbf{s})}{q}} \tag{2.3.20}$$

Then if $1 \leq M < q^{\frac{\epsilon}{k}}$, one has the following estimate

$$|S_{\mathbf{a},q}(M, \mathbf{s})| \lesssim_\tau q^{d - \frac{K}{(1+\epsilon)r(k-1)} + \tau} \quad (\forall \tau > 0) \tag{2.3.21}$$

In particular, taking $M = 1, \epsilon \to 0$ in (2.3.21), one has

$$S_{\mathbf{a},q}(1, \mathbf{s}) = s_{\mathbf{a},q}(1, 0) \lesssim_\tau q^{d - \frac{K}{r(k-1)} + \tau}$$

Proof. Note that

$$S_{\mathbf{a},q}(M, \mathbf{s}) = S_{Mq}(M, \mathbf{s}, \mathbf{a}/q)$$

Also, if $r(k-1)\theta < 1$ then for all $1 \leq q' \leq q^{r(k-1)\theta} < q$ and $(\mathbf{a}', q') = 1$,

$$|\frac{\mathbf{a}}{q} - \frac{\mathbf{a}'}{q'}| \geq \frac{1}{qq'} > \frac{1}{q'} q^{-k+r(k-1)\theta}$$

This implies $\frac{\mathbf{a}}{q} \notin \mathcal{M}(\theta)$. Since $M < q^{\epsilon/k}$, choose θ so that $r(k-1)\theta < 1$, since $M \leq (N/M)^{r(k-1)\theta}$, one has

$$M < q^{\frac{\epsilon}{k}} < (N/M)^{\epsilon r(1-k^{-1})\theta}$$

Hence $M < (\frac{N}{M})^\eta$ where $\eta := \epsilon r(1-k^{-1})\theta$ which is the assumption of Lemma 2.3.9. Applying this lemma, we have for each $\tau > 0$,

$$|S_{\mathbf{a},q}(M, \mathbf{s})| \lesssim_\tau q^{d - \frac{K}{1+\epsilon}\theta + \tau} \lesssim_\tau q^{d - \frac{K}{(1+\epsilon)r(k-1)} + \tau}$$

$\square$

Now we will show that the minor arcs contribute little to the integral; this holds if $codim(V_{\mathcal{F}}^*)$ is large enough which we will assume. We will apply Lemma 4.4 in [11] (see Lemma 2.3.11 below) to our situation, we make the following assumption

$$K := \frac{codim(V_{\mathcal{F}}^*)}{2^{k-1}} > (1+\epsilon)r(r+1)(k-1) \tag{2.3.22}$$

Then we can choose small positive numbers δ, θ_0 satisfying the conditions

$$\delta + 2r(r+2)\theta_0 < 1 \tag{2.3.23}$$

$$2\delta\theta_0^{-1} < K(1+\epsilon)^{-1} - r(r+1)(k-1) \tag{2.3.24}$$

If θ is small enough so that $k > 2r(k-1)\theta$ then, as in [11], we can verify the following facts from the definition of the major arcs.

- If $a/q \neq a'/q'$ then $\mathcal{M}_{\mathbf{a},q}, \mathcal{M}_{\mathbf{a}',q'}$ are disjoint (from condition (2.3.23)).
- $|\mathcal{M}(\theta)| \leq (N/M)^{-rk+r(r+1)(k-1)\theta}$

We choose $\theta = \theta_T < \theta_{T-1} < \cdots < \theta_0$ and write

$$\mathcal{M}(\theta)^C = \mathcal{M}(\theta_0)^C \cup \bigcup_{i=0}^{T-1} \mathcal{M}(\theta_i)\backslash\mathcal{M}(\theta_{i+1})$$

Using the two facts above and the size of $S_N(M, \mathbf{s}, \underline{\alpha})$ to bound the integral as in [11]. We use the bound on $S_N(M, \mathbf{s}, \underline{\alpha})$ and (2.3.24) to show that integral over $\mathcal{M}(\theta_0)$ is $O((N/M)^{d-kr-\delta})$. We can also show that the integral over $\mathcal{M}(\theta_i)\backslash\mathcal{M}(\theta_{i+1})$ is $O((N/M)^{d-rk-\frac{3}{2}\delta})$, say, using the bound of $|\mathcal{M}(\theta_i)|$ and size of $S_N(M, \mathbf{s}, \underline{\alpha})$ on $\mathcal{M}(\theta_{i+1})$. Choose $|\theta_{i+1} - \theta_i|$ not too big and $T \ll_\delta 1$, see [11] Lemma 4.3. We obtain

Lemma 2.3.11 ([11], Lemma 4.4). *Let δ, θ_0 satisfy (2.3.23)-(2.3.24) and $0 < \eta \leq \epsilon(1-k^{-1})\theta_0$. Then for $1 \leq M \leq N^{\frac{\eta}{1+\eta}}$ and $\mathbf{s} \in \mathbb{Z}^d$, one has*

$$\int_{\alpha \notin \mathcal{M}(\theta_0)} |S_N(M, \mathbf{s}, \alpha)| d\alpha \lesssim_\delta (N/M)^{d-Mr-\delta}$$

Additionally, if

$$\eta < \delta/Mr \tag{2.3.25}$$

which is equivalent to $-\delta + (Mr+\delta)(1-\frac{1}{1+\eta}) < 0$. Then as in the above lemma,

$$(N/M)^{d-Mr-\delta} = N^{d-Mr}M^{-d}N^{-\delta}M^{Mr+\delta} \leq N^{d-Mr}d^{-n}N^{-\delta+(Mr+\delta)\eta(1+\eta)^{-1}} \leq N^{d-rM-\delta'}M^{-d}$$

for some $\delta' > 0$. Hence under assumptions of Lemma 2.3.11 and equation (2.3.25), one has

$$\mathcal{R}_N(M,\mathbf{s},\mathbf{v}) = \int_{S^1} e^{-2\pi i\alpha\cdot\mathbf{v}} S_N(M,\mathbf{s};\alpha)d\alpha = \int_{\mathcal{M}'(\theta_0)} e^{-2\pi i\alpha\cdot\mathbf{v}} S_N(M,\mathbf{s};\alpha)d\alpha + O(N^{d-rM-\delta'}M^{-d}) \tag{2.3.26}$$

for any set $\mathcal{M}'(\theta_0) \supseteq \mathcal{M}(\theta_0)$. Define $\mathcal{M}'(\theta_0)$ as follow:

$$\mathcal{M}'(\theta_0) := \bigcup_{1\le q\le (N/M)^{r(k-1)\theta_0}} \bigcup_{(\mathbf{a},q)=1} \mathcal{M}'_{\mathbf{a},q}(\theta_0) \tag{2.3.27}$$

$$\mathcal{M}'_{\mathbf{a},q}(\theta_0) := \{\underline{\alpha} \in [0,1]^r; |\alpha_i - a_i/q| \le (N/M)^{-k+r(k-1)\theta_0}, 1 \le i \le r\} \tag{2.3.28}$$

Now for given $\underline{\alpha} \in \mathcal{M}'_{\mathbf{a},q}(\theta_0)$, we can write

$$\underline{\alpha} = \mathbf{a}/q + \underline{\beta},\ \ |\underline{\beta}|_\infty \le |(N/M)|^{-k+r(k-1)\theta_0}$$

Then if $\underline{\alpha} \in \mathcal{M}'_{\mathbf{a},q}(\theta_0)$ we can give an estimation of $S_N(M,\mathbf{s};\alpha)$ in terms of the *singular series* and the *singular integral*.

Lemma 2.3.12. *Let $0 < \eta \le \frac{1}{2}$, $M \le N^{\frac{\eta}{1+\eta}}$, $\mathbf{s} \in \mathbb{Z}^d$. Then for $\underline{\alpha} \in \mathcal{M}'_{\mathbf{a},q}(\theta_0)$, we have*

$$S_N(M,\mathbf{s};\underline{\alpha}) = N^d M^{-d} q^{-d} S_{\mathbf{a},q}(M,\mathbf{s}) I(N^k\underline{\beta}) + O(N^{d-1+2\eta+r(k-1)\theta_0}M^{-d}) \tag{2.3.29}$$

where

$$I(\underline{\gamma}) := \int_{\mathbb{R}^d} e^{2\pi i\underline{\gamma}\cdot\mathcal{F}(\mathbf{y})}\mathbf{1}_{[0,1]^d}(\mathbf{y})d\mathbf{y}$$

Proof. Write $\mathbf{x} := q\mathbf{y} + \mathbf{z}$ with $\mathbf{z} \in [0,q)^d$. We have

$$S_N(M,\mathbf{s},;\underline{\alpha}) = \sum_{\mathbf{z}\in\mathbb{Z}_q^d} e^{2\pi i\frac{\mathbf{a}\cdot\mathcal{F}(M\mathbf{z}+\mathbf{s})}{q}} \sum_{\mathbf{y}\in\mathbb{Z}^d} e^{2\pi i\underline{\beta}\cdot\mathcal{F}(qM\mathbf{y}+M\mathbf{z}+\mathbf{s})}\mathbf{1}_{[0,N]^d}(qM\mathbf{y}+M\mathbf{z}+\mathbf{s}) \tag{2.3.30}$$

Now for $\mathbf{t} \in [0,1]^d$, using that the arguments in the functions are bounded by $\lesssim N$ and recall that $q \le (N/M)^{r(k-1)\theta_0}$, $M \le (N/M)^\eta$, we have

$$\begin{aligned}
&|e(\underline{\beta}\cdot\mathcal{F}(\mathbf{s}+M\mathbf{z}+qM(\mathbf{y}+\mathbf{t}))) - e(\underline{\beta}\cdot\mathcal{F}(\mathbf{s}+M\mathbf{z}+qM\mathbf{y}))| \\
&\lesssim |\underline{\beta}\cdot(\mathcal{F}(\mathbf{s}+M\mathbf{z}+qM\mathbf{y}+qM\mathbf{t}) - \mathcal{F}(\mathbf{s}+M\mathbf{z}+qM\mathbf{y}))|_\infty \\
&\lesssim_k |\underline{\beta}|_\infty N^{k-1}qM \\
&\le (N/M)^{-k+(k-1)r\theta_0}N^{k-1}(N/M)^{r(k-1)\theta_0}(N/M)^\eta \\
&\le N^{-k+(k-1)r\theta_0+(k-1)+r(k-1)\theta_0+\eta} = N^{-1+2(k-1)r\theta_o+\eta}
\end{aligned}$$

where we used $N/M \leq N$. Now observe that

$$\begin{aligned}&\phi_N(\mathbf{s}+M\mathbf{z}+qM\mathbf{y}) \neq \phi(\mathbf{s}+M\mathbf{z}+qM\mathbf{y}+qM\mathbf{t})\\ &\iff \mathbf{y} \in E := \Big([0,N/qM]^d - (\mathbf{s}+M\mathbf{z})/qM\Big)\Delta\Big([0,N/qM]^d - (\mathbf{s}+M\mathbf{z})/qM - \mathbf{t}\Big).\end{aligned}$$

Since we can write E as a union of d boxes with one side has length $O(1)$, the number of $\mathbf{y}$ for which $\phi_N(\mathbf{s}+M\mathbf{z}+qM\mathbf{y}) \neq \phi_N(\mathbf{s}+M\mathbf{z}+qM\mathbf{y}+qM\mathbf{t})$ is bounded above by $|E| \lesssim (N/qM)^{d-1}$.

Hence we can replace the inner sum with the integral, with the error

$$\begin{aligned}&\Big|\int_{\mathbf{y}\in\mathbb{R}^d} e^{2\pi i\underline{\beta}\cdot\mathcal{F}(qM\mathbf{y}+M\mathbf{z}+\mathbf{s})}\mathbf{1}_{[0,N]^d}(qM\mathbf{y}+M\mathbf{z}+\mathbf{s})d\mathbf{y} - \sum_{\mathbf{y}\in\mathbb{Z}^d} e^{2\pi i\underline{\beta}\cdot\mathcal{F}(qM\mathbf{y}+M\mathbf{z}+\mathbf{s})}\mathbf{1}_{[0,N]^d}(qM\mathbf{y}+M\mathbf{z}+\mathbf{s})\Big|\\
&= \Big|\sum_{\mathbf{y}\in\mathbb{Z}^d}\int_{\mathbf{t}\in[0,1]^d}\Big(e(\underline{\beta}\cdot\mathcal{F}(\mathbf{s}+M\mathbf{z}+qM\mathbf{y}+qM\mathbf{t}))\phi_N(\mathbf{s}+M\mathbf{z}+qM\mathbf{y}+qM\mathbf{t})\\
&- e(\underline{\beta}\cdot\mathcal{F}(\mathbf{s}+M\mathbf{z}+qM\mathbf{y})\phi_N(\mathbf{s}+M\mathbf{z}+qM\mathbf{y})\Big)d\mathbf{t}\Big|\\
&\lesssim \sum_{\mathbf{y}\in E} 1 + \sum_{\mathbf{y}\notin E}\Big|\int_{\mathbf{t}\in[0,1]^d}(e(\underline{\beta}\cdot\mathcal{F}(\mathbf{s}+M\mathbf{z}+qM\mathbf{y}+qM\mathbf{t}) - e(\underline{\beta}\cdot\mathcal{F}(\mathbf{s}+M\mathbf{z}+qM\mathbf{y}))d\mathbf{t}\phi_N(\mathbf{s}+M\mathbf{z}+Mq\mathbf{y})\Big|\\
&\lesssim (N/qM)^{d-1} + (N/qM)^d N^{-1+2r(k-1)\theta_0+\eta} = O((N/qM)^d N^{-1+2r(k-1)\theta_0+\eta}).\end{aligned}$$

By a change of variables $N^{-1}(qM\mathbf{y}+M\mathbf{z}+\mathbf{s}) \mapsto \mathbf{y}$, we have

$$\int_{\mathbf{y}\in\mathbb{R}^r} e^{2\pi i\beta\cdot\mathcal{F}(qM\mathbf{y}+M\mathbf{z}+\mathbf{s})}\mathbf{1}_{[0,N]^r}(qM\mathbf{y}+M\mathbf{z}+\mathbf{s})d\mathbf{y} = N^d M^{-d} q^{-d} I(N^k\underline{\beta})$$

Substituting the estimate in (2.3.30) and summing over $\mathbf{z} \in \mathbb{Z}_q^d$, we have (2.3.29). □

The Singular Integral

Let $\mu \in \mathbb{R}^r$ and $\Phi > 0$. Recall $I(\underline{\gamma}) := \int_{\mathbb{R}^d} e^{2\pi i\underline{\gamma}\cdot\mathcal{F}(\mathbf{y})}\mathbf{1}_{[0,1]^d}(\mathbf{y})d\mathbf{y}$, write

$$\begin{aligned}J(\mu;\Phi) &= \int_{|\underline{\gamma}|_\infty\leq\Phi} I(\underline{\gamma})e^{-2\pi i\underline{\gamma}\cdot\mu}d\underline{\gamma}\\ J(\mu) &:= \lim_{\Phi\to\infty} J(\mu;\Phi)\end{aligned} \tag{2.3.31}$$

Lemma 2.3.13 ([11], Lemma 5.2, Lemma 5.3, section 6)**.** *$J(\mu)$ exists, continuous and uniformly bounded by*

$$\int_{R^r} |I(\underline{\gamma})|d\underline{\gamma} < \infty$$

Furthermore if $\mathcal{F}(M\mathbf{x}+\mathbf{s})=\mathbf{v}$ has a nonsingular real solutions in $[\delta,1-\delta]^d$ then $J(N^{-k}\mathbf{v})\geq c(\delta)>0$.

Proof. (Sketch): Let B be a box of size less than 1 and $\underline{\lambda}\in\mathbb{R}^r$. Define

$$I(B,\underline{\lambda})=\int_B e^{2\pi i\underline{\lambda}\cdot\mathcal{F}(\mathbf{y})}d\mathbf{y}=(N/M)^{-d}\int_{(N/M)B} e^{2\pi i(N/M)^{-k}\underline{\lambda}\cdot\mathcal{F}(\mathbf{y})}d\mathbf{y}$$

Then the claim in Lemma 2.3.13 follows from the bound

$$|I(B,\underline{\lambda})|\lesssim C_\varepsilon(1+|\underline{\lambda}|_\infty)^{-\frac{K}{(k-1)r}+\varepsilon} \tag{2.3.32}$$

To see this bound, assume $|\underline{\lambda}|_\infty>1$ in the $I(B,\underline{\lambda})$. Let $\underline{\alpha}=(N/M)^{-k}\underline{\lambda}$ and choose θ so that $|\underline{\lambda}|_\infty=(N/M)^{r(k-1)\theta}$. Thus, we have $\underline{\alpha}\in\mathcal{M}'_{0,1}(\theta)$ is on the edge and $\underline{\alpha}\notin\mathcal{M}(\theta')\ \forall\theta'<\theta$. Apply the bound on minor arc of θ' and use the fact that $\underline{\alpha}\in\mathcal{M}_{0,1}(\theta)$ with (2.3.29) to estimate the sum by the integral. Note that $r(k-1)\theta\leq 1$ if $N/M\geq|\underline{\lambda}|_\infty^{2/k}$. Since $\underline{\alpha}\notin\mathcal{M}(\theta')$, we have

$$|S_N(M,\mathbf{s},\underline{\alpha})|\lesssim(N/M)^{d+\varepsilon}((N/M)^k|\underline{\alpha}|_\infty)^{-\frac{K}{(k-1)r}}\lesssim(N/M)^{d+\varepsilon}|\underline{\lambda}|_\infty^{-\frac{K}{(k-1)r}}$$

and as in the proof of Lemma 2.3.12,

$$|(N/M)^dI(B,\underline{\lambda})-S_N(M,\mathbf{s},\underline{\alpha})|\lesssim(|\underline{\alpha}|_\infty(N/M)^{k-1})(N/M)^d+(N/M)^{d-1}\lesssim(\frac{|\underline{\lambda}|_\infty}{N/M})(N/M)^d.$$

Choosing large enough N that is $N/M\geq|\underline{\lambda}|_\infty^{1+\frac{K}{(k-1)r}}$ so that $\frac{|\underline{\lambda}|_\infty}{N/M}\leq|\underline{\lambda}|_\infty^{-\frac{K}{(k-1)r}}$, we obtain

$$|I(B,\underline{\lambda})|\lesssim_\varepsilon|\underline{\lambda}|_\infty^{-\frac{K}{(k-1)r}+\varepsilon}$$

□

To see that $J(\mathbf{u})$ is positive, first consider the contribution of singular points; cover $[0,1]^d$ with boxes of sidelength δ. cover $C=V^*_{\mathcal{F}}\cap[0,1]^d$ with $\lesssim\delta^{-dim(V^*_{\mathcal{F}})}$ cubes of size δ. We can show using (2.3.32) that the contribution from these cubes is $\lesssim\delta^{d-k-dim(V^*_{\mathcal{F}})}$ which is negligible if $d>k+dim(V^*_{\mathcal{F}})$ which can be verified. Now we consider only non-singular points $\mathbf{y}$. Let $B=[0,1]^d\backslash C$. Take local coordinate $u_{r+1},\dots,u_d$ such that the Jacobian

$$|Jac_{\mathcal{F}}|=\left|\frac{\partial(f_1,\dots,f_r,u_{r+1},\dots,u_d)}{\partial(x_1,\dots x_d)}\right|$$

is a nonzero. Following the calculations in [11] Lemma 6.3 (skipping some technical details),we have

$$J(\mathbf{u}) = \int_{\mathcal{F}(\mathbf{y})=\mathbf{u},\mathbf{y}\in B} \phi(\mathbf{y})d\sigma(\mathbf{y})$$

where $d\sigma_{\mathcal{F}} = dS_{\mathcal{F}}/|Jac_{\mathcal{F}}|$. Here $dS_{\mathcal{F}}$ is the surface measure of $\mathcal{F}(\mathbf{y}) = \mathbf{u}$.

Let $\overline{B}_\delta$ be the closed ball, suppose $\mathbf{y}_0 \in (0,1)^d$,the interior of $[0,1]^d$, $\mathbf{y}_0 + \overline{B}_\delta \subseteq [0,1]^d$ is such that $\mathcal{F}(\mathbf{y}_0) = \mathbf{u}, |Jac_{\mathcal{F}}(\mathbf{y}_0)| \neq 0$. Then $|Jac_{\mathcal{F}}(\mathbf{y})| \geq c(\delta, \mathbf{y}_0) > 0$ for all $\mathbf{y} \in \mathbf{y}_0 + \overline{B}_\delta$ and $\{\mathcal{F}(\mathbf{y}) = \mathbf{u}\} \cap \{\mathbf{y}_0\} + \overline{B}_\delta$ is a $d-r$ dimensional surface with positive measure. Hence

$$J(\mathbf{u}) \geq \int_{\{\mathcal{F}(\mathbf{y})=\mathbf{u}\}\cap\{\mathbf{y}_0\}+\overline{B}_\delta} \frac{dS_{\mathcal{F}}(\mathbf{y})}{|Jac_{\mathcal{F}}(\mathbf{y})|} \geq \eta(\mathcal{F}, \mathbf{y}_0) > 0.$$

To get the bound independent of $\mathbf{y}_0$, we need such non-singular $\mathbf{y}_0 \in [\delta, 1-\delta]^d$ (a closed set in the interior).

The Singular Series

We define the *singular series*

$$\mathfrak{S}(M, \mathbf{s}; \mathbf{v}) := \sum_{q=1}^{\infty} q^{-d} \sum_{(\mathbf{a},q)=1} e^{-2\pi i \frac{\mathbf{a}\cdot\mathbf{v}}{q}} S_{\mathbf{a},q}(M, \mathbf{s}) \tag{2.3.33}$$

We have by the assumption (2.3.24)

$$\frac{2\delta}{r(k-1)\theta_0} < \frac{K}{(1+\epsilon)r(k-1)} - r - 1$$

Then using the Gauss sum estimate (2.3.21) and recalling $M \leq N^{\frac{\eta}{1+\eta}}$, one has

$$\sum_{q\geq(N/M)^{r(k-1)\theta_0}} \sum_{(\mathbf{a},q)=1} q^{-d}|S_{\mathbf{a},q}(d,\mathbf{s})| \lesssim_\tau \sum_{q\geq(N/M)^{r(k-1)\theta_0}} q^{-\frac{2\delta}{r(k-1)\theta_0}+\tau} \lesssim_\tau (N/M)^{-2\delta+\tau} \lesssim_\tau N^{-2\delta+\delta\eta+\tau} \lesssim N^{-\delta} \tag{2.3.34}$$

Hence the infinite sum defining $\mathfrak{S}(M, \mathbf{s}; \mathbf{v})$ is absolutely convergent.

Finally we analyze the singular series.To express it in terms of the density of solutions in $\mathbb{Z}_{p^l}$.

Theorem 2.3.14 (Singular Series)**.** *Consider the singular series*

$$\mathfrak{S}(M, \mathbf{s}; \mathbf{v}) := \sum_{q=1}^{\infty} \sum_{(\mathbf{a},q)=1} q^{-d} e^{-2\pi i \frac{\mathbf{a}\cdot\mathbf{v}}{q}} S_{\mathbf{a},q}(M, \mathbf{s}), \quad S_{\mathbf{a},q}(M, \mathbf{s}) := \sum_{\mathbf{x}\in\mathbb{Z}_q^d} e^{2\pi i \frac{\mathbf{a}\cdot\mathcal{F}(M\mathbf{x}+\mathbf{s})}{q}}. \tag{2.3.35}$$

We have

$$\mathfrak{G}(M,\mathbf{s},\mathbf{v}) = \prod_{p \text{ prime}} \sigma_p(M,\mathbf{s},\mathbf{v}) \ \ \textit{where} \ \ \sigma_p(M,\mathbf{s};\mathbf{v}) = \lim_{l\to\infty} \sigma_p^{(l)}(M,\mathbf{s};\mathbf{v})$$

where

$$\sigma_p^{(l)}(M,\mathbf{s},\mathbf{v}) = p^{-l(d-r)} |\{\mathbf{x} \in \mathbb{Z}_{p^l}^d ; \mathcal{F}(M\mathbf{x}+\mathbf{s}) = \mathbf{v} \pmod{p^l}\}|$$

The infinite product converges absolutely and uniformly in $\mathbf{v}$ *and the singular series is positive if* σ_p *is positive for all* p*. A sufficient condition is that* $\mathcal{F}(M\mathbf{x}+\mathbf{s}) = \mathbf{v}$ *has a non-singular solution* $\pmod{p}$ *for every* p*.*

Proof. (Sketch) Since the summand

$$q^{-d} e^{-2\pi i \frac{\mathbf{a}\cdot\mathbf{v}}{q}} S_{\mathbf{a},q}(M,\mathbf{s})$$

is multiplicative in q (by multiplicativity of Gauss sum), we can formally write

$$\mathfrak{G}(M,\mathbf{s},\mathbf{v}) = \prod_{p \text{ prime}} \sigma_p(M,\mathbf{s},\mathbf{v})$$

where

$$\sigma_p(M,\mathbf{s},\mathbf{v}) = \sum_{m=0}^{\infty} p^{-md} \sum_{(\mathbf{a},p^m)=1} e^{-2\pi i \frac{\mathbf{a}\cdot\mathbf{v}}{p^m}} S_{\mathbf{a},p^m}(M,\mathbf{s})$$

If p is sufficiently large,we may apply the Gauss sum estimate (2.3.21) together with assumption (2.3.22), we have

$$\sigma_p(M,\mathbf{s},\mathbf{v}) = 1 + \sum_{m=1}^{\infty} O(p^{(-1-\frac{1}{(1+\varepsilon)r(k+1)}+\tau)m}) = 1 + O(p^{-\delta'}) \tag{2.3.36}$$

for some $\delta' > 1$ and hence the product is absolutely and uniformly convergent in $\mathbf{v}$. Finally, we do a routine calculation as in [11],

$$\begin{aligned} \sigma_p^{(l)}(M,\mathbf{s};\mathbf{v}) &= \sum_{m=0}^{l} p^{-md} \sum_{(\mathbf{a},p^m)=1} e^{-2\pi i \frac{\mathbf{a}\cdot\mathbf{v}}{p^m}} S_{\mathbf{a},p^m}(M,\mathbf{s}) \\ &= p^{-l(d-r)} |\{\mathbf{x} \in \mathbb{Z}_{p^l}^d ; \mathcal{F}(M\mathbf{x}+\mathbf{s}) = \mathbf{v} \pmod{p^l}\}| \end{aligned} \tag{2.3.37}$$

Now to verify positivity, by the bound (2.3.36), we have

$$|\prod_{p\geq P} \sigma_p(M,\mathbf{s},\mathbf{v}) - 1| \lesssim \sum_{p\geq P} |\sigma_p(M,\mathbf{s},\mathbf{v}) - 1| \leq p^{-\delta''}, \ \delta'' > 0$$

Next we check the positivity of $\sigma_p(M,\mathbf{s},\mathbf{v})$ for small p if there is a nonsingular solution $\pmod p$. Arguing as in [11] or in the proof of Lemma 2.4.5 below shows us that this is indeed the case. $\square$

Now we can prove the main theorem of this section.

Proof of Theorem 2.3.4. First we claim that if $0<\epsilon\le 1$ satisfies $\epsilon<\frac{K}{r(r+1)(k-1)}-1$ and $\eta>0$ such that

$$\eta<\frac{1}{4r(r+2)k}\min\{\epsilon,\frac{K-(1+\epsilon)r(r+1)(k-1)}{rk(1+\epsilon)}\} \tag{2.3.38}$$

then (2.3.10) holds for $1\le M\le N^{\frac{\eta}{1+\eta}}$ and $\mathbf{s}\in\mathbb{Z}^d$.

If we have this claim, since $K>r(r+1)(k-1)$ we have $\epsilon>\frac{K}{r(r+1)(k-1)}-1\ge\frac{1}{r(r+1)(k-1)}$ then

$$\frac{K-(1+\epsilon)r(r+1)(k-1)}{rk(1+\epsilon)}\ge\frac{1}{rk}$$

Hence choosing ϵ slightly larger than $\frac{1}{r(r+1)(k-1)}$, one has $\eta\le\frac{1}{4r^2(r+1)(r+2)k^2}$ by (2.3.38). Now under assumption (2.3.11) that $K>2r(r+1)(k-1)+2rk$, we take ϵ slightly larger than 1 then $\eta\le\frac{1}{4r(r+2)k}$. So we now only need to verify the claim.

Set the parameters θ_0 and δ as

$$\theta_0:=\frac{1}{2r(r+2)k+1},\quad \delta:=\frac{\theta_0}{2}\min\{1,\frac{K}{1+\epsilon}-r(r+1)(k-1)\}$$

Then θ_0,δ satisfy (2.3.23), (2.3.24). Set η as (2.3.38) above. We have

$$\eta<\epsilon(1-k^{-1})\theta_0\ \text{ and }\ \eta<\delta k^{-1}r^{-1}$$

Hence the condition of Lemma 2.3.11 and the condition (2.3.25) are satisfied. Also note that

$$2r(k-1)\theta_0+\eta\le\frac{k-1}{k(r+2)}+\frac{1}{4r(r+2)k}\le\frac{1}{r+2}(1-\frac{3}{4k})<\frac{1}{3}.$$

Hence

$$|\mathcal{M}'(\theta_0)|\le(N/M)^{(r+1)r(k-1)\theta_0-rk+r\eta}\le N^{-rk+2/3} \tag{2.3.39}$$

By (2.3.26) we have (for some $\delta'=\delta'(r,k)>0$) ,

$$\mathcal{R}_N(M,\mathbf{s},\mathbf{v})=\sum_{q\le(N/M)^{r(k-1)\theta_0}}\ \sum_{\mathbf{a},(\mathbf{a},q)=1}\int_{\mathcal{M}'(\mathbf{a},q)}e^{-2\pi i\underline{\alpha}\cdot\mathbf{v}}S_N(M,s;\alpha)d\alpha+O(N^{d-rk-\delta'}M^{-d})$$

By Lemma 2.3.12 and the size of major arc (2.3.39), this becomes

$$N^d M^{-d}\Big(\sum_{q\le (N/M)^{r(k-1)\theta_0}} q^{-d} e^{-2\pi i\frac{\mathbf{a}\cdot\mathbf{v}}{q}} S_{\mathbf{a},q}(M,\mathbf{s}) \int_{|\beta_i|\le (N/M)^{-k+r(k-1)\theta_0}} e^{-2\pi i\underline{\beta}\cdot\mathbf{v}} I(N^k\underline{\beta})d\underline{\beta}+O(N^{rk-\frac{1}{3}+2r(k-1)\theta_0+\cdots}$$

Rescaling $\underline{\beta} := N^k\underline{\beta}$, this becomes

$$N^{d-rk}M^{-d}\Big(\sum_{q\le (N/M)^{r(k-1)\theta_0}} q^{-d} e^{-2\pi i\frac{\mathbf{a}\cdot\mathbf{v}}{q}} S_{\mathbf{a},q}(M,\mathbf{s}) J(N^{-k}\mathbf{v}; M^k(N/M)^{r(k-1)\theta_0}) + O(N^{-\delta'})\Big)$$

Applying (2.3.31) and (2.3.34) we have

$$\mathcal{R}_N(d,\mathbf{s},\mathbf{v}) = N^{d-rk}M^{-d}\mathfrak{S}(d,\mathbf{s};\mathbf{v})J(N^{-k}\mathbf{v}) + O(N^{d-rk-\delta'}M^{-d}) \tag{2.3.40}$$

as required. □

2.4 Almost Prime Solutions to Diophantine Equations

For $0<\varepsilon<1$ and $N\ge 1$ let $\mathcal{P}_\varepsilon[N]$ denote the set of natural numbers $m\le N$ such that each prime divisor of m is at least N^ε. Note that each $m\in\mathcal{P}_\varepsilon[N]$ at most $\lfloor 1/\varepsilon\rfloor$ prime factors. We call sets of the form $\mathcal{P}_\varepsilon[N]$ "*almost prime*". For given $\mathbf{v}\in\mathbb{Z}^d$, let

$$\mathcal{M}^\varepsilon_{\mathcal{F}}[N] := |\{\mathbf{x}\in\mathcal{P}_\varepsilon[N]^d;\ \mathcal{F}(\mathbf{x})=\mathbf{v}\}|,$$

denote the number of almost prime solutions $x\in[1,N]^d$ to the system $\mathcal{F}(\mathbf{x})=\mathbf{v}$. Let $\mathbb{U}^d_{p^t}$ denote the multiplicative group of reduced residue classes $\pmod{p^t}$. Let $M(p^t,\mathbf{v})$ represents the number of solutions to the equation $\mathcal{F}(\mathbf{x})=\mathbf{v}$ in $\mathbb{U}^d_{p^t}$.

For each prime p, define the local density

$$\sigma^*_p(\mathbf{v}) := \lim_{t\to\infty}\frac{(p^t)^r\mathcal{M}(p^t,\mathbf{v})}{\phi(p^t)^d} \tag{2.4.1}$$

provided the limit exists. As almost primes are concentrated in reduced residue classes, the general *local to global principle* suggests that

$$\mathcal{M}^\varepsilon_{\mathcal{F}}[N] \approx_\varepsilon N^{d-kr}\,(\log\ N)^{-d}J(N^{-k}\mathbf{v})\prod_p\sigma^*_p(\mathbf{v}), \tag{2.4.2}$$

as $N\to\infty$ where $J(\mathbf{u})$ is the singular integral. Our main result in this section is the following.

Theorem 2.4.1. *Let $\mathcal{F}=(F_1,\ldots,F_r)$ be a system of r integral forms of degree $k\ge 2$ in d variables such that*

$$Rank(\mathcal{F}) > r(r+1)(k-1)2^{k-1}. \tag{2.4.3}$$

Then there exists a constant $\varepsilon = \varepsilon(n,k) > 0$ such that

$$\mathcal{M}^{\varepsilon}_{\mathcal{F}}[N] \geq c_{d,k}\, N^{d-kr}\,(\log\, N)^{-d} J(N^{-k}\mathbf{v}) \prod_p \sigma^*_p(\mathbf{v}). \tag{2.4.4}$$

Moreover, if $\mathcal{F}(\mathbf{x}) = \mathbf{v}$ has a nonsingular solution in $\mathbb{U}_p$, the p-adic integer units, for all primes p, then

$$\prod_p \sigma^*_p(\mathbf{v}) > 0.$$

The key to prove Theorem 2.4.1 is to study a weighted sum over the solutions with weights that are concentrated on numbers having few prime factors. Such weights have been mentioned in section 2.2 which we recall here. For given $0 < \eta < 1$, let $R := N^\eta$ and χ is some smooth compactly supported functions, define

$$\Lambda_R(m) := \sum_{d|m} \mu(d)\chi(\frac{\log d}{\log R}).$$

We will also employ the "W-trick" to bypass the contribution of small primes in our initial asymptotic formulas. Let $\omega = \omega_{\mathcal{F}} > 1$ be a fixed positive integer depending only on the system $\mathcal{F}$ and let $W := \prod_{p\leq\omega} p$, the product of primes up to ω. Note that if $\mathbf{x} \in \mathcal{P}_\varepsilon[N]^d$ and $p|x_i$ implies $p \geq N^\varepsilon > \omega_{\mathcal{F}}$ for sufficiently large N, hence $(x_i, W) = 1$ for each $1 \leq i \leq d$. We will write $(\mathbf{x}, W) = 1$ in this case. Under the conditions of Theorem 2.4.1 our key estimates are the following

Theorem 2.4.2. *Let $\mathcal{F} = (F_1, \ldots, F_r)$ be a system of r integral forms of degree $k \geq 2$ in d variables satisfying the rank condition (2.3.8). Let $0 < \eta < \frac{1}{4r^2(r+1)(r+2)k(k+1)}$, $R = N^\eta$ and $W = \prod_{p\leq\omega} p$. Then one has*

$$\sum_{\substack{\mathbf{x}\in[N]^d \\ (\mathbf{x},W)=1,\ \mathcal{F}(\mathbf{x})=\mathbf{v}}} \Lambda^2_R(x_1x_2\cdots x_d) = N^{d-rk}(\log\, R)^d J(N^{-k}\mathbf{v}) \prod_{p|W} \sigma^*_p(\mathbf{v})\,(1 + o_{N,W\to\infty}(1)), \tag{2.4.5}$$

moreover for given $0 < \varepsilon < \eta$

$$\sum_{\substack{\mathbf{x}\in[N]^d,\ \mathbf{x}\notin\mathcal{P}^\varepsilon[N]^d \\ \mathcal{F}(\mathbf{x})=\mathbf{v}}} \Lambda^2_R(x_1x_2\cdots x_d) \lesssim \frac{\varepsilon}{\eta}\, N^{d-rk}(\log\, R)^d J(N^{-k}\mathbf{v}) \prod_{p|W} \sigma^*_p(\mathbf{v}). \tag{2.4.6}$$

In the proof of Theorem 2.4.2 we will use the asymptotic for the number of integer solutions $\mathbf{x} \in [N]^d$

to $\mathcal{F}(\mathbf{x}) = \mathbf{v}$ subject to the congruence condition $\mathbf{x} \equiv \mathbf{s} \pmod{M}$, where M is a small modulus bounded by a sufficiently small power N This is summarized in Theorem 2.3.4.

2.4.1 Local Factors of Integral Forms

The following proposition summarizes the properties of the Euler factors we will need. Recall the density of solutions in p-adic numbers,

$$\sigma_P(M, \mathbf{s}, \mathbf{v}) = \lim_{l\to\infty} \sigma_p^{(l)}(M, \mathbf{s}, \mathbf{v})$$

where

$$\sigma_p^{(l)}(M, \mathbf{s}, \mathbf{v}) = p^{-l(d-r)}|\{\mathbf{x} \in \mathbb{Z}_{p^l}^d : \mathcal{F}(M\mathbf{x} + \mathbf{s}) \equiv \mathbf{v} \pmod{p^l}\}|$$

Define the Euler's factor that will appear in the asymptotic of the sum of in (2.4.5).

$$\gamma_p(\mathbf{v}) := \frac{p^{-d}}{\sigma_p(\mathbf{v})} \sum_{\substack{\mathbf{s}\in\mathbb{Z}_p^d \\ \mathcal{F}(s)\equiv\mathbf{v} \pmod{p}}} \mathbf{1}_{p|s_1\ldots s_d}\sigma_p(p, \mathbf{s}, \mathbf{v})$$

The key property we will need is

Proposition 2.4.3. *If $\mathcal{F}$ is a family of r integral forms of degree k such that rank$(\mathcal{F}) > r(r+1)(k-1)2^k$ then for all sufficiently large primes $p > \omega_{\mathcal{F}}$ we have*

$$\gamma_p(\mathbf{v}) = \frac{d}{p} + O(p^{-2}) \tag{2.4.7}$$

We start with a simple observation on the local densities of solutions.

Lemma 2.4.4. *Let M, W be square free numbers such that $(M, W) = 1$ and let p be a prime. If $(M, W) = 1$ then*

$$\sigma_p(MW, \mathbf{t}, \mathbf{v}) = \sigma_p(\mathbf{v}) \tag{2.4.8}$$

If $p|M$ and $\mathbf{t} \equiv \mathbf{s} \pmod{M}$ then one has

$$\sigma_p(MW, \mathbf{t}, \mathbf{v}) = \sigma_p(p, \mathbf{t}, \mathbf{v}) = \sigma_p(p, \mathbf{s}, \mathbf{v}) \tag{2.4.9}$$

The analogue statement indeed holds when we interchange M and W.

Proof. For any $l \in \mathbb{Z}_+$, since $(p^l, MW) = 1$, we have the transformation $\mathbf{x} \mapsto MW\mathbf{x} + \mathbf{t}$ is a bijection on $\mathbb{Z}_{p^l}^d$. Hence

$$|\{\mathbf{x} \in \mathbb{Z}_{p^l}^d : \mathcal{F}(\mathbf{x}) \equiv \mathbf{v} \pmod{p^l}\}| = |\{\mathbf{x} \in \mathbb{Z}_{p^l}^d : \mathcal{F}(MW\mathbf{x} + \mathbf{t}) \equiv \mathbf{v} \pmod{p^l}\}|$$

We have (2.4.8). Next assume $p|M$ then $M = pM'$ with $(p, M'W) = 1$ then one may write $MW\mathbf{x} + \mathbf{t} = p(M'W\mathbf{x}) + \mathbf{t}$ and note that $\mathbf{x} \mapsto M'W\mathbf{x}$ is a bijection on $\mathbb{Z}_{p^l}^d$ hence

$$|\{\mathbf{x} \in \mathbb{Z}_{p^l}^d : \mathcal{F}(MW\mathbf{x} + \mathbf{t}) \equiv \mathbf{v} \pmod{p^l}\}| = |\{\mathbf{x} \in \mathbb{Z}_{p^l}^d : \mathcal{F}(p\mathbf{x} + \mathbf{t}) \equiv \mathbf{v} \pmod{p^l}\}|$$

which establishes the first equality in (2.4.9). To see the second equality of (2.4.9) we write $p\mathbf{y} + \mathbf{t} = p(\mathbf{y} + \mathbf{u}) + \mathbf{s}$ where $\mathbf{y} \mapsto \mathbf{y} + \mathbf{u}$ is a bijection on $\mathbb{Z}_{p^l}^d$. $\square$

Recall that if the local factor $\sigma_p(p, \mathbf{s}, \mathbf{v})$ does not vanish then $\mathcal{F}(\mathbf{s}) \equiv \mathbf{v} \pmod{p}$. We call a point $\mathbf{s} \in \mathbb{Z}_p^d$ non-singular if the Jacobian $Jac_F(\mathbf{s})$ has full rank r over $\mathbb{Z}_p$. We show that under this rank condition, it is easy to calculate $\sigma_p^{(l)}(p, \mathbf{s}, \mathbf{v})$ explicitly.

Lemma 2.4.5. *Let* $\mathbf{s}$ *be a non-singular solution to* $\mathcal{F}(\mathbf{s}) \equiv \mathbf{v} \pmod{p}$. *Then for all* l*,*

$$\sigma_p^{(l)}(p, \mathbf{s}, \mathbf{v}) = p^r$$

Proof. We do induction on l. For $l = 1$, we have $\mathcal{F}(p\mathbf{x} + \mathbf{s}) \equiv \mathcal{F}(\mathbf{s}) \equiv \mathbf{v} \pmod{p}$ for all $\mathbf{x} \in \mathbb{Z}_p^d$. Hence $\sigma_p^{(1)}(p, \mathbf{s}, \mathbf{v}) = p^r$. For $l = 2$, we count $\mathbf{x} \in \mathbb{Z}_{p^2}^d$ satisfying

$$\mathcal{F}(p\mathbf{x} + \mathbf{s}) \equiv \mathcal{F}(\mathbf{s}) + pJac_{\mathcal{F}}(\mathbf{s}) \cdot \mathbf{x} \equiv \mathbf{v} \pmod{p^2}.$$

Since $\mathcal{F}(\mathbf{s}) - \mathbf{v} = p\mathbf{u}$ for some $\mathbf{u} \in \mathbb{Z}_p^d$, this is

$$Jac_{\mathcal{F}}(\mathbf{s}) \cdot \mathbf{x} \equiv -\mathbf{u} \pmod{p}.$$

Since $Jac_{\mathcal{F}}(\mathbf{s})$ has full rank, the above equation has p^{n-r} solutions in $\mathbb{Z}_p^d$ and p^{2n-r} solutions in $\mathbb{Z}_{p^2}^d$. Hence $\sigma_p^{(2)}(p, \mathbf{s}, \mathbf{v}) = p^r$.

For $l \geq 3$, we show

$$\sigma_p^{(l)}(p, \mathbf{s}, \mathbf{v}) = \sigma_p^{(l-1)}(p, \mathbf{s}, \mathbf{v})$$

Now if $\mathbf{x} \equiv \mathbf{y} \pmod{p^{l-1}}$ then $p\mathbf{x} + \mathbf{s} \equiv p\mathbf{x} + \mathbf{y} \pmod{p^l}$ then $\mathcal{F}(p\mathbf{x} + \mathbf{s}) \equiv \mathcal{F}(p\mathbf{y} + \mathbf{s}) \pmod{p^l}$. For given $\mathbf{y} \in \mathbb{Z}_{p^{l-1}}^d$, we can uniquely write $\mathbf{y} = p^{l-2}\mathbf{u} + \mathbf{z}$ with $\mathbf{z} \in \mathbb{Z}_{p^{l-2}}^d$ and $\mathbf{u} \in \mathbb{Z}_p^d$. Then

$$\mathcal{F}(p\mathbf{y} + \mathbf{s}) \equiv \mathcal{F}(p^{l-1}\mathbf{u} + p\mathbf{z} + \mathbf{s}) \equiv \mathcal{F}(p\mathbf{z} + \mathbf{s}) + p^{l-1}Jac_{\mathcal{F}}(\mathbf{s}) \cdot \mathbf{u} \pmod{p^l} \tag{2.4.10}$$

Hence $\mathcal{F}(p\mathbf{y} + \mathbf{s}) \equiv \mathbf{v} \pmod{p^l}$ implies

$$\mathcal{F}(p\mathbf{z} + \mathbf{s}) \equiv \mathbf{v} \pmod{p^{l-1}}. \tag{2.4.11}$$

The number of such $\mathbf{z} \in \mathbb{Z}_{p^{l-2}}^d$ is $p^{-d} \times p^{(l-1)(d-r)}\sigma_p^{(l-1)}(p, \mathbf{s}, \mathbf{v})$. For a given $\mathbf{z}$ satisfying (2.4.10), write $\mathcal{F}(p\mathbf{z}+\mathbf{s}) = p^{l-1}\mathbf{b}+\mathbf{v}$, then (2.4.10) holds if and only if

$$Jac_{\mathcal{F}}(\mathbf{s}) \cdot \mathbf{u} \equiv -\mathbf{b} \pmod p \tag{2.4.12}$$

Since $Jac_{\mathcal{F}}(\mathbf{s})$ has full rank r over $\mathbb{Z}_p^d$, the number of solutions of (2.4.12) is p^{d-r}. Since the decomposition $\mathbf{y} = p^{l-2}\mathbf{u}+\mathbf{z}$ is unique it follows that

$$\begin{aligned}
\sigma_p^{(l)}(p,\mathbf{s},\mathbf{v}) &= p^{-l(d-r)} \times |\{\mathbf{x} \in \mathbb{Z}_{p^l}^d ; \mathcal{F}(p\mathbf{x}+\mathbf{s}) \equiv 0 \pmod{p^l}\}| \\
&= p^{-l(d-r)}p^d \times |\{\mathbf{x} \in \mathbb{Z}_{p^{l-1}}^d : \mathcal{F}(p\mathbf{x}+\mathbf{s}) \equiv 0 \pmod{p^l}\}| \\
&= p^{-l(d-r)}p^d p^{-d} p^{(l-1)(d-r)} \sigma_p^{(l-1)}(p,\mathbf{s},\mathbf{v}) p^{d-r} \\
&= \sigma_p^{(l-1)}(p,\mathbf{s},\mathbf{v})
\end{aligned}$$

as required.

$\square$

For singular values we can only obtain an upper bound for $\sigma_p(p,\mathbf{s},\mathbf{v})$. If $\mathbf{s} = \mathbf{v} = \mathbf{0}$, we have $\mathcal{F}(p\mathbf{x}) = p^k\mathcal{F}(\mathbf{x}) \equiv 0 \pmod{p^l}$ which has $\approx p^{(l-k)(d-r)+kd}$ solutions in $\mathbb{Z}_{p^l}^d$ and hence $\sigma_p^{(l)}(p,\mathbf{s},\mathbf{v}) \approx p^{kr}$.

Lemma 2.4.6. *Let $\mathcal{F}$ be a family of r integral linear forms of degree k, assume the rank condition*

$$codim(V_{\mathcal{F}}^*) \geq r(r+1)(k-1)2^k + 1 \tag{2.4.13}$$

then uniformly in $l \in \mathbb{N}$ and $\mathbf{s} \in \mathbb{Z}_p^d$, one has

$$\sigma_p^{(l)}(p,\mathbf{s},\mathbf{v}) \lesssim p^{r^2k} \tag{2.4.14}$$

Proof. By (2.3.37), we have

$$\sigma_p^{(l)}(p,\mathbf{s},\mathbf{v}) = \sum_{m=0}^{l} \sum_{\substack{\mathbf{b}\in\mathbb{Z}_{p^m}^r \\ (b_i,p)=1\exists i}} p^{-md} e^{2\pi i \frac{\mathbf{b}\cdot\mathcal{F}(p\mathbf{x}+\mathbf{s})}{p^m}} S_{\mathbf{b},p^m}(p,\mathbf{s})$$

here $S_{\mathbf{b},p^m}$ is the exponential sum defined in (2.3.20). If $m > rk$ then the Gauss sum estimate (Lemma 2.3.10) applied with $\epsilon = 1/r$ and $K = codim(V_{\mathcal{F}}^*)/2^{k-1}$ gives

$$\sum_{m>rk} \sum_{\substack{\mathbf{b}\in\mathbb{Z}_{p^m}^r \\ (b_i,p)=1\exists i}} p^{-mn}|S_{\mathbf{b},p^m}(p,\mathbf{s})| \lesssim \sum_{m>rk} p^{mr}p^{-\frac{mK}{(r+1)(k-1)}+\tau} \lesssim \sum_{m>rk} p^{-m\tau/2} \lesssim 1$$

for sufficiently small $\tau = \tau(r,k) := \frac{K}{(r+1)(k-1)} - r > 0$ (i.e. sufficiently small $V_{\mathcal{F}}^*$). Here we apply the condition (2.4.13); here τ is the constant such that

$$codim(V_{\mathcal{F}}^{*}) = (r-\tau)(r+1)(k-1)2^{k-1}$$

Now using the trivial bound $p^{-md}|S_{\mathbf{b},p^m}(p,\mathbf{s})| \leq 1$. We have

$$\sigma_p^{(l)}(p,\mathbf{s},\mathbf{v}) = \sum_{m\leq rk} \sum_{\substack{\mathbf{b}\in\mathbb{Z}_{p^m}^r \\ (b_i,p)=1\exists i}} p^{-md} e^{2\pi i \frac{\mathbf{b}\cdot\mathcal{F}(p\mathbf{x}+\mathbf{s})}{p^m}} S_{\mathbf{b},p^m}(p,\mathbf{s}) + O(1) \leq \sum_{m\leq rk} p^{mr} \lesssim p^{r^2 k}$$

This proves (2.4.14).

$\square$

2.4.2 Proof of Theorem 2.4.3

Since $\sigma_p(\mathbf{v})^{-1} = 1 + O(p^{-2})$ for all sufficiently large prime p, it suffices to show that (2.4.7) holds for $\sigma_p(\mathbf{v})\gamma_p(\mathbf{v})$. Now use Lemma 2.4.5 to write

$$\begin{aligned}
\sigma_p(\mathbf{v})\gamma_p(\mathbf{v}) &= p^{-d} \sum_{\mathcal{F}(\mathbf{s})\equiv\mathbf{v} \pmod p} \mathbf{1}_{p|s_1\dots s_d}\sigma_p(p,\mathbf{s},\mathbf{v}) \\
&= p^{-d} \sum_{\mathcal{F}(\mathbf{s})\equiv\mathbf{v} \pmod p} \mathbf{1}_{p|s_1\dots s_d}\sigma_p(p,\mathbf{s},\mathbf{v}) \\
&= p^{-d} \sum_{\substack{\mathcal{F}(\mathbf{s})\equiv 0 \pmod{\mathbf{s}} \\ \mathbf{s}\text{ non-singular}}} \mathbf{1}_{p|s_1\dots s_d}\sigma_p(p,\mathbf{s},\mathbf{v}) + p^{-d} \sum_{\substack{\mathcal{F}(\mathbf{s})\equiv 0 \pmod{\mathbf{s}} \\ \mathbf{s}\text{ singular}}} \mathbf{1}_{p|s_1\dots s_d}\sigma_p(p,\mathbf{s},\mathbf{v}) \\
&= p^{-d+r} \sum_{\mathcal{F}(\mathbf{s})\equiv 0 \pmod p} \mathbf{1}_{p|s_1\dots s_d} - p^{-d+r} \sum_{\substack{\mathcal{F}(\mathbf{s})\equiv 0 \pmod p \\ \mathbf{s}\text{ singular}}} \mathbf{1}_{p|s_1\dots s_d} \\
&+ p^{-d} \sum_{\substack{\mathcal{F}(\mathbf{s})\equiv 0 \pmod p \\ \mathbf{s}\text{ singular}}} \mathbf{1}_{p|s_1\dots s_d}\sigma_p(p.\mathbf{s},\mathbf{v}) \\
&:= \gamma_p^1(\mathbf{v}) + \gamma_p^2(\mathbf{v}) + \gamma_p^3(\mathbf{v})
\end{aligned}$$

Now we need the following facts from algebraic geometry on the singular variety when consider reduced $\pmod p$.

- [95] The codimension of singular variety does not change when the equation defining the variety are considered $\pmod p$. Let $V_{\mathcal{F}}^{*}(p)$ denote the locus of singular points $\mathbf{s} \in \mathbb{Z}_p^d$ of the $\pmod p$−reduced singular variety $V_{\mathcal{F}}(p) = \{\mathbf{s} \in \mathbb{Z}_p^d : \mathcal{F}(\mathbf{s}) = \mathbf{v}\}$ then

$$codim(V_{\mathcal{F}}^{*}(p)) = codim(V_{\mathcal{F}}^{*})$$

for all but finitely many primes p.

– ([61], Prop. 12.1) The number of points over $\mathbb{Z}_p$ on a homogeneous algebraic set V is bounded above by its degree times $p^{\dim V}$.

From these two facts, one has

$$|V^*_{\mathcal{F}}(p)| \lesssim p^{d-codim(V^*_{\mathcal{F}})}$$

where the implicit constant may depend on n, k, r. For sufficiently large $p > \omega$. Also we state some more facts from algebraic geometry which we use later.

Lemma 2.4.7 ([19] Cor 4.)**.** *If $\mathcal{F}$ is a system of r forms then for any subspace M_J of codimension $|J|$ one has*

$$rank(\mathcal{F}|_{M_J}) \geq rank(\mathcal{F}) - r|J|$$

Lemma 2.4.8 ([20], Prop 4.)**.** *Let $v \in \mathbb{F}_p^r$ and $S = \mathcal{F}^{-1}(v)$ where $\mathcal{F} : \mathbb{F}_p^d \to \mathbb{F}_p^r$ is a homogeneous polynomial map of degree k then*

$$\|\mathbf{1}_S - p^{-r}\|_{U^k} \leq (k-1)^{2^{-k}d} p^{2^{-k}(r-codim(V^*_{\mathcal{F}}))}$$

In particular

$$|p^{-d}|S| - p^{-r}| = \|\mathbf{1}_S - p^{-r}\|_{U^1} \leq \|\mathbf{1}_S - p^{-r}\|_{U^k} \lesssim_{k,d} p^{2^{-k}(r-codim(V^*_{\mathcal{F}}))} \tag{2.4.15}$$

Apply lemma 2.4.6 one has for $i = 2, 3$,

$$|\gamma_p^i(\mathbf{v})| \lesssim p^{-d+r^2k} \sum_{\substack{\mathcal{F}(\mathbf{s})\equiv 0 \pmod p \\ \mathbf{s} \text{ singular}}} \mathbf{1}_{p|s_1\ldots s_d} \lesssim p^{-d+r^2k} p^{d-codim(V^*_{\mathcal{F}})} \lesssim p^{r^2k - r(r+1)(k-1)2^{k-1}-1} \lesssim p^{-2}$$

For each $J \subseteq [1, d]$ define the coordinate subspace $M_J = \{\mathbf{s} = (s_1, \ldots, s_d) \in \mathbb{Z}_p^d : s_j = 0\ \forall j \in J\}$. by inclusion-exclusion principle, one has

$$\gamma_p^1(\mathbf{v}) = p^{-d+r} \sum_{j=1}^{d} (-1)^{j-1} \sum_{|J|=j} \sum_{\mathbf{s}\in M_J} \mathbf{1}_{\mathcal{F}(\mathbf{s})=\mathbf{v}} \tag{2.4.16}$$

From Lemma 2.4.7 and our assumption on the rank of the system $\mathcal{F}$, one has that for $2 \leq |J| \leq r+1$,

$$rank(\mathcal{F}|_{M_J}) - r \geq r(r+1)(k-1)2^k - r(r+2) = r[(r+1)(k-1)2^k - (r+2)] - r \geq r2^k \tag{2.4.17}$$

Applying Lemma 2.4.8 to the system $\mathcal{F}$ restricted to the subspace $M_j \cong \mathbb{Z}_p^{d-j}$, then one obtains

$$p^{-(d-j)+r}|\{\mathbf{s} \in M_J; \mathcal{F}(\mathbf{s}) = \mathbf{v}\}| = 1 + O_{k,d}(p^{-\frac{rank(\mathcal{F}|_{M_J})-r}{2^k}+r}) = 1 + O(p^{-\varepsilon'}) \tag{2.4.18}$$

for some $\varepsilon' \geq 0$. Hence from (2.4.16), we have

$$\begin{aligned}\sigma_p^1(\mathbf{v}) = p^{-d+r} \sum_{k=1}^{d} \sum_{\mathbf{s} \in M_{\{k\}}} \mathbf{1}_{\mathcal{F}(\mathbf{s})=\mathbf{v}} &+ p^{-d+r} \sum_{j=2}^{r+1} (-1)^{j-1} \sum_{|J|=j} \sum_{\mathbf{s} \in M_j} \mathbf{1}_{\mathcal{F}(\mathbf{s})=\mathbf{v}} \\ &+ p^{-d+r} \sum_{j=r+2}^{d} (-1)^{j-1} \sum_{|J|=j} \sum_{\mathbf{s} \in M_j} \mathbf{1}_{\mathcal{F}(\mathbf{s})=\mathbf{v}}\end{aligned}$$

For the second term corresponding to $2 \leq j \leq r+1$, by (2.4.18), the total sum contributes $O(p^{-j}) = O(p^{-2})$. For the third term we use the the trivial fact $|M_j| = p^{d-j} \leq p^{d-r-2}$ and hence the third term also contributes $O(p^{-2})$. Hence

$$\sigma_p^1(\mathbf{v}) = p^{-d+r} \sum_{k=1}^{d} \sum_{\mathbf{s} \in M_{\{k\}}} \mathbf{1}_{\mathcal{F}(\mathbf{s})=\mathbf{v}} + O(p^{-2}) = \frac{d}{p} + O(p^{-2})$$

as required.

2.4.3 Sums of Multiplicative Functions

Let $(\mathbf{b}, W) = 1$ and furthermore let us assume the conditions of Theorem 2.4.2 holds. Define

$$S_{N,W,\mathbf{b}}(\mathbf{v}) := \sum_{\substack{\mathbf{x} \equiv \mathbf{b} \pmod{W} \\ \mathcal{F}(\mathbf{x})=v}} \Lambda_R^2(x_1 x_2 \ldots x_d) \mathbf{1}_{[0,N]^d}(\mathbf{x}) \tag{2.4.19}$$

and for a prime $q > \omega$,

$$S_{N,W,\mathbf{b},q}(\mathbf{v}) := \sum_{\substack{\mathbf{x} \equiv \mathbf{b} \pmod{W} \\ \mathcal{F}(\mathbf{x})=v}} \mathbf{1}_{q|x_1 \ldots x_d} \Lambda_R^2(x_1 x_2 \ldots x_d) \mathbf{1}_{[0,N]^d}(\mathbf{x}) \tag{2.4.20}$$

First we show that these sums could be written in terms of Euler factors and Goldston-Yidirim sums. The proof invokes Theorem 2.3.4.

Lemma 2.4.9.

$$S_{W,\mathbf{b}}(N) = N^{d-kr} J(N^{-k}\mathbf{v}) W^{-d} \mathfrak{G}_{W,\mathbf{b}}(\mathbf{v}) \sideset{}{'}\sum_{D;(D,W)=1} h_D(R) \gamma_D(\mathbf{v}) + O(N^{d-rk-\delta}) \tag{2.4.21}$$

$$S_{W,q,\mathbf{b}}(N) = N^{d-kr} J(N^{-k}\mathbf{v}) W^{-d} \mathfrak{G}_{W,\mathbf{b}}(\mathbf{v}) \sideset{}{'}\sum_{D;(D,W)=1} h_D(R)\gamma_{[D,q]}(\mathbf{v}) + O(N^{d-rk-\delta}) \quad (2.4.22)$$

where

$$\mathfrak{G}_{W,\mathbf{b}}(\mathbf{v}) := \prod_{p|W} \sigma_p(p,\mathbf{b},\mathbf{v}) \prod_{p\nmid W} \sigma_p(\mathbf{v}) \quad (2.4.23)$$

$$\gamma_D(\mathbf{v}) := D^{-d} \sum_{\substack{\mathbf{s}\in\mathbb{Z}_D^d \\ \mathcal{F}(\mathbf{s})\equiv\mathbf{v} \pmod D}} \mathbf{1}_{D|s_1\dots s_d} \prod_{p|D} \frac{\sigma_p(p,\mathbf{s},\mathbf{v})}{\sigma_p(\mathbf{v})} \quad (2.4.24)$$

$$h_D(R) := \sum_{[d_1,d_2]=D} \mu(d_1)\mu(d_2)\chi(\frac{\log d_1}{\log R})\chi(\frac{\log d_2}{\log R}) \quad (2.4.25)$$

Proof. By definition (2.4.19),

$$\begin{aligned} S_{W,\mathbf{b}}(N) : &= \sum_{\substack{\mathbf{x}\equiv\mathbf{b}\ (mod\ W) \\ \mathcal{F}(\mathbf{x})=\mathbf{v}}} \Lambda_R^2(x_1x_2\cdots x_d)\, \mathbf{1}_{[0,N]^d}(\mathbf{x}) \\ &= \sum_{\substack{\mathbf{x}\equiv\mathbf{b}\ (mod\ W) \\ \mathcal{F}(\mathbf{x})=\mathbf{v}}} \mathbf{1}_{[0,N]^d}(\mathbf{x}) \sideset{}{'}\sum_{\substack{d_1,d_2 \\ [d_1,d_2]|x_1\cdots x_d}} \mu(d_1)\mu(d_2)\,\chi(\frac{\log d_1}{\log R})\chi(\frac{\log d_2}{\log R}) \\ &= \sideset{}{'}\sum_{D} \sum_{\substack{d_1,d_2 \\ [d_1,d_2]=D}} \mu(d_1)\mu(d_2)\,\chi(\frac{\log d_1}{\log R})\chi(\frac{\log d_2}{\log R}) \sum_{\substack{\mathbf{x}\equiv\mathbf{b}\ (mod\ W) \\ \mathcal{F}(\mathbf{x})=\mathbf{v}}} \mathbf{1}_{D|x_1\cdots x_d}\mathbf{1}_{[0,N]^d}(\mathbf{x}) \end{aligned} \quad (2.4.26)$$

Since $(\mathbf{b},W)=1$, the inner sum of the last line of (2.4.26) is zero unless $(D,W)=1$ which we assume from now on. The condition $\mathbf{x}\equiv\mathbf{b} \pmod W$ and $D|x_1\dots x_d$ depends only on $\mathbf{x} \pmod{DW}$ thus one may write

$$\sum_{\substack{\mathbf{x}\equiv\mathbf{b}\ (mod\ W) \\ \mathcal{F}(\mathbf{x})=\mathbf{v}}} \mathbf{1}_{D|x_1\cdots x_d}\mathbf{1}_{[0,N]^d}(\mathbf{x}) = \sum_{\substack{\mathbf{t}\in\mathbb{Z}_{DW}^d,\mathbf{t}\equiv\mathbf{b} \pmod W \\ \mathcal{F}(\mathbf{t})\equiv\mathbf{v} \pmod{DW}}} \mathbf{1}_{D|t_1\dots t_d} \sum_{\substack{\mathbf{x}\equiv\mathbf{t} \pmod{DW} \\ \mathcal{F}(\mathbf{x})=\mathbf{v}}} \mathbf{1}_{[0,N]^d}(\mathbf{x}) \quad (2.4.27)$$

□

Since $D \leq R^2 \leq N^{\frac{\eta}{1+\eta}}$, apply Theorem 2.3.4,we can write (2.4.27) as

$$\sum_{\substack{\mathbf{t}\in\mathbb{Z}_{DW}^d, \mathbf{t}\equiv\mathbf{b} \pmod W \\ \mathcal{F}(\mathbf{t})\equiv\mathbf{v} \pmod{DW}}} \mathbf{1}_{D|t_1\ldots t_d}\left(N^{d-kr}(DW)^{-n}J(N^{-k}\mathbf{v})\prod_p \sigma_p(DW,\mathbf{t},\mathbf{v}) + O(N^{d-kr-\delta'}D^{-d})\right) \tag{2.4.28}$$

First we estimate the contribution of error term in (2.4.28) to the sum $S_{W,\mathbf{b}}(N)$. By the standard number of divisors estimate, we have that the number of pairs d_1, d_2 such that $[d_1, d_2] = D$ is $\lesssim_\tau D^\tau$ for any $\tau > 0$. Also since χ is supported on $x \leq 1$, the sum in D is restricted to $D \leq R^2$. The contribution of error term is given by

$$\lesssim_\tau N^{d-kr-\delta'}W^d \sum_{D\leq R^2} D^\tau \lesssim N^{d-kr-\delta'}W^d R^2 N^\tau \lesssim N^{d-kr-\delta'/2} \tag{2.4.29}$$

for a sufficiently small τ. (here we may think of W as a fixed large constant.)

Now to calculate the main term using the Chinese Remainder Theorem. For each $\mathbf{t} \in \mathbb{Z}_{DW}^d$ satisfying $\mathbf{t} \equiv \mathbf{b} \pmod W$, there is a unique $\mathbf{s} \in \mathbb{Z}_D^d$ such that $\mathbf{t} \equiv \mathbf{s} \pmod D$. Hence suppose $\mathcal{F}(\mathbf{b}) \equiv \mathbf{v} \pmod D$ then $\mathcal{F}(\mathbf{t}) \equiv \mathbf{v} \pmod{DW}$ is equivalent to $\mathcal{F}(\mathbf{s}) \equiv \mathbf{v} \pmod W$. Hence, applying Lemma 2.4.4, we have

$$\begin{aligned}
&\sum_{\substack{\mathbf{t}\in\mathbb{Z}_{DW}^d, \mathbf{t}\equiv\mathbf{b} \pmod W \\ \mathcal{F}(\mathbf{t})\equiv\mathbf{v} \pmod{DW}}} \mathbf{1}_{D|t_1\ldots t_d}\prod_{p|W}\sigma_p(p,\mathbf{b},\mathbf{v})\prod_{p|D}\sigma_p(p,\mathbf{s},\mathbf{v})\prod_{p\nmid DW}\sigma_p(\mathbf{v}) \\
&= \prod_{p|W}\sigma_p(p,\mathbf{b},\mathbf{v})\prod_{p\nmid W}\sigma_p(\mathbf{v}) \sum_{\substack{\mathbf{s}\in\mathbb{Z}_D^d \\ \mathcal{F}(\mathbf{s})\equiv\mathbf{v} \pmod D}} \mathbf{1}_{D|s_1\ldots s_d}\prod_{p|D}\frac{\sigma_p(p,\mathbf{s},\mathbf{v})}{\sigma_p(\mathbf{v})}
\end{aligned} \tag{2.4.30}$$

Hence (2.4.19) follows from (2.4.26)-(2.4.30).

Now to show (2.4.20), we do the same calculation with D is replaced by $Dq = [D, q]$ in (2.4.26). Hence (2.4.27)-(2.4.30) remain valid with D replaced by $[D, q]$. Now we have to calculate (asymptotically) the sum

$$S_W(f,\gamma) := \sum_{D:(D,W)=1}{}' \gamma_D(\mathbf{v})h_D(R) \tag{2.4.31}$$

This could be done by sieve methods, as in [42], [106]. We will follow the approach in [106] and then adapt it to give the asymptotic for the sum

$$S_{W,q}(f,\gamma) := \sum_{D:(D,W)=1}{}' \gamma_{[D,q]}(\mathbf{v})h_D(R) \tag{2.4.32}$$

which will be needed in the concentration estimate (2.4.5).

Lemma 2.4.10 ([106], Proposition 10)**.** *Let $\gamma_D(\mathbf{v})$ be a multiplicative function (in D) satisfying the estimate (2.4.7). Let $\chi(x) = f(x) = (1-x)_+^{10d}$ then*

$$S_W(f,\gamma) = (\frac{\phi(W)}{W}\log R)^{-d}\int_0^\infty f^{(d)}(x)^2\frac{x^{d-1}}{(d-1)!}dx + o_{\omega\to\infty}(1) \tag{2.4.33}$$

Furthermore, for a prime $q > \omega$,

$$S_{W,q}(f,\gamma) = \frac{d}{q}(\frac{\phi(W)}{W}\log R)^{-d}\int_0^\infty (f^{(d)}(x) - f^{(d)}(x+\frac{\log q}{\log R}))^2\frac{x^{d-1}}{(d-1)!}dx + o_{\omega\to\infty}(1) \tag{2.4.34}$$

Proof. Since $f(x) = (1-x)_+^{10d}$ hence $e^x f(x)$ is compactly supported and $10d-1$ continuously differentiable. Denoted $\hat{f}(t)$ the Fourier transform of $e^x f(x)$. Hence

$$|\hat{f}(t)| \lesssim (1+|t|)^{-10d}$$

Recall the formula

$$d^{-\frac{1}{\log R}}f(\frac{\log d}{\log R}) = \int_{\mathbb{R}} d^{\frac{-it}{\log R}}\hat{f}(t)dt.$$

Substitute this into (2.4.25), swapping the sum and integral due to rapid decay of $\hat{f}_1$,$\hat{f}_2$, one has

$$h_D(R) = \int_{\mathbb{R}}\int_{\mathbb{R}}\sum_{\substack{d_1,d_2\\ [d_1,d_2]=D}}\mu(d_1)\mu(d_2)d_1^{-\frac{1+it_1}{\log R}}d_2^{-\frac{1+it_2}{\log R}}\hat{f}(t_1)\hat{f}(t_2)dt_1dt_2 = \int_{\mathbb{R}}\int_{\mathbb{R}} g_D(t_1,t_2)\hat{f}(t_1)\hat{f}(t_2)dt_1dt_2. \tag{2.4.35}$$

The function $g_D(t_2,t_2)\gamma_D(\mathbf{v})$ is multiplicative in D and by rapid decay of $g_D(t_1,t_2)$ in t_1,t_2, one has

$$\sum_{D:(D,W)=1}{}' g_D(t_1,t_2)\gamma_D(\mathbf{v}) = \prod_{p>\omega}\left(1+g_p(t_1,t_2)\gamma_p(\mathbf{v})\right).$$

Substitute this into (2.4.31) gives

$$S_W(f,\gamma) = \int_{\mathbb{R}}\int_{\mathbb{R}}\prod_{p>\omega}\left(1-\frac{\gamma_p(\mathbf{v})}{p^{\frac{1+it_1}{\log R}}}-\frac{\gamma_p(\mathbf{v})}{p^{\frac{1+it_2}{\log R}}}+\frac{\gamma_p(\mathbf{v})}{p^{\frac{2+it_1+it_2}{\log R}}}\right)\hat{f}(t_1)\hat{f}(t_2)dt_1dt_2 \tag{2.4.36}$$

Using the asymptote of $\gamma_p(\mathbf{v})$ (2.4.7), one has the following estimate via Taylor's series of $\log(1+\epsilon)$,

$$\log\left|1-\frac{\gamma_p(\mathbf{v})}{p^{\frac{1+it_1}{\log R}}}-\frac{\gamma_p(\mathbf{v})}{p^{\frac{1+it_2}{\log R}}}+\frac{\gamma_p(\mathbf{v})}{p^{\frac{2+it_1+it_2}{\log R}}}\right| \le 3dp^{-1-\frac{1}{\log R}} + O(p^{-2})$$

Now by the well-known asymptotic[15]

$$\sum_p p^{-1-\frac{1}{\log R}} = \log\log R + O(1)$$

Hence the integrand in (2.4.36) is bounded by

$$C(\log R)^{3d}(1+|t_1|)^{-10d}(1+|t_2|)^{-10d}$$

Integrating over $|t_2| > \sqrt{\log R}$,

$$\int_{|t_2|>\sqrt{\log R}} \int_{\mathbb{R}} (\log R)^{3d}(1+|t_1|)^{-10d}(1+|t_2|)^{-10d} dt_2 = O(\log^{-d} R \int_{|t_2|>\sqrt{\log R}} (1+|t_2|)^{-2d} dt_2) = O(\log^{-d} R)$$

The same holds for $|t_1| > \sqrt{\log R}$. Hence

$$S_W(f,\gamma) = \int_{|t_1|\le\sqrt{\log R}} \int_{|t_2|\le\sqrt{\log R}} \prod_{p>\omega} \left(1 - \frac{\gamma_p(\mathbf{v})}{p^{\frac{1+it_1}{\log R}}} - \frac{\gamma_p(\mathbf{v})}{p^{\frac{1+it_2}{\log R}}} + \frac{\gamma_p(\mathbf{v})}{p^{\frac{2+it_1+it_2}{\log R}}}\right) \hat{f}(t_1)\hat{f}(t_2) dt_1 dt_2 + O(\log^{-d} R) \tag{2.4.37}$$

□

For $\Re s > 1$, define

$$\zeta_W(s) := \prod_{p>\omega} (1-p^{-s})^{-1} = \zeta(s) \prod_{p\le\omega} (1-p^{-s})$$

Apply (2.4.7) (asymptote for $\gamma_p(\mathbf{v})$), we have that

$$\begin{aligned}
\prod_{p>\omega} \left(1 - \frac{\gamma_p(\mathbf{v})}{p^{\frac{1+it_1}{\log R}}} - \frac{\gamma_p(\mathbf{v})}{p^{\frac{1+it_2}{\log R}}} + \frac{\gamma_p(\mathbf{v})}{p^{\frac{2+it_1+it_2}{\log R}}}\right) &= \prod_{p>\omega} \left(1 - \frac{d}{p^{1+s_1}} - \frac{d}{p^{1+s_2}} + \frac{d}{p^{2+s_1+s_2}} + O(p^{-2})\right) \\
&= \prod_{p>\omega} \left(1 - \frac{d}{p^{1+s_1}} - \frac{d}{p^{1+s_2}} + \frac{d}{p^{2+s_1+s_2}}\right)(1+O(p^{-2})) \\
&= \prod_{p>\omega} \left(1 - \frac{d}{p^{1+s_1}} - \frac{d}{p^{1+s_2}} + \frac{d}{p^{2+s_1+s_2}}\right) \prod_{p>\omega}(1+O(p^{-2})) \\
&= \frac{\zeta_W(1+s_1+s_2)^d}{\zeta_W(1+s_1)^d \zeta_W(1+s_2)^d}(1+o_{\omega\to\infty}(1)) \qquad (2.4.38)
\end{aligned}$$

where $s_1 = 1 + \frac{1+it_1}{\log R}$, $s_2 = 1 + \frac{1+it_2}{\log R}$. On the range $|t_1|, |t_2| \le \sqrt{\log R}$, we have $s = 1 + O(\frac{1}{\sqrt{\log R}})$.

[15] This can be seen by taking log of the following equation obtained from the simple pole with residue 1 at 1 of the Riemann's Zeta Function.

$$\prod_p \left(1 - \frac{1}{p^{1+\frac{1}{\log R}}}\right) = \frac{1}{\zeta(1+\frac{1}{\log R})} = \frac{1}{\log R + O(1)}$$

For each fixed ω, letting N and hence R goes to infinity, we have

$$\prod_{p\le\omega}(1-p^{-s})=\prod_{p\le\omega}(1-p^{-1})+o(1)=\frac{\phi(W)}{W}+o(1)$$

Hence using that $\zeta(s)=(s-1)^{-1}+O(1)$

$$\begin{aligned}\zeta_W(s)=\prod_{p\le\omega}(1-p^{-s})\zeta(s)&=\big(\frac{\phi(W)}{W}+o(1)\big)\big(\frac{1}{s-1}+O(1)\big)\\&=\frac{1}{s-1}\frac{\phi(W)}{W}+o(\frac{1}{s-1})+O(\frac{\phi(W)}{W})+o(1)\\&=\frac{1}{s-1}\frac{\phi(W)}{W}(1+o_{w\to\infty}(1))\end{aligned}$$

Substitute this into (2.4.38) and (2.4.37) gives

$$\begin{aligned}S_W(f,\gamma)&=(\frac{\phi(W)}{W}\log R)^{-d}\int_{|t_1|,|t_2|\le\sqrt{\log R}}\frac{(1+it_1)^d(1+it_2)^d}{(2+it_1+it_2)^d}(1+o_{\omega\to\infty}(1))\hat{f}(t_1)\hat{f}(t_2)dt_1dt_2+O(\log^{-d}R)\\&=(\frac{\phi(W)}{W}\log R)^{-d}\int_{|t_1|,|t_2|\le\sqrt{\log R}}\frac{(1+it_1)^d(1+it_2)^d}{(2+it_1+it_2)^d}\hat{f}(t_1)\hat{f}(t_2)dt_1dt_2(1+o_{\omega\to\infty}(1))\\&=(\frac{\phi(W)}{W}\log R)^{-d}\int_{\mathbb{R}}\int_{\mathbb{R}}\frac{(1+it_1)^d(1+it_2)^d}{(2+it_1+it_2)^d}\hat{f}(t_1)\hat{f}(t_2)dt_1dt_2(1+o_{\omega\to\infty}(1))\\&=(\frac{\phi(W)}{W}\log R)^{-d}\int_{\mathbb{R}}\int_{\mathbb{R}}\frac{(1+it_1)^d(1+it_2)^d}{(2+it_1+it_2)^d}\hat{f}(t_1)\hat{f}(t_2)dt_1dt_2+o_{\omega\to\infty}(1)\end{aligned}\tag{2.4.39}$$

Here in the last line we use that $\hat{f}$ is rapidly decays and extending the integral to $\mathbb{R}$ causes an error term $o(1)$. Now recall the value of the Gamma function at positive integers k:

$$(k-1)!=\Gamma(k)=\int_0^\infty e^{-x}x^{k-1}dx$$

Since Γ is analytic on $\{z:\Re z\ge 0\}$ and $e^{-z}z^{k-1}$ decays for large $\Re z$. Hence we shift the contour from $\mathbb{R}$ to $(s+it)\mathbb{R}$ with $s>0$, which are lines in the first quadrant with starting point at origin. That is we have that for $y=(s+it)x$,

$$(k-1)!=\int_0^\infty e^{-y}y^{k-1}dy$$

That is

$$(s+it)^{-k}=\int_0^\infty e^{-x(s+it)}\frac{x^{k-1}}{(k-1)!}dx\tag{2.4.40}$$

In our case, we have

$$(2+it_1+it_2)^{-d} = \int_0^\infty e^{-x(2+it_1+it_2)}\frac{x^{d-1}}{(d-1)!}dx \tag{2.4.41}$$

Also recall the Fourier inversion formula

$$f(x) = \int_{\mathbb{R}} e^{-(1+it)x}\hat{f}(t)dt$$

Differentiate n times, one get

$$f^{(d)}(x) = (-1)^d \int_{\mathbb{R}} e^{-x(1+it)}(1+it)^d \hat{f}(t)dt \tag{2.4.42}$$

Hence one could write (2.4.39) as

$$S_W(f,\gamma) = (\frac{\phi(W)}{W}\log R)^{-d}\int_0^\infty f^{(d)}(x)^2\frac{x^{d-1}}{(d-1)!}dx + o_{\omega\to\infty}(1)$$

This shows (2.4.33).

Now we modify the above arguments to show (2.4.34). Fix a prime $q>\omega$, we have

$$S_{W,q}(f,\gamma) := \int_{\mathbb{R}}\int_{\mathbb{R}} \sideset{}{'}\sum_{D:(D,W)=1} \gamma_{[D,q]}(\mathbf{v})g_D(t_1,t_2)dt_1dt_2 \tag{2.4.43}$$

Now we separate the inner sum in D into cases $q\nmid D$ and $q|D$,

$$\begin{aligned}\sideset{}{'}\sum_{D:(D,W)=1} \gamma_{[D,q]}(\mathbf{v})g_D(t_1,t_2) &= \gamma_q(\mathbf{v})(1+g_1(t_1,t_2))\sideset{}{'}\sum_{\substack{D,q\nmid D,\\(D,W)=1}} g_D(t_1,t_2)\gamma_D(\mathbf{v})\\ &= \gamma_q(\mathbf{v})(1+g_q(t_1,t_2))\prod_{p>\omega,p\neq q}\left(1-\frac{\gamma_p(\mathbf{v})}{p^{\frac{1+it_1}{\log R}}}-\frac{\gamma_p(\mathbf{v})}{p^{\frac{1+it_2}{\log R}}}+\frac{\gamma_p(\mathbf{v})}{p^{\frac{2+it_1+it_2}{\log R}}}\right)\\ &= \frac{\gamma_q(\mathbf{v})(1+g_q(t_1,t_2))}{1+g_q(t_1,t_2)\gamma_q(\mathbf{v})}\prod_{p>\omega}\left(1-\frac{\gamma_p(\mathbf{v})}{p^{\frac{1+it_1}{\log R}}}-\frac{\gamma_p(\mathbf{v})}{p^{\frac{1+it_2}{\log R}}}+\frac{\gamma_p(\mathbf{v})}{p^{\frac{2+it_1+it_2}{\log R}}}\right)\end{aligned}$$

Hence this differs from the previous case in the sense that we have the additional factor $\frac{\gamma_q(\mathbf{v})(1+g_q(t_1,t_2))}{1+g_q(t_1,t_2)\gamma_q(\mathbf{v})}$. Since we assume $q>\omega$ we have

$$\gamma_q(\mathbf{v}) = \frac{d}{q}(1+o(1))$$

Using this estimate, we have

$$\frac{\gamma_q(\mathbf{v})(1+g_q(t_1,t_2))}{1+g_q(t_1,t_2)\gamma_q(\mathbf{v})} = \frac{d}{q}(1-q^{-\frac{1+it_1}{\log R}})(1-q^{-\frac{1+it_2}{\log R}})(1+o(1))$$

Hence we have analogue of (2.4.39),

$$\begin{aligned} S_{W,q}(f,\gamma) =& (\frac{\phi(W)}{W}\log R)^{-d}\int_{\mathbb{R}}\int_{\mathbb{R}}\frac{(1+it_1)^d(1+it_2)^d}{(2+it_1+it_2)^d} \\ &\times (1-e^{-\frac{(1+it_1)\log q}{\log R}})(1-e^{-\frac{(1+it_2)\log q}{\log R}})\hat{f}(t_1)\hat{f}(t_2)dt_1dt_2 + o_{\omega\to\infty}(1) \end{aligned} \tag{2.4.44}$$

Applying (2.4.41), we write (2.4.39) as

$$\int_0^\infty \left(\int_{\mathbb{R}}(e^{-x(1+it)} - e^{-(x+\frac{\log q}{\log R})(1+it)})(1+it)^d\hat{f}(t)dt\right)^2 \frac{x^{d-1}}{(d-1)!}dx$$

Applying (2.4.42), this becomes

$$\int_0^\infty (f^{(d)}(x) - f^{(d)}(x+\frac{\log q}{\log R}))^2\frac{x^{d-1}}{(d-1)!}dx + o_{\omega\to\infty}(1)$$

as required.

2.4.4 Proof of the Main Theorem

Proof of Theorem 2.4.2. Let $\eta \leq \frac{\eta(r,k)}{2(1+\eta(r,k))}$ where $\eta(r,k)$ is as in the assumption of Theorem 2.4.2. Then by (2.4.21) and (2.4.30), one has

$$\begin{aligned} &\sum_{\substack{\mathbf{x}\in[N]^d \\ (\mathbf{x},W)=1,\ \mathcal{F}(\mathbf{x})=\mathbf{v}}} \Lambda_R^2(x_1x_2\cdots x_d) = \sum_{\substack{\mathbf{b}\in\mathbb{Z}_W^d \\ (\mathbf{b},W)=1}} S_{N,W,\mathbf{b}}(\mathbf{v}) \\ &= c_d(f)N^{d-kr}J(N^{-k}\mathbf{v})(\log R)^{-d}\phi(W)^{-d}(1+o_{\omega\to\infty}(1))\sum_{\substack{\mathbf{b}\in\mathbb{Z}_W^d \\ (\mathbf{b},W)=1}} \mathfrak{G}_{W,\mathbf{b}}(\mathbf{v}) + O(N^{d-kr-\delta'}) \end{aligned} \tag{2.4.45}$$

By Lemma 2.4.4 and the Chinese Remainder Theorem,

$$\phi(W)^{-d}\sum_{\substack{\mathbf{b}\in\mathbb{Z}_W^d \\ (\mathbf{b},W)=1}} \mathfrak{G}_{W,\mathbf{b}}(\mathbf{v}) = \prod_{p|W}(\phi(p)^{-d}\sum_{\substack{\mathbf{b}\in\mathbb{Z}_p^d \\ (\mathbf{b},p)=1}}\sigma_p(p,b,\mathbf{v}))\prod_{p\nmid W}\sigma_p(\mathbf{v}) \tag{2.4.46}$$

Recall that $\sigma_p(\mathbf{v}) = 1+O(p^{-2})$, hence $\prod_{p\nmid W}\sigma_p(\mathbf{v}) = 1+o_{\omega\to\infty}(1)$. For a fixed $l\in\mathbb{N}$ and primes

$p \le \omega$, one has[16]

$$\begin{aligned}
&\phi(p)^{-d}p^{-l(d-r)} \sum_{\substack{\mathbf{b}\in\mathbb{Z}_p^d,(\mathbf{b},p)=1\\ \mathcal{F}(\mathbf{b})\equiv\mathbf{v} \pmod p}} |\{\mathbf{x}\in\mathbb{Z}_{p^l}^d; \mathcal{F}(p\mathbf{x}+\mathbf{b})=\mathbf{v}|\}| \\
&= \phi(p)^{-d}p^{-l(d-r)}p^d|\{\mathbf{y}\in\mathbb{Z}_{p^l}^d; (\mathbf{y},p)=1, \mathcal{F}(\mathbf{y})=\mathbf{v}\}| \\
&= \frac{p^{rd}}{\phi(p^l)^d}\mathcal{M}(p^l;\mathbf{v}) \qquad (2.4.47)
\end{aligned}$$

where $\mathcal{M}(p^l;\mathbf{v})$ is the number of solutions to $\mathcal{F}(\mathbf{y}) \equiv \mathbf{v} \pmod{p^l}$ in the reduced residue class $\mathbf{y}\in\mathbb{Z}_{p^l}^d$, $(\mathbf{y},p)=1$. Taking limit $l\to\infty$, one has

$$\phi(p)^{-d} \sum_{\substack{\mathbf{b}\in\mathbb{Z}_p^d,(\mathbf{b},p)=1\\ \mathcal{F}(\mathbf{b})\equiv\mathbf{v} \pmod p}} \sigma_p(p,\mathbf{b};\mathbf{v}) = \sigma_p^*(\mathbf{v})$$

and by (2.4.46), one has

$$\phi(W)^{-d} \sum_{\substack{(\mathbf{b},W)=1\\ \mathcal{F}(\mathbf{b})\equiv v \pmod W}} \mathfrak{G}_{W,b}(\mathbf{v}) = \prod_{p|W}\sigma_p^*(\mathbf{v})(1+o_{\omega\to\infty}(1)) = \mathfrak{G}^*(\mathbf{v})(1+o_{\omega\to\infty}(1))$$

This proves (2.4.5).

Now we prove (2.4.6). Note that to estimate the sum over $\mathbf{x}\in[N]^d\backslash\mathcal{P}^\epsilon[N]^d$ under the restriction $(\mathbf{x},W)=1$, we only need to sum over $\mathbf{x}=(x_1,\dots,x_d)$ for which $q|x_1\dots x_d$ for which $q|x_1\dots x_d$ for some prime $\omega<q\le N^\epsilon$. Hence

$$\sum_{\substack{\mathbf{x}\in[N]^d\backslash\mathcal{P}^\epsilon(N)^d\\ (\mathbf{x},W)=1,\mathcal{F}(\mathbf{x})=\mathbf{v}}} \Lambda_R^2(x_1\dots x_d) \le \sum_{\omega<q\le N^\epsilon}\sum_{\substack{(\mathbf{x},W)=1\\ \mathcal{F}(\mathbf{x}=\mathbf{v})}} \mathbf{1}_{q|x_1\dots x_d}\Lambda_R^2(x_1\dots x_d)\mathbf{1}_{[0,N]^d}(\mathbf{x}) = \sum_{\omega<q\le N^\epsilon} S_{W,q,\mathbf{b}}(N)$$

Now choose $f(x)=(1-x)_+^{10d}$ then recall $a^n-b^n=(a-b)(a^{n-1}+a^{n-2}b+\dots+b^{n-1})$, we observe directly that for $0\le x,\tau\le 1$, we have

$$|f^{(d)}(x)-f^{(d)}(x+\tau)| \le \tau|f^{(d+1)}(x)| \qquad (2.4.48)$$

Then by estimates (2.4.22), (2.4.34), and (2.4.48),

$$\sum_{\omega<q\le N^\epsilon} S_{W,q,\mathbf{b}}(N)$$

[16] By p^d to one correspondence between $\{p\mathbf{x}+\mathbf{b} : \mathbf{x}\in\mathbb{Z}_{p^l}^d, (\mathbf{b},p)=1\}$ and $\{\mathbf{y}\in\mathbb{Z}_{p^l}^d : (\mathbf{y},p)=1\}$.

$$\leq \sum_{\omega<q\leq N^\epsilon} \frac{d}{q}(\frac{\phi(W)}{W}\log R)^{-d}\int_0^\infty (f^{(d)}(x)-f^{(d)}(x+\frac{\log q}{\log R}))^2\frac{x^{d-1}}{(d-1)!}dx + o_{\omega\to\infty}(1)+O(N^{d-kr-\delta})$$
$$\leq \frac{d}{q}\left(\frac{\log q}{\log R}\right)^2 c_{d+1}(f)N^{d-kr}(\log R)^{-d}\mathfrak{G}^*(N,\mathbf{v})(1+o_{\omega\to\infty}(1))+O(N^{d-kr-\delta})$$

Write $\epsilon'=\epsilon/\eta$ so $N^\epsilon=R^{\epsilon'}$ where $R=N^\eta$. Then using dyadic decomposition and the Prime Numner Theorem, one can bound the sum over primes $q, \omega<q\leq R^{\epsilon'}$,

$$\sum_{\omega<q\leq R^{\epsilon'}} q^{-1}(\log q)^2 = \sum_{\omega\leq 2^j<R^{\epsilon'}}\sum_{2^{j-1}<q\leq 2^j} q^{-1}(\log q)^2$$
$$\lesssim \sum_{\omega\leq 2^j<R^{\epsilon'}}(\frac{2^j}{j}+o_{\omega\to\infty}(1))\frac{j^2}{2^{j-1}} \quad \leq (2+o_{\omega\to\infty}(1))\sum_{j\leq\epsilon'\frac{\log R}{\log 2}} j\leq 2(\epsilon')^2$$

as required. We will choose $\epsilon>0$ to ensure that (2.4.5) dominates (2.4.6). for that we need to compare $c'_{d+1}(f)$ with $c_d(f)$ defined in the Theorem. Here $f(x)=(1-x)_+^{10d}$ so $f^{(d)}(x)=\alpha_d(1-x)_+^{9d}$ and $f^{(d+1)}(x)=9d\alpha_d(1-x)_+^{9d-1}$ with $\alpha_d=(10d)!/(9d)!$. By the beta function identity,

$$\int_0^1(1-x)^a x^b dx=\frac{a!b!}{(a+b+1)!}$$

we have

$$c'_{d+1}(f)<16d^2c_d(f)$$

Hence if

$$32d^3(\epsilon/\eta)^2\leq\frac{1}{2} \tag{2.4.49}$$

then for sufficiently large N,ω.

$$\sum_{\substack{\mathbf{x}\in\mathcal{P}^\epsilon(N)\\ \mathcal{F}(\mathbf{x})=\mathbf{v}}}\Lambda_R^2(x_1\dots x_d)\mathbf{1}_{[0,N]^d}(\mathbf{x})\geq c_dN^{d-kr}(\log R)^{-d}\mathfrak{G}^*(N,\mathbf{v}) \tag{2.4.50}$$

for some positive constant $c_d=c_d(f)>0$.

Finally if $\mathbf{x}\in\mathcal{P}^\epsilon[N]^d$ then each coordinate x_i could have at most $\frac{1}{\epsilon}$ prime factors. Hence

$$\Lambda_R(x_1\dots x_d)\leq \text{the number of squarefree divisors of } x_1\dots x_d\leq 2^{d/\epsilon}.$$

Thus by (2.4.50), the numbers of solutions to $\mathcal{F}(\mathbf{x})=\mathbf{v}$ with $\mathbf{x}\in\mathcal{P}^\epsilon[N]^d$ satisfies

$$\mathcal{M}^\epsilon_{\mathcal{F}}(N)\geq c(d,k,r)N^{d-kr}(\log N)^{-d}\mathfrak{G}^*(N,\mathbf{v})$$

where $c(d,k,r):=c_d2^{-2d/\epsilon}$ for some $\epsilon=\epsilon(d,k,r)>0$. In fact we may choose $\epsilon:=(4d)^{-3/2}\eta(r,k)$

to satisfy (2.4.49) with $\eta(r,k) = (8r^2(r+1)(r+2)k(k+1))^{-1}$ (chosen from conditions in Theorem 2.3.4). This proves Theorem 2.4.2.

2.5 Concluding Remarks

Some ideas from additive combinatorics are also used along this line where solutions are restricted to some special sets like primes or almost primes. For example, Bourgain-Gamburd-Sarnak's result [13] on almost primes uses the idea of affine sieves. Their results are different from us that they give better bound on ranks but only applicable to some classes of equations with high degrees of symmetry. This is non generic. Cook-Magyar proves analogue result [19] on prime solutions to diophantine system whose rank is tower-exponential with respect to its degree.

The first application of inverse Gowers norm theorem to number theory is to study the asymptotic of the number of prime solutions of systems of linear equations of finite complexity by Green-Tao [52]. Basically what is shown in [52] is that $\|\tilde{\Lambda} - 1\|_{U^k}$ is small where $\tilde{\Lambda}$ is W-tricked Mangoldt function. To show this, one would need an explicit form of all structures that make the U^k norm large; to apply the results of [53] [54]. We obtain a decomposition of Mangoldt function: $\tilde{\Lambda} = \Lambda^\sharp + \Lambda^\flat$ Here $\Lambda^\sharp = \sum_{d|n, d\leq R} \mu(d)\log(n/d)$ with R a small power of N which will contribute to the main term (major arc). Any term involving $\Lambda^\flat$(sum over $d > R$) will be a small error term (minor arc) that does not correlate with U^k-obstructions. This follows from smallness of $\|\mu\|_{U^s}$. Recently, so-called *nilpotent circle method* where nilsequences could play a role of the linear phase, is also used to find the asymptotic of average of $f(L_1(u,v))\dots f(L_k(u,v))$ for some classes of arithmetic functions f and L_j are binary linear forms. Some arithmetic functions that are orthogonal to polynomial nilsequences are studied in [76]. In [64] Frantzikinakis and Host prove a structure theorem for bounded multiplicative functions using higher order Fourier analysis with some new applications in number theory and Ramsey theory.

Chapter 3

Corners in Dense Subsets of Primes via a Transference Principle

Recall that a set $A \subseteq \mathcal{P}^d$ has upper relative density α if

$$\limsup_{N\to\infty} \frac{|A \cap \mathcal{P}_N^d|}{|\mathcal{P}_N^d|} = \alpha$$

Let us state our main result in this section.

Theorem 3.0.1. *Let $A \subset (\mathcal{P}_N)^d$ of positive relative upper density $\alpha > 0$. Then A contains at least $C(\alpha)\frac{N^{d+1}}{(\log N)^{2d}}$ corners for some (computable) constant $C(\alpha)$.*

Indeed this is not the most general known results and there are also other modern approaches to this problem known as the densification trick [16]. We demonstrate the application the original approach of Green-Tao [51] to attack this problem. This is the first result in the direction of extending the theorem of Green and Tao to the multidimensional setting. The key ingredient is to move our set up to translate the problem to the setting of a hypergraph system and to prove an appropriate version of the so-called "correlation conditions" of Green and Tao.

In higher dimensions, the direct product of primes $\mathcal{P}^d$ is not a random subset of $\mathbb{Z}^d$ and one reason is the correlation from direct product structure. For example, if we want to count corners $\{(a, b), (a + d, b), (a, b + d)\}$ in $\mathcal{P}^2$, suppose $(a + d, b), (a, b + d) \in \mathcal{P}^2$ then the remaining vertex (a, b) must also be in $\mathcal{P}^2$. Thus the probability that all three vertices are in $\mathcal{P}^2$ (or in the direct product of the almost primes) is not $(\log N)^{-6}$ as one would expect, but roughly $(\log N)^{-4}$. Due to this correlation, the obvious generalization of ν, the $d-$folds tensor product $\nu \otimes \nu \cdots \otimes \nu(x_1, \dots, x_d) = \nu(x_1)\nu(x_2)\dots\nu(x_d)$, could not behaves pseudorandomly on its support $\mathcal{A}^d$ where $\mathcal{A}$ is the support of ν. For example in the corner $\mathcal{P}^2$, if we calculate

$$\mathbb{E}_{a,b,d}(\nu\otimes\nu)(a,b)(\nu\otimes\nu)(a+d,b)(\nu\otimes\nu)(a,b+d)=\mathbb{E}_{a,b,d}\nu(a)^2\nu(b)^2\nu(a+d)\nu(b+d),$$

we have to deal with higher moments of ν where we don't have control. This happens exactly when there is a correlation, that is there are points P_1, P_2 with a projection π_i to a coordinate axis such that $\pi_i(P_1) = \pi_i(P_2)$. Such a correlation does not happen in the case of Gaussian prime ; $a + ib, c + id$ being Gaussian primes do not imply that $a + id$ is Gaussian prime (but there is a milder correlation to its conjugate).

Our approach is transfer our problem to a corresponding problem in hypergraph to get rid of *strong* correlation from direct product structure. This approach partly used already in [102], where one reduces the problem to that of proving a hypergraph removal lemma for weighted uniform hypergraphs. Then we use an appropriate form of the so-called transference principle [37], [87] to remove the weights and apply the removal lemmas for "un-weighted" hypergraphs, obtained in [36], [82], [104]. An interesting feature is that in our situation the so-called dual function estimates [50] are naturally handled only by the linear forms conditions.

In our weighted setting, this method allows us to distribute the weights such that we can avoid dealing with higher moments of the Green-Tao measure ν. We will define the notion of independent (pseudorandom) weight systems on hypergraph which will be used to count prime configurations. The reason that we cannot handle more general constellations is that we don't quite have a suitable regularity or removal lemma for general weight systems on non-uniform hypergraphs which allow us to do transference arguments. We will apply different methods to overcome this difficulty in the next chapter.

3.1 Hypergraph Setting and Weighted Hypergraph System.

First let us parameterize any affine copies of a corner as follow.

Definition 3.1.1. *A non-degenerate corner is given by the following set of $d-$tuples of size $d+1$ in $\mathbb{Z}^d$ (or $\mathbb{Z}_N^d$):*

$$\{(x_1,...,x_d),(x_1+s,x_2,...,x_d),...,(x_1,...,x_{d-1},x_d+s), s\neq 0\}$$

or equivalently,

$$\{(x_1,...,x_d),(z-\sum_{\substack{1\le j\le d\\ j\neq 1}}x_j,x_2,...,x_d),(x_1,z-\sum_{\substack{1\le j\le d\\ j\neq 2}}x_j,x_3,...,x_d),...,(x_1,...,x_{d-1},z-\sum_{\substack{1\le j\le d\\ j\neq d}}x_j)\}$$

with $z\neq\sum_{1\le i\le d}x_i$

Now to a given set $A \subseteq \mathbb{Z}_N^d$, we assign a $(d+1)-$ partite hypergraph $\mathcal{G}_A$ as follows:

Let $X_1 = ... = X_{d+1} := \mathbb{Z}_N$ be the vertex sets, and for $1 \leq j \leq d$. Let an element $a \in X_j$ represent the hyperplane $x_j = a$, and an element $a \in X_{d+1}$ represent the hyperplane $a = x_1 + .. + x_d$. We join these d vertices (which represent d hyperplanes) if all of these d hyperplanes intersect in a single point in A. Then a simplex in $\mathcal{G}_A$ corresponds to a corner in A. Note that this includes trivial corners which consist of a single point where they are negligible in order of magnitude.

For each $I \subseteq [d+1]$ let $E(I)$ denote the set of hyperedges whose elements are exactly from vertices set $V_i,\ i \in I$. In order to count corners in A, we will place some weights on some of these hyperedges that will represent the coordinates of the corner. To be more precise we define the weights on $1-$edges:

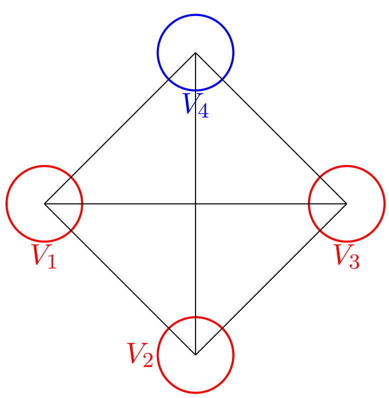
$$\nu_j(a) = \nu(a), a \in X_j, j \leq d, \quad \nu_{d+1}(a) = 1,\ a \in X_{d+1},$$

and on $d-$hyperedges:

$$\nu_I(a) = \nu(a_{d+1} - \sum_{j \in I \setminus \{d+1\}} a_j),\ \ a \in E(I),\ |I| = d,\ d+1 \in I$$

$$\nu_{[1,d]}(a) = 1,\ a \in E([1,d])$$

In particular the weights are 1 or of the form $\nu_I(L_I(\mathbf{x}_I))$ where all linear forms $\{L_I(\mathbf{x}_I)\}$ are pairwise linearly independent. This is an example of what we will call *independent weight system*. The parametrization of corner is indeed a special case of general parametrization (4.2.11) in next chapter but let us just describe this explicitly here. Important special features here is that all linear forms either depends on 1 or d variables.



weight for corner in $\mathbb{Z}^3$ on $4-$partite $3-$regular hypergraph:
$\nu(x_4 - x_1 - x_2)$ on (x_1, x_2, x_4)
$\nu(x_4 - x_2 - x_3)$ on (x_2, x_3, x_4)
$\nu(x_4 - x_1 - x_3)$ on (x_1, x_3, x_4)
$\nu(x_1), \nu(x_2), \nu(x_3)$ on x_1, x_2, x_3 respectively

Measure space $(V_e, \mu_e(x_e))$.
$V_e = \prod_{j \in e} V_j, x_e = (x_i)_{i \in e}$

Figure 3.1: Weighted hypergraph system. In general, on each edge e of a hypergraph, we attach the weight$\prod \nu(L_e)$ on e where the product is taking over all linear forms depending on exactly x_e. In our case, these linear forms will be pairwise linearly independent.

Definition 3.1.2 (Independent weight system)**.** *An independent weight system is a family of weights* *on the edges of a $d+1-$partite hypergraph such that for any $I \subseteq [d+1], |I| \leq d$, $\nu_I(\mathbf{x}_I)$ is either 1 or of the form $\prod_{j=1}^{K(I)} \nu(L_I^j(\mathbf{x}_I))$ where all distinct linear forms $\{L_I^j\}_{1\leq j\leq K(I)}$ are pairwise linearly independent, moreover the form L_I^j depends exactly on the variables $\mathbf{x}_I = (x_j)_{j\in I}$.*

In fact for a weight system that arised from parametrizing affine copies of configurations in $\mathbb{Z}^d$, it is easy to see from the construction that for any $I \subseteq [d+1], |I| = d$ all distinct linear forms $\{L_J^k\}_{J\subseteq I, 1\leq k\leq K(J)}$ are pairwise linearly independent.
Now for each $I = [d+1]\backslash\{j\}, 1 \leq j \leq d$ let

$$f_I = \mathbf{1}_A(x_1, ..., x_{j-1}, x_{d+1} - \sum_{\substack{1\leq i\leq d \\ i\neq j}} x_i, x_{j+1}, ..., x_d) \cdot \nu_I$$

and for $I = [d]$, let $f_I = \mathbf{1}_A(x_1, ..., x_d)$. As the coordinates of a corner contained in $\mathcal{P}^d$ are given by $2d$ prime numbers. Recall $\nu(p) \approx \log N$ if p is a prime in $[\varepsilon_1 N, \varepsilon_2 N]$ (in residue class $b \pmod W$). We define a multi-linear form

$$\begin{aligned}\Lambda := \Lambda_{d+1}(f_I, |I| = d) := \mathbb{E}_{\mathbf{x}_{[d+1]}} \prod_{|I|=d} f_I \prod_{i=1}^{d} \nu(x_i) &= N^{-d-1} \sum_{\substack{p_i\in A, 1\leq i\leq 2d \\ (p_i)_{1\leq i\leq 2d} \text{ constitutes a corner}}} \prod_{i=1}^{2d} \nu(p_i) \\ &\approx \frac{\log^{2d} N}{N^{d+1}} |\text{number of corners in } A|\end{aligned}$$

Hence Λ can be used to estimate the numbers of corners. Indeed if $\Lambda \geq C_1$ then

$$\text{number of corners in } A \geq C_2 \frac{N^{d+1}}{\log^{2d} N}.$$

We define measure spaces associated to our system of measure as follows. For $1 \leq i \leq d$, let $(X_i, \mu_{X_i}) = (\mathbb{Z}_N, \nu)$ where ν is the Green-Tao measure, and let $\mu_{X_{d+1}}$ be the normalized counting measure on $X_{d+1} = \mathbb{Z}_N$. With this notation one may write

$$\Lambda := \Lambda_{d+1}(f_I, |I| = d) = \int_{X_1} \cdots \int_{X_{d+1}} \prod_{|I|=d} f_I \, d\mu_{X_1} \cdots d\mu_{X_{d+1}}.$$

We define a measure on $X_I, I \subseteq [d+1], |I| = d$ associated to our weight system by

$$\int_{X_I} f d\mu_{X_I} := \mathbb{E}_{\mathbf{x}_I} f_I \cdot \prod_{J\subseteq I, |J|<d} \nu_J(\mathbf{x}_J),$$

also on $X_{[d+1]}$ by

$$\int_{X_{[d+1]}} f d\mu_{X_{[d+1]}} := \mathbb{E}_{\mathbf{x}_{[d+1]}} f \cdot \prod_{I\subseteq[d+1], |I|<d} \nu_I(\mathbf{x}_I),$$

and the associated multi-linear form by

$$\Lambda := \Lambda(f_I, |I| = d) := \int_{X_{[d+1]}} \prod_{|I|=d} f_I d\mu_{X_{[d+1]}} \tag{3.1.1}$$

Remark 3.1.3. *For general configurations, we will use same weighted hypergraph to count the prime configurations but will not attach weight to the function and so the measure space is constructed a bit differently. For the corner case in this chapter, we will apply the transference principle technique hence we will attach weights of size* d *to the functions, only weights of size 1 left on the hypergraph. This strategy will not work in general in particular if there is an intermediate weight of size* $\mathrm{d}', 1 < \mathrm{d}' < \mathrm{d}$. *We don't have an appropriate version of hypergraph removal for transference principle. If we attach that weight to the function, then we don't have control on the size of dual functions.*

3.2 Weighted Box Norm and Weighted Generalized von-Neumann's Inequality

In this section we describe the weighted version of Gowers's uniformity norm on $(d+1)-$partite hypergraph (box norms) and the so-called Gowers's inner product associated to the hypergraph $\mathcal{G}_A$ endowed with a weight system $\{\nu_I\}_{I \subseteq [d+1], |I| \leq d}$. We describe the analogue properties of weighted box norm as in unweighted case i.e. Gowers-Cauchy-Schwartz's inequalities and generalized von-Neumann inequalities in this setting. Here we may recall the index notations stated at the starting of Chapter 1.

Definition 3.2.1. *For each* $1 \leq j \leq d$, *let* X_j, Y_j *be finite set (in this thesis, we will take* $X_j = Y_j := \mathbb{Z}_N$*) with a weight system* ν *on* $X_{[d]} \times Y_{[d]}$. *For* $f : X_{[d]} \to \mathbb{R}$, *define*

$$\begin{aligned} \|f\|_{\square_\mu^d}^{2^d} &:= \int_{X_{[d]} \times Y_{[d]}} \prod_{\underline{\omega}_{[d]}} f(P_{\underline{\omega}_{[d]}}(\mathbf{x}_{[d]}, \mathbf{y}_{[d]})) d\mu_{X_{[d]} \times Y_{[d]}} \\ &:= \mathbb{E}_{\mathbf{x}_{[d]}} \mathbb{E}_{\mathbf{y}_{[d]}} \prod_{\underline{\omega}_{[d]}} f(P_{\underline{\omega}_{[d]}}(\mathbf{x}_{[d]}, \mathbf{y}_{[d]})) \times \prod_{|I|<d} \prod_{\underline{\omega}_I} \nu_I(P_{\underline{\omega}_I}(\mathbf{x}_I, \mathbf{y}_I)) \end{aligned}$$

and define the corresponding Gowers's inner product of 2^d *functions,*

$$\begin{aligned} \left\langle f_{\underline{\omega}}, \underline{\omega} \in \{0,1\}^d \right\rangle_{\square_\mu^d} &:= \int_{X_{[d]} \times Y_{[d]}} \prod_{\underline{\omega}_{[d]}} f_{\underline{\omega}_{[d]}}(P_{\underline{\omega}_{[d]}}(\mathbf{x}_{[d]}, \mathbf{y}_{[d]})) d\mu_{X_{[d]} \times Y_{[d]}} \\ &:= \mathbb{E}_{\mathbf{x}_{[d]}} \mathbb{E}_{\mathbf{y}_{[d]}} \prod_{\underline{\omega}_{[d]}} f_{\underline{\omega}_{[d]}}(P_{\underline{\omega}_{[d]}}(\mathbf{x}_{[d]}, \mathbf{y}_{[d]})) \prod_{|I|<d} \prod_{\underline{\omega}_I} \nu_I(P_{\underline{\omega}_I}(\mathbf{x}_I, \mathbf{y}_I)) \end{aligned}$$

So $\left\langle f, \underline{\omega} \in \{0,1\}^d \right\rangle_{\square_\mu^d} = \|f\|_{\square_\mu^d}^{2^d}$.

For each $e \in \mathcal{H}$, we may define have a measure space on V_e with weighted from all edges $f \subseteq e$. We can define the box norm $\|f\|_{\square_{\mu_e}}$ for $f : V_e \to \mathbb{R}$ as well. If e is clear from the context, we may write this as $\|f\|_{\square_\mu^{d'}}$ where $|e| = d'$.

Remark 3.2.2. *To prove weighted Cauchy-Schwartz's inequality (Theorem 3.2.3) or Generalized von Neumann's theorem (Theorem 3.2.5) below. We will apply Cauchy-Schwartz's inequality and linear forms conditions. The way we apply linear forms condition we will only consider the set of variables they depend on, and if they are different, linear forms condition is applicable. This will be how we apply linear forms conditions here.*

Theorem 3.2.3 (Gowers-Cauchy-Schwartz's Inequality)**.**

$$|\left\langle f_{\underline{\omega}}; \underline{\omega} \in \{0,1\}^d \right\rangle_{\square_\mu^d}| \leq \prod_{\underline{\omega}_{[d]}} \|f_{\underline{\omega}}\|_{\square_\mu^d}.$$

Proof. We will use Cauchy-Schwartz's inequality and linear form condition. Write

$$\begin{aligned}\left\langle f_\omega; \omega \in \{0,1\}^d \right\rangle_{\square_\mu^d} = \mathbb{E}_{\mathbf{x}_{[2,d]},\mathbf{y}_{[2,d]}} \Bigg[\Bigg(& \prod_{|I|<d, 1\notin I} \prod_{\underline{\omega}_I} \nu_I(P_{\underline{\omega}_I}(\mathbf{x}_I, \mathbf{y}_I)) \Bigg)^{1/2} \\ & \Bigg(\mathbb{E}_{x_1} \nu(x_1) \prod_{\underline{\omega}_{[2,d]}} f_{\underline{\omega}_{(0,[2,d])}}(x_1, P_{\underline{\omega}_{[2,d]}}(\mathbf{x}_{[2,d]}, \mathbf{y}_{[2,d]})) \prod_{|I|<d-1, 1\notin I} \nu_{\{1\}\cup I}(x_1, P_{\underline{\omega}_I}(\mathbf{x}_I, \mathbf{y}_I)) \Bigg) \\ & \times \Bigg(\prod_{|I|<d, 1\notin I} \prod_{\underline{\omega}_I} \nu_I(P_{\underline{\omega}_I}(\mathbf{x}_I, \mathbf{y}_I)) \Bigg)^{1/2} \\ & \Bigg(\mathbb{E}_{y_1} \nu(y_1) \prod_{\underline{\omega}_{[2,d]}} f_{\underline{\omega}_{(1,[2,d])}}(y_1, P_{\underline{\omega}_{[2,d]}}(\mathbf{x}_{[2,d]}, \mathbf{y}_{[2,d]})) \prod_{|I|<d-1, 1\notin I} \nu_{\{1\}\cup I}(y_1, P_{\underline{\omega}_I}(\mathbf{x}_I, \mathbf{y}_I)) \Bigg) \Bigg]\end{aligned}$$

Applying the Cauchy Schwartz inequality in the $\mathbf{x}_{[2,d]}, \mathbf{y}_{[2,d]}$ variables, one has

$$|\langle f_{\underline{\omega}}; \underline{\omega} \in \{0,1\}^d \rangle_{\square_\mu^d}|^2 \leq A \cdot B$$

here,

$$\begin{aligned}A = \mathbb{E}_{\mathbf{x}_{[2,d]},\mathbf{y}_{[2,d]}} \Bigg[& \prod_{|I|<d, 1\notin I} \prod_{\underline{\omega}_I} \nu_I(P_{\underline{\omega}_I}(\mathbf{x}_I, \mathbf{y}_I)) \\ \times \Bigg(& \mathbb{E}_{x_1, y_1} \nu(x_1)\nu(y_1) \prod_{\underline{\omega}_{[2,d]}} f_{\underline{\omega}_{(0,[2,d])}}(x_1, P_{\underline{\omega}_{[2,d]}}(\mathbf{x}_{[2,d]}, \mathbf{y}_{[2,d]})) f_{\underline{\omega}_{(0,[2,d])}}(y_1, P_{\underline{\omega}_{[2,d]}}(\mathbf{x}_{[2,d]}, \mathbf{y}_{[2,d]})) \\ \times & \prod_{|I|<d-1, 1\notin I} \prod_{\underline{\omega}_I} \nu_{\{1\}\cup I}(x_1, P_{\underline{\omega}_I}(\mathbf{x}_I, \mathbf{y}_I)) \nu_{\{1\}\cup I}(y_1, P_{\underline{\omega}_I}(\mathbf{x}_I, \mathbf{y}_I)) \Bigg) \Bigg] \\ = \Big\langle & f_{\underline{\omega}}^{(0)}(P_{\underline{\omega}}(\mathbf{x}_{[d]}, \mathbf{y}_{[d]})) \Big\rangle_{\square_\mu^d}\end{aligned}$$

where $f^{(0)}_{\tilde{\underline{\omega}}} = f_{(0,\tilde{\underline{\omega}}\cap[2,d])}$ for any $\tilde{\underline{\omega}}_{[1,d]}$. And,

$$
\begin{aligned}
B = \mathbb{E}_{\mathbf{x}_{[2,d]},\mathbf{y}_{[2,d]}}\Bigg[&\prod_{|I|<d, 1\notin I}\prod_{\underline{\omega}_I} \nu_I(P_{\underline{\omega}_I}(\mathbf{x}_I,\mathbf{y}_I)) \\
&\times \Bigg(\mathbb{E}_{x_1,y_1}\nu(x_1)\nu(y_1)\prod_{\underline{\omega}_{[2,d]}} f_{\underline{\omega}_{(1,[2,d])}}(x_1, P_{\underline{\omega}_{[2,d]}}(\mathbf{x}_{[2,d]},\mathbf{y}_{[2,d]}))f_{\underline{\omega}_{(1,[2,d])}}(y_1, P_{\underline{\omega}_{[2,d]}}(\mathbf{x}_{[2,d]},\mathbf{y}_{[2,d]})) \\
&\times \prod_{|I|<d, 1\notin I}\prod_{\underline{\omega}_I} \nu_{\{1\}\cup I}(x_1, P_{\underline{\omega}_I}(\mathbf{x}_I,\mathbf{y}_I))\nu_{\{1\}\cup I}(y_1, P_{\underline{\omega}_I}(\mathbf{x}_I,\mathbf{y}_I))\Bigg)\Bigg] \\
&= \left\langle f^{(1)}_{\underline{\omega}}(P_{\underline{\omega}}(\mathbf{x}_{[d]},\mathbf{y}_{[d]}))\right\rangle_{\square^d_\mu}
\end{aligned}
$$

where $f^{(1)}_{\tilde{\underline{\omega}}} = f_{(1,\tilde{\underline{\omega}}\cap[2,d])}$for any $\tilde{\underline{\omega}}_{[1,d]}$.
In the same way, we apply Cauchy-Schwartz's inequality in $(\mathbf{x}_{[3,d]},\mathbf{y}_{[3,d]})$ variables to end up with

$$
|\langle f_{\underline{\omega}}; \underline{\omega}\in\{0,1\}^d\rangle_{\square^d_\mu}|^4 \leq \prod_{\underline{\omega}_{[0,1]}} \langle f^{\underline{\omega}_{[1,2]}}_{\underline{\omega}}; \underline{\omega}\in\{0,1\}^d\rangle_{\square^d_\mu}
$$

Continue applying Cauchy-Schwartz's inequality consecutively in $(\mathbf{x}_{[4,d]},\mathbf{y}_{[4,d]})$, ..., $(\mathbf{x}_{[d,d]},\mathbf{y}_{[d,d]})$ variables, we end up with

$$
\begin{aligned}
|\langle f_{\underline{\omega}}; \underline{\omega}\in\{0,1\}^d\rangle_{\square^d_\mu}|^{2^d} &\leq \prod_{\underline{\omega}_{[d]}} \langle f^{\underline{\omega}}, ..., f^{\underline{\omega}}\rangle_{\square^d_\mu}, f^{\underline{\omega}} = f_{\underline{\omega}} \\
&\leq \prod_{\underline{\omega}_{[d]}} \|f_{\underline{\omega}}\|^{2^d}_{\square^d_\mu}
\end{aligned}
$$

□

Corollary 3.2.4. $\|\cdot\|_{\square^d_\mu}$ *is a norm for N is sufficiently large.*

Proof. First we show nonnegativity. By the linear forms condition, $\|1\|_{\square_\nu} = 1 + o(1)$. Hence by the Gowers-Cauchy-Schwartz inequality, we have $\|f\|_{\square^d_\mu} \gtrsim |\langle f, 1, ..., 1\rangle_{\square^d_\mu}| \geq 0$ for all sufficiently large N. Now

$$
\begin{aligned}
\|f+g\|_{\square^d_\mu} = \langle f+g, ..., f+g\rangle_{\square^d_\mu} &= \sum_{\underline{\omega}\in\{0,1\}^d} \langle h^{\omega_1}, ..., h^{\omega_d}\rangle_{\square^d_\mu}, h^\omega = \begin{cases} f & ,\omega = 0 \\ g & ,\omega = 1 \end{cases} \\
&\leq \sum_{\underline{\omega}\in\{0,1\}^d} \|h^{\omega_1}\|_{\square^d_\mu} \cdots \|h^{\omega_d}\|_{\square^d_\mu} \\
&= (\|f\|_{\square^d_\mu} + \|g\|_{\square^d_\mu})^{2^d}
\end{aligned}
$$

Also it follows directly from the definition that $\|\lambda f\|^{2^d}_{\square^d_\mu} = \lambda^{2^d}\|f\|^{2^d}_{\square^d_\mu}$. Since the norm are nonnegative, we have $\|\lambda f\|_{\square^d_\mu} = |\lambda|\,\|f\|_{\square^d_\mu}$. $\square$

Now we will prove Generalized von Neumann inequalities. The generalized von-Neumann inequality says that the average $\Lambda := \Lambda_{d+1,\mu}(f_I, I \subseteq [d+1], |I| = d)$, see (3.1.1), is controlled by the weighted box norm. We show this inequality in the general settings of an independent weight system.

Theorem 3.2.5 (Weighted generalized von-Neumann inequality for corner)**.** *Let $I \subseteq [d+1], |I| = d$, $f_I : X_I \to [0,1]$ bounded by 1. Write $f_{e_j} = f_{[d+1]\backslash\{j\}}$. Let $\underline{\nu}$ be an independent system of measure on $X_{[d+1]}$ that satisfies linear form conditions.*[1] *then*

$$|\Lambda_{d+1,\mu}(f_{e_1}, ..., f_{e_{d+1}})| \lesssim \min\{\|f_{e_1}\|_{\square_{\mu_{e_1}}}, ..., \|f_{e_{d+1}}\|_{\square_{\mu_{e_{d+1}}}}\} \tag{3.2.1}$$

Proof of Weighted Generalized von Neumann. Let $\mathcal{H}' = \{f \in \mathcal{H};\ |f| < d\}$, and write the left side of (3.2.1) as

$$\Lambda_{d+1,\mu} = \mathbb{E}_{x \in V_J} \prod_{e \in \mathcal{H}_d} f_e(x_e)\nu_e(x_e) \prod_{f \in \mathcal{H}'} \nu_f(x_f).$$

Write $e_j := [d+1]\backslash\{j\}, 1 \le j \le d+1$ for the faces. The idea is to apply the Cauchy-Schwartz inequality successively in the $x_1, x_2, \ldots, x_d$ variables to eliminate the functions[2] and weights $(f_{e_1}, \nu_{e_1}), \ldots, (f_{e_d}, \nu_{e_d})$, using the linear forms condition at each step, leaving $f_{e_{d+1}}$ on RHS.

Write $E := \Lambda_{d+1,\mu}$. To eliminate f_{e_1}, ν_{e_1} we have

$$|E| \le \mathbb{E}_{x_2,...,x_{d+1}} \nu_{e_1}(x_{e_1}) \prod_{1 \notin f \in \mathcal{H}'} \nu_f(x_f) \Big|\mathbb{E}_{x_1} \prod_{j \neq 2} f_{e_j}(x_j) \prod_{1 \in f \in \mathcal{H}'} \nu_f(x_f)\Big|.$$

By the linear forms condition $\mathbb{E}_{x_2,...,x_{d+1}} \nu_{e_1}(x_{e_1}) \prod_{1 \notin f \in \mathcal{H}'} \nu_f(x_f) = 1 + o_{N\to\infty}(1)$, thus by the Cauchy-Schwartz inequality

$$E^2 \lesssim \mathbb{E}_{x_2,...,x_{d+1}} \nu_{e_1}(x_{e_1}) \prod_{1 \notin f \in \mathcal{H}'} \nu_f(x_f)\, \mathbb{E}_{x_1,y_1} \prod_{j \neq 2} f_{e_j}\nu_{e_j}(x_1, x_{e_j\backslash\{1\}}) f_{e_j}\nu_{e_j}(y_1, x_{e_j\backslash\{1\}}) \tag{3.2.2}$$

$$\times \prod_{1 \in f \in \mathcal{H}'} \nu_f(y_1, x_{f\backslash\{1\}})\, \nu_f(x_1, x_{f\backslash\{1\}})$$

This eliminates f_{e_1}, ν_{e_1} and doubles the variable x_1 to the pair of variables (x_1, y_1) and also doubled each factor of the form $G_e(x_e)$ (which is either $f_e(x_e)$ or $\nu_e(x_e)$, for $e \in \mathcal{H}$) depending on the x_1 variable. To keep track of these changes as we continue with the rest of that variables, let us introduce

[1] This lemma indeed directly applicable to corner system where $f_I \le \nu_I$ by reinterpreting the weight attached to f^I as weight attached to X_I.

[2] Indeed, f is bounded by 1 an ν is positive so f_e can be trivially eliminated but they are also naturally the same way as ν_e

some notations. Let $g \subseteq [d]$ and for a function $G_e(x_e)$ define

$$G_e^*(x_{e\cap g}, y_{e\cap g}, x_{e\backslash g}) := \prod_{\omega_e \in \{0,1\}^{e\cap g}} G_e(\omega_e(x_{e\cap g}, y_{e\cap g}), x_{e\backslash g}). \tag{3.2.3}$$

We claim that after applying the Cauchy-Schwartz inequality in the $x_1, \ldots, x_i$ variables we have[3] with $g = [i]$

$$E^{2^i} \lesssim \mathbb{E}_{x_{[i]}, y_{[i]}, x_{J\backslash[i]}} \prod_{j \leq i} \nu_{e_j}^*(x_{[i]\cap e_j}, y_{[i]\cap e_j}, x_{e_j\backslash[i]}) \prod_{j>i} f_{e_j}^* \nu_{e_j}^*(x_{[i]\cap e_j}, y_{[i]\cap e_j}, x_{e_j\backslash[i]}) \tag{3.2.4}$$
$$\times \prod_{f \in \mathcal{H}'} \nu_f^*(x_{f\cap[i]}, y_{f\cap[i]}, x_{f\backslash[i]}).$$

For $i = 1$ this can be seen from (3.2.2). Note that the linear forms appearing in any of these factors are pairwise linearly independent. Assuming it holds for i separating the factors independent of the x_{i+1} variable, and eliminate $f_{e_{i+1}}$ and applying the Cauchy-Schwartz inequality we double the variable x_{i+1} to the pair (x_{i+1}, y_{i+1}) and each factor $G_e^*(x_{e\cap[i]}, y_{e\cap[i]}, x_{e\backslash[i]})$ depending on it, to obtain the factor

$$G_e^*(x_{e\cap[i+1]}, y_{e\cap[i+1]}, x_{e\backslash[i+1]}),$$

thus the formula holds for $i+1$. After finishing this process we have

$$E^{2^d} \lesssim \mathbb{E}_{x_{[d]}, y_{[d]}} \prod_{\omega \in \{0,1\}^d} f_{e_{d+1}}(\omega(x_{[d]}, y_{[d]}))\nu_{e_{d+1}}(\omega(x_{[d]}, y_{[d]})) \prod_{f \subseteq [d], f \neq e_0} \prod_{\omega_f \in \{0,1\}^f} \nu_f(\omega_f(x_f, y_f))\, \mathcal{W}(x_{[d]}, y_{[d]}),$$

where

$$\mathcal{W}(x_{[d]}, y_{[d]}) = \mathbb{E}_{x_{d+1}} \prod_{d+1 \in e \in \mathcal{H}} \prod_{\omega_e \in \{0,1\}^{e\cap[d]}} \nu_e(\omega_e(x_{e\cap[d]}, y_{e\cap[d]}, x_{e\backslash[d]})).$$

Thus to prove (3.2.1), it is enough to show that

$$\mathbb{E}_{x_{[d]}, y_{[d]}} \prod_{f \subseteq [d]} \prod_{\omega_f \in \{0,1\}^f} \nu_f(\omega_f(x_f, y_f))\, |\mathcal{W}(x_{[d]}, y_{[d]}) - 1| = o_{N\to\infty}(1).$$

This can be done with one more application of the Cauchy-Schwartz inequality in x_{d+1} variable leading to 4 terms involving the "big" weight functions $\mathcal{W}$ and $\mathcal{W}^2$. Each terms is however $1 + o_{N\to\infty}(1)$ by the linear forms condition, as the underlying linear forms are pairwise linearly independent: the forms $L_f(\omega_f(x_f, y_f))$ are pairwise independent for $f \subseteq [d]$, and depend on a different set of variables then the forms $L_e(\omega_e(x_{e\cap[d]}, y_{e\cap[d]}, x_{e\backslash[d]}))$ for $e \nsubseteq [d]$ defining the weight function $\mathcal{W}$. The new forms appearing in $\mathcal{W}^2$ are copies of the forms in $\mathcal{W}$ with the x_{d+1} variable replaced by a new variable y_{d+1} hence are independent of each other and the rest of the forms. This proves the proposi-

[3] Noting that for $j \leq i : [i] \cap e_j = [i]\backslash\{j\}$. For $j > i : [i] \cap e_j = [i]$.

tion for $f_{e_{d+1}}$ and we can prove for other f_{e_j} in the same way. $\square$

3.3 Dual Function Estimates

Definition 3.3.1 (Dual Function). *For $f, g : \mathbb{Z}_N^d \to \mathbb{R}$ define the weight inner product*

$$\langle f, g\rangle_\mu := \int_{X_{[d]}} f \cdot g d\mu_{X_{[d]}} = \mathbb{E}_{\mathbf{x}\in\mathbb{Z}_N^d} f(\mathbf{x})g(\mathbf{x}) \prod_{|I|<d} \nu_I(\mathbf{x}_I).$$

Define the dual function of f by

$$\mathcal{D}f = \mathcal{D}_\mu f := \mathbb{E}_{\mathbf{y}\in\mathbb{Z}_N^d} \prod_{\underline{\omega}\neq\underline{0}} f(P_{\underline{\omega}}(\mathbf{x},\mathbf{y})) \prod_{|I|<d} \prod_{\underline{\omega}_I\neq\underline{0}} \nu_I(P_{\underline{\omega}_I}(\mathbf{x}_I,\mathbf{y}_I))$$

So

$$\begin{aligned}\|f\|_{\square_\mu}^{2^d} &= \mathbb{E}_{\mathbf{x}\in\mathbb{Z}_N^d} f(\mathbf{x}) \left[\mathbb{E}_{\mathbf{y}\in\mathbb{Z}_N^d} \prod_{\underline{\omega}\neq\underline{0}} f(P_{\underline{\omega}}(\mathbf{x},\mathbf{y})) \prod_{|I|<d} \prod_{\underline{\omega}_I\neq\underline{0}} \nu_I(P_{\underline{\omega}_I}(\mathbf{x}_I,\mathbf{y}_I))\right] \prod_{|I|<d} \nu_I(\mathbf{x}_I)\\ &= \langle f, \mathcal{D}f\rangle_\mu\end{aligned}$$

In this section we prove the dual function estimate in our hypergraph system. We uses product of these dual functions as an uniformity obstruction in soft inverse theorem arguments as in [51], [36]. In [51], they allow K to be arbitrary large[4] and employ the correlation condition to avoid the infinite linear forms conditions which was not available at that time (this is the only place where they used correlation condition in [51]). We only needs linear form condition here as the parameter K could be arbitrarily finite but fixed constant, depending on α. This makes the number of linear form conditions involved depends on α.

Theorem 3.3.2. *For all $K \le K(\alpha)$ any independent measure system and any fixed $J \subseteq [d+1], |J| = d$, let $F_1, ..., F_K : X_J \to \mathbb{R},\ F_j(\mathbf{x}_J) \le \nu_J(\mathbf{x}_J)$ be given functions. Then for each $1 \le K \le K(\alpha)$ we have that*

$$\|\prod_{j=1}^{K} \mathcal{D}F_j\|_{\square_\mu^d}^* = O_\alpha(1)$$

Proof. Denote by I the subsets of a fixed set $J \subseteq [d+1],\ |J| = d$. First, for each $1 \le j \le K$, write

$$\mathcal{D}F_j(\mathbf{x}) = \mathbb{E}_{\mathbf{y}^j\in\mathbb{Z}_N^d} \prod_{\underline{\omega}\neq\underline{0}} F_j(P_{\underline{\omega}}(\mathbf{x},\mathbf{y}^j)) \prod_{|I|<d} \prod_{\underline{\omega}_I\neq\underline{0}} \nu_I(P_{\underline{\omega}_I}(\mathbf{x}_I,\mathbf{y}_I^j))$$

[4] In [50], K is the number of iterations in energy increment which is $2^{2^K}/\varepsilon$.

Now assume $\|f\|_{\square^d_\mu} \leq 1$ then

$$\left\langle f, \prod_{j=1}^K \mathcal{D}F_j \right\rangle_\mu = \mathbb{E}_{\mathbf{x}\in\mathbb{Z}_N^d} f(\mathbf{x}) \prod_{j=1}^K \mathcal{D}F_j(\mathbf{x}) \prod_{|I|<d} \nu_I(\mathbf{x}_I)$$

$$= \mathbb{E}_{\mathbf{x}\in\mathbb{Z}_N^d} f(\mathbf{x}) \mathbb{E}_{\mathbf{y}^1,\ldots \mathbf{y}^K\in\mathbb{Z}_N^d} \prod_{j=1}^K \Big(\prod_{\underline{\omega}\neq\underline{0}} F_j(P_{\underline{\omega}}(\mathbf{x},\mathbf{y}^j)) \prod_{|I|<d} \Big[\prod_{\underline{\omega}_I\neq\underline{0}} \nu_I(P_{\underline{\omega}_I}(\mathbf{x}_I,\mathbf{y}_I^j)) \Big] \nu_I(\mathbf{x}_I) \Big)$$

We will compare this to the box norm to exploit the fact that $\|f\|_{\square^d_\nu} \leq 1$. To compare this to the Gowers's inner product, let us introduce the following change of variables:

For a fixed $\mathbf{y}\in\mathbb{Z}_N^d$, write $\mathbf{y}^j \mapsto \mathbf{y}^j + \mathbf{y}$ for $1 \leq j \leq K$ then our expression takes the form

$$\left\langle f, \prod_{j=1}^K \mathcal{D}F_j \right\rangle_\mu = \mathbb{E}_{\mathbf{y}^1,\ldots,\mathbf{y}^K} \mathbb{E}_{\mathbf{x}} f(\mathbf{x}) \prod_{j=1}^K \Big[\prod_{\underline{\omega}\neq\underline{0}} F_j(P_{\underline{\omega}}(\mathbf{x},\mathbf{y}+\mathbf{y}^j)) \prod_{|I|<d} \prod_{\underline{\omega}_I\neq\underline{0}} \Big[\nu_I(P_{\underline{\omega}_I}(\mathbf{x}_I,\mathbf{y}_I^j+\mathbf{y}_I)) \Big] \nu_I(\mathbf{x}_I) \Big]$$

Since $\mathbb{Z}_N^d$ is cyclic. This is equal to the average

$$\mathbb{E}_{\mathbf{y}^1,\ldots,\mathbf{y}^K\in\mathbb{Z}_N^d} \mathbb{E}_{\mathbf{x},\mathbf{y}\in\mathbb{Z}_N^d} f(\mathbf{x}) \prod_{j=1}^K \Big[\prod_{\underline{\omega}\neq\underline{0}} F_j(P_{\underline{\omega}}(\mathbf{x},\mathbf{y}+\mathbf{y}^j)) \prod_{|I|<d} \prod_{\underline{\omega}_I\neq\underline{0}} \Big[\nu_I(P_{\underline{\omega}_I}(\mathbf{x}_I,\mathbf{y}_I^j+\mathbf{y}_I)) \Big] \nu_I(\mathbf{x}_I) \Big]$$

For $\omega\in\{0,1\}^d, Y=(y^1,\ldots,y^k)\in(\mathbb{Z}_N^d)^k$. We will define functions $G_{\underline{\omega},Y}(\mathbf{x}) : \mathbb{Z}_N^d \to \mathbb{R}$ such that

$$\left\langle f, \prod_{j=1}^K \mathcal{D}F_j \right\rangle_\mu = \mathbb{E}_{\mathbf{y}^1,..,\mathbf{y}^K} \left\langle G_{\underline{\omega},Y}; \underline{\omega}\in\{0,1\}^d \right\rangle_{\square^d_\mu}$$

To do this, let $G_{\underline{0}}(\mathbf{x}) := f(\mathbf{x})$ and for each $\underline{\tilde{\omega}} \neq \underline{0}, Y$, define

$$G_{\underline{\tilde{\omega}},Y}(\mathbf{x}) := \prod_{j=1}^K \left[F_j(\mathbf{x}+\mathbf{y}^j_{1(\underline{\tilde{\omega}})}) \Big(\prod_{|I|<d} \nu_I((\mathbf{x}+\mathbf{y}^j_{1(\underline{\tilde{\omega}})})\big|_I) \Big)^{\frac{1}{2^{d-|I|}}} \right] \times \prod_{|I|<d} \nu_I(\mathbf{x}_I)^{-\frac{1}{2^{d-|I|}}}$$

Hence for $\underline{\tilde{\omega}} \neq \underline{0}$

$$G_{\underline{\tilde{\omega}},Y}(P_{\underline{\tilde{\omega}}}(\mathbf{x},\mathbf{y})) = \prod_{j=1}^K \left[F_j(P_{\underline{\tilde{\omega}}}(\mathbf{x},\mathbf{y}+\mathbf{y}^j)) \big(\prod_{|I|<d} \nu_I((P_{\underline{\tilde{\omega}}}((\mathbf{x},\mathbf{y}+\mathbf{y}^j)\big|_I)^{\frac{1}{2^{d-|I|}}} \right] \times \prod_{|I|<d} \nu_I(P_{\underline{\tilde{\omega}}}(\mathbf{x},\mathbf{y})\big|_I)^{-\frac{1}{2^{d-|I|}}}$$

Remark 3.3.3. *For each $I \subseteq [d]$ and fixed $\underline{\omega}_I$, the number of $\underline{\omega}_{[d]}$ such that $\underline{\omega}_{[d]}|_I = \underline{\omega}_I$ is $2^{d-|I|}$ and*

$$P_{\underline{\omega}}(\mathbf{x},\mathbf{y})|_I = P_{\underline{\omega}_I}(\mathbf{x}_I,\mathbf{y}_I) \Longleftrightarrow \underline{\omega}|_I = \underline{\omega}_I$$

Hence

$$\left\langle G_{\underline{\omega},Y};\underline{\omega}\in\{0,1\}^d\right\rangle_{\square^d_\mu}=\mathbb{E}_{\mathbf{x},\mathbf{y}\in\mathbb{Z}_N^d}\prod_{\underline{\omega}}G_{\underline{\omega},Y}(P_{\underline{\omega}}(\mathbf{x},\mathbf{y}))\times\prod_{|I|<d}\prod_{\underline{\omega}_I}\nu_I(P_{\underline{\omega}_I}(\mathbf{x}_I,\mathbf{y}_I))$$

$$=\mathbb{E}_{\mathbf{x},\mathbf{y}\in\mathbb{Z}_N^d}\prod_{\underline{\omega}}\Big[\prod_{j=1}^K\Big[F_j(P_{\underline{\omega}}(\mathbf{x},\mathbf{y}+\mathbf{y}^j))(\prod_{|I|<d}\nu_I(P_{\underline{\omega}}((\mathbf{x},\mathbf{y})+\mathbf{y}^j_{1(\underline{\omega})})\big|_I)^{\frac{1}{2^{d-|I|}}}\Big]$$

$$\times\prod_{|I|<d}\nu_I(P_{\underline{\omega}}(\mathbf{x}_I,\mathbf{y}_I)|_I)^{-\frac{1}{2^{d-|I|}}}\Big]\times\prod_{|I|<d}\prod_{\underline{\omega}_I}\nu_I(P_{\underline{\omega}_I}(\mathbf{x}_I,\mathbf{y}_I))$$

$$=\mathbb{E}_{\mathbf{x},\mathbf{y}\in\mathbb{Z}_N^d}f(\mathbf{x})\Big[\prod_{j=1}^K\prod_{\underline{\omega}\neq\underline{0}}F_j(P_{\underline{\omega}}(\mathbf{x},\mathbf{y}+\mathbf{y}^j))\Big]\times\prod_{|I|<d}\Big[\Big(\prod_{j=1}^K\prod_{\underline{\omega}_I\neq\underline{0}}\nu_I(P_{\underline{\omega}_I}(\mathbf{x}_I,\mathbf{y}^j_I+\mathbf{y}_I))\Big)\nu_I(\mathbf{x}_I)\Big]$$

Hence we have

$$\langle f,\prod_{j=1}^K\mathcal{D}F_j\rangle_\mu=\mathbb{E}_{\mathbf{y}^1,..,\mathbf{y}^K}\left\langle G_{\underline{\omega},Y};\underline{\omega}\in\{0,1\}^d\right\rangle_{\square^d_\mu}$$

Then by Gowers-Cauchy-Schwartz's and arithmetic-geometric mean inequality, we have

$$\big|\langle f,\prod_{j=1}^K\mathcal{D}F_j\rangle_\mu\big|\leq\mathbb{E}_{\mathbf{y}^1,\ldots,\mathbf{y}^K}\|f\|_{\square^d_\mu}\prod_{\underline{\omega}\neq\underline{0}}\|G_{\underline{\omega},Y}\|_{\square^d_\mu}\lesssim\mathbb{E}_{\mathbf{y}^1,\ldots,\mathbf{y}^K}\big(1+\sum_{\underline{\omega}_{[d]}\neq\underline{0}}\|G_{\underline{\omega},Y}\|^{2^d}_{\square^d_\mu}\big)$$

Hence to prove the dual function estimate, it is enough to show that

$$\mathbb{E}_{\mathbf{y}^1,\ldots,\mathbf{y}^K}\|G_{\underline{\tilde\omega},Y}\|^{2^d}_{\square^d_\mu}=O(1)$$

For any fixed $\underline{\tilde\omega}\neq\underline{0}$. Now

$$\mathbb{E}_{\mathbf{y}^1,\ldots,\mathbf{y}^K}\|G_{\underline{\tilde\omega},Y}\|^{2^d}_{\square^d_\mu}=\mathbb{E}_{\mathbf{y}^1,\ldots,\mathbf{y}^K}\mathbb{E}_{\mathbf{x},\mathbf{y}}\prod_{\underline{\omega}}G_{\underline{\tilde\omega},Y}(P_{\underline{\omega}}(\mathbf{x},\mathbf{y}))\prod_{|I|<d}\prod_{\underline{\omega}_I}\nu_I(P_{\underline{\omega}_I}(\mathbf{x}_I,\mathbf{y}_I))$$

$$\leq\mathbb{E}_{\mathbf{y}^1,\ldots,\mathbf{y}^K}\mathbb{E}_{\mathbf{x},\mathbf{y}}\prod_{\underline{\omega}}\prod_{j=1}^K\Big[\nu_{[d]}(P_{\underline{\omega}}(\mathbf{x},\mathbf{y})+\mathbf{y}^j_{1(\underline{\tilde\omega})})\prod_{|I|<d}\nu_I((P_{\underline{\omega}}(\mathbf{x},\mathbf{y})+\mathbf{y}^j_{1(\underline{\tilde\omega})})\big|_I)^{\frac{1}{2^{d-|I|}}}$$

$$\times\prod_{|I|<d}\nu_I(P_{\underline{\omega}}(\mathbf{x},\mathbf{y})\big|_I)^{-\frac{1}{2^{d-|I|}}}\Big]\times\prod_{|I|<d}\prod_{\underline{\omega}_I}\nu_I(P_{\underline{\omega}_I}(\mathbf{x}_I,\mathbf{y}_I))$$

$$=\mathbb{E}_{\mathbf{y}^1,\ldots,\mathbf{y}^K}\mathbb{E}_{\mathbf{x},\mathbf{y}}\prod_{j=1}^K\Big[\prod_{\underline{\omega}}\nu_{[d]}(P_{\underline{\omega}}((\mathbf{x},\mathbf{y})+\mathbf{y}^j_{1(\underline{\tilde\omega})}\big|_I))\prod_{|I|<d}\prod_{\underline{\omega}_I}\nu_I(P_{\underline{\omega}_I}(\mathbf{x}_I,\mathbf{y}_I)+\mathbf{y}^j_{1(\underline{\tilde\omega})}\big|_I)\Big]$$

by remark 3.3.3 above. As the linear forms appearing in the above expression are pairwise linearly independent this is $O_\alpha(1)$ (in fact $1+o_\alpha(1)$) by the linear forms condition as required. □

3.4 Transference Principle

In this section, we will modify the transference principle in [37](see Theorem 3.4.6). We will work on the set on which our functions have bounded dual, treating the contributions of the remaining set as error terms.

We will work on functions $f : X_I \to \mathbb{R}$, dominated by ν_I. WLOG $I = [d]$. Let $\langle \cdot \rangle$ be any inner product on $\mathcal{F} := \{f : X_{[d]} \to \mathbb{R}\}$ written as $\langle f, g \rangle = \int f \cdot g \, d\mu$ for some measure μ on $X_{[d]}$. In this section we will need the explicit description of the set $\Omega(T)$ that the dual function is bounded by T and for this we will need correlation condition (in particular, Proposition 3.4.3 below. This is the only place in this thesis that we will need correlation condition).

We will do this on the set on which our functions have bounded dual, and treat the contributions of the remaining set as error terms. We will need the explicit description of the set $\Omega(T)$ that the dual function is bounded by T using the correlation condition (In particular, that this set is of lower complexity, missing one of the variable). In general, T will depend on ϵ and $T \to \infty$ as $\epsilon \to 0$ but when we apply removal lemma we will choose ϵ to be some small number depending on α and hence if α is fixed then we can regard ϵ as a fixed small constant and T as a fixed large constant.

Definition 3.4.1. *For each $T > 1$ we have th set $\Omega(T)$ and define the following sets*

$$\begin{aligned}
\mathcal{F} &:= \{f : X_{[d]} \to \mathbb{R}\} \\
\mathcal{F}_T &:= \{f \in \mathcal{F} : \mathit{supp}(f) \subseteq \Omega(T)\} \\
\mathcal{S}_T &:= \{f \in \mathcal{F}_T : |f| \leq \nu_{[d]}(\mathbf{x}_{[d]}) + 2\}
\end{aligned}$$

We define the following (basic anti-correlation) norm on $\mathcal{F}_T$

$$\|f\|_{BAC} := \|f\|_{BAC,\mu} = \max_{g \in \mathcal{S}_T} |\langle f, \mathcal{D}g \rangle_\mu| \geq \|f\|_{\square_\mu}^{2^d}.$$

If $f \in \mathcal{S}_T$, this norm measures how much the functions on $\mathcal{F}_T$ correlate with function on $\mathcal{S}_T$ and our weighted box norm is denominated by this norm. This will allow us to work on $BAC-$norm (rather than the box norm) which has some useful algebraic properties. Recall

$$\mathcal{D}f = \mathbb{E}_{\mathbf{y}} \prod_{\underline{\omega} \neq \mathbf{0}} f(P_{\underline{\omega}}(\mathbf{x}, \mathbf{y})) \prod_{|I| < d} \prod_{\underline{\omega}_I \neq \mathbf{0}} \nu_I(P_{\underline{\omega}_I}(\mathbf{x}_I, \mathbf{y}_I))$$

Also recall that $L^I(\mathbf{x})$ are linear forms appeared in the definition of ν_I. Write $h_{\underline{\omega}_I} = L^I(\mathbf{x})|_{0(\underline{\omega}_I)}$ hence using correlation condition (Definition 2.2.8), we have

$$|\mathcal{D}f| \leq \prod_{\emptyset \neq J \subseteq [d]} \sum_{(\underline{\omega}_{I_1}, \underline{\omega}_{I_2}) \in T_J} \tau(W \cdot (a_{\underline{\omega}_{I_1}} h_{\underline{\omega}_{I_1}} - a_{\underline{\omega}_{I_2}} h_{\underline{\omega}_{I_2}}) + (a_{\underline{\omega}_{I_1}} - a_{\underline{\omega}_{I_2}}) b) \tag{3.4.1}$$

where for each $J \subsetneq [d], J \neq \emptyset$

$$T_J := \{(\underline{\omega}_{I_1}, \underline{\omega}_{I_2}), \underline{\omega}_{I_1}, \underline{\omega}_{I_2} \neq \underline{0}, 1(\underline{\omega}_{I_1}) = 1(\underline{\omega}_{I_2}) = J : \exists c \in \mathbb{Q}, L^{I_1}(\mathbf{y}_{1(\underline{\omega}_{I_1})}) = cL^{I_2}(\mathbf{y}_{1(\underline{\omega}_{I_2})})\}$$

where $a_{\underline{\omega}_{I_j}} \in \mathbb{Z}$. Define

$$\Omega_J(T) = \{(\mathbf{x}_{[d]} : \sum_{(\underline{\omega}_{I_1}, \underline{\omega}_{I_2}) \in T_J} \tau(W \cdot (a_{\underline{\omega}_{I_1}} h_{\underline{\omega}_{I_1}} - a_{\underline{\omega}_{I_2}} h_{\underline{\omega}_{I_2}}) + (a_{\underline{\omega}_{I_1}} - a_{\underline{\omega}_{I_2}})b)) \leq T^{1/2^d}\} \quad (3.4.2)$$

$$\Omega(T) = \bigcap_{\emptyset \neq J \subsetneq [d]} \Omega_J(T) \quad (3.4.3)$$

So $\mathcal{D}f$ is bounded by T on $\Omega(T)$ for any fixed $T > 1$. Indeed, when T is large, the set $\Omega(T) \neq \emptyset$. Explicit description of $\Omega(T)$ is only used in the proof of property 1 in Proppposition 3.4.3 below.

Example 3.4.2. *In 2-dimension corner, let $f : X_1 \times X_2 \to \mathbb{R}$, we have*

$$\begin{aligned}
\mathcal{D}f(x_1, x_2) &= \mathbb{E}_{y_1,y_2} f(x_1, y_2) f(y_1, x_2) f(y_1, y_2) \nu(y_1) \nu(y_2) \\
&\leq \mathbb{E}_{y_1,y_2} \nu(x_2 - y_1) \nu(y_2 - x_1) \nu(y_2 - y_1) \nu(y_1) \nu(y_2) \\
&\leq \tau(W \cdot x_1) \tau(W \cdot x_2)
\end{aligned}$$

Then define $\Omega_1(T) := \{(x_1, x_2) : \tau(W \cdot x_2) \leq T^{1/2}\}, \Omega_2(T) := \{(x_1, x_2) : \tau(W \cdot x_1) \leq T^{1/2}\}$. Let $\Omega_{\{1,2\}}(T) := \Omega_1(T) \cap \Omega_2(T)$ then $\mathcal{D}f$ is bounded by T on $\Omega_1(T) \cap \Omega_2(T)$.

We have the following basic properties of this norm.

Proposition 3.4.3.

1. *$g \in \mathcal{F}_T \Rightarrow \mathcal{D}g \in \mathcal{F}_T$*
2. *$\|\cdot\|_{BAC}$ is a norm on $\mathcal{F}_T$ and can be extended to be a seminorm on $\mathcal{F}$. Furthermore, we have $\|f\|_{BAC} = \|f \cdot \mathbf{1}_{\Omega(T)}\|_{BAC}, f \in \mathcal{F}$.*
3. *Span$\{\mathcal{D}g : g \in \mathcal{S}_T\} = \mathcal{F}_T$*
4. *$\|f\|^*_{BAC} = \inf\{\sum_{i=1}^k |\lambda_i|, f = \sum_{i=1}^k \lambda_i \mathcal{D}g_i; g_i \in \mathcal{S}_T\}$ for $f \in \mathcal{F}_T$. Hence the dual of BAC norm measure how one can write f as a linear combination of $\mathcal{D}g_i; g_i \in \mathcal{S}_T$.*

Remark 3.4.4. *If $f \notin \mathcal{F}_T$ then supp$(f) \nsubseteq \Omega(T)$ so f is not of the form $\sum_{i=1}^k \lambda_i \mathcal{D}g_i; g_i \in \mathcal{F}_T$ as RHS is zero.*

Proof. 1. Suppose $(\tilde{x}_1, ..., \tilde{x}_d) \in \Omega(T)^C$ then there is an $J \subsetneq [d]$ such that $K_J(\tilde{\mathbf{x}}_{[d]\setminus J}) > T$ where K_J is the function in the definition of $\Omega_J(T)$ for some J. Let $g \in \mathcal{F}_T$ then $g(\tilde{\mathbf{x}}_{[d]\setminus J}, \mathbf{x}_J) = 0$ for all $\mathbf{x}_J \in X_J$ So

$$\mathcal{D}g(\tilde{\mathbf{x}}_{[d]\setminus J}, \mathbf{x}_J) = g(\tilde{\mathbf{x}}_{[d]\setminus J}, \mathbf{x}_J)E(x) = 0$$

for some function E so $\mathcal{D}g \in \mathcal{F}_T$.

2. It follows directly from the definition that $\|f+g\|_{\text{BAC}} \leq \|f\|_{\text{BAC}} + \|g\|_{\text{BAC}}$ and $\|\lambda f\|_{\text{BAC}} = |\lambda| \, \|f\|_{\text{BAC}}$ for any $\lambda \in \mathbb{R}$. Now suppose $f \in \mathcal{F}_T$, f is not identically zero then we need to show that $\|f\|_{\text{BAC}} \neq 0$. Since f is defined on a finite set, we have that $\|f\|_\infty < \infty$. Let $g = \gamma f$ where γ is a constant such that $\|g\|_\infty < 2$ then $g \in \mathcal{S}_T$ and $\langle f, \mathcal{D}g \rangle_\mu = \langle f, \mathcal{D}\gamma f \rangle_\mu = \gamma^{2^d-1} \langle f, \mathcal{D}f \rangle_\mu > 0$ so $\|f\|_{\text{BAC}} > 0$.
   Now since $\operatorname{supp}(\mathcal{D}g) \subseteq \Omega(T)$ we have for any $f \in \mathcal{F}$
   $$\|f\|_{\text{BAC}} = \sup_{g \in \mathcal{S}_T} |\langle f, \mathcal{D}g \rangle_\mu| = \sup_{g \in \mathcal{S}_T} |\langle f \cdot \mathbf{1}_{\Omega(T)}, \mathcal{D}g \rangle_\mu| = \left\| f \cdot \mathbf{1}_{\Omega(T)} \right\|_{\text{BAC}}$$
3. If there is an $f \in \mathcal{F}_T$, f is not identically zero and $f \notin \operatorname{span}\{\mathcal{D}g : g \in \mathcal{S}_T\}$ So $f \in \operatorname{span}\{\mathcal{D}g : g \in \mathcal{S}_T\}^\perp$ then $\langle f, \mathcal{D}g \rangle = 0$ for all $g \in \mathcal{S}_T$. So $\|f\|_{BAC} = 0$ which is a contradiction.
4. This is a standard argument. Define $\|f\|_D = \inf\{\sum_{i=1}^k |\lambda_i| : f = \sum_{i=1}^k \lambda_i \mathcal{D}g_i, g_i \in \mathcal{S}_T\}$ which can be easily verified to be a norm on $\mathcal{F}_T$. Now let $\phi, f \in \mathcal{F}_T$, $f = \sum_{i=1}^k \lambda_i \mathcal{D}g_i, g_i \in \mathcal{S}_T$, then
   $$|\langle \phi, f \rangle| = \sum_{i=1}^k |\lambda_i| |\langle \phi, \mathcal{D}g_i \rangle| \leq \|\phi\|_{\text{BAC}} \sum_{i=1}^k |\lambda_i| \leq \|\phi\|_{\text{BAC}} \|f\|_D$$
   so
   $$\|f\|_{\text{BAC}}^* \leq \|f\|_D$$
   Next for all $g \in \mathcal{S}_T$, we have $\|\mathcal{D}g\|_D \leq 1$ then
   $$\|f\|_{\text{BAC}} = \sup_{g \in \mathcal{S}_T} |\langle f, \mathcal{D}g \rangle| \leq \sup_{\|h\|_D \leq 1} |\langle f, h \rangle| = \|f\|_D^*$$
   so $\|f\|_{\text{BAC}} \leq \|f\|_D^*$ i.e. $\|f\|_{\text{BAC}}^* \geq \|f\|_D$. So $\|f\|_{\text{BAC}}^* = \|f\|_D$.

$\square$

Now let us prove the following lemma whose proof relies on the dual function estimate. From here we consider our inner product $\langle \, \cdot \, \rangle_\mu$ and the norm $\| \cdot \|_{\square_\mu}$. This argument also works for any norm for which one has the dual function estimate.

Lemma 3.4.5. *Let $\phi \in \mathcal{F}_T$ be such that $\|\phi\|_{BAC}^* \leq C$ and $\eta > 0$. Let $\phi_+ := \max\{0, \phi\}$. Then there is a polynomial $P(u) = a_m u^m + ... + a_1 u + a_0$ such that*

1. $\|P(\phi) - \phi_+\|_\infty \leq \eta$
2. $\|P(\phi)\|_{\square_\mu^d}^* \leq \rho(C, T, \eta)$

where
$$\rho(C, T, \eta) := 2 \inf R_P(C)$$

where the infimum is taken over polynomials P such that $\|P - \phi_+\|_\infty \leq \eta$ on $[-CT, CT]$ and

$$R_P(x) = \sum_{j=0}^{m} C(j)|a_j|x^j, \text{ where } C(m) \text{ is the constant in the dual function estimate}$$

Proof. First, recall that if $(x_1, ..., x_d) \in \text{supp}(\mathcal{D}g_i) \subseteq \Omega(T)$ then

$$|\mathcal{D}g(x_1, ..., x_d)| \leq T$$

Now suppose $\|\phi\|^*_{\text{BAC}} \leq C$ then there exist $g_1, .., g_k \in \mathcal{S}_T$ and $\lambda_1, ..., \lambda_k$ such that $\phi = \sum_{i=1}^{k} \lambda_i \mathcal{D}g_i$ and $\sum_{1\leq i\leq k} |\lambda_i| \leq C$. Hence applying the boundedness of the duals,

$$|\phi(x_1, ..., x_d)| \leq (\sum_{i=1}^{k} |\lambda_i|)(\max_{1\leq i\leq k} |\mathcal{D}g_i(x_1, .., x_d)|) \leq CT$$

Hence *the Range of* $\phi = \phi(\Omega(T)) \subseteq [-CT, CT]$. Then by Weierstrass approximation theorem, there is a polynomial P (which may depend on C, T, η) such that $R_P(C) \leq \rho$ and

$$|P(u) - u_+| \leq \eta \quad \forall |u| \leq CT$$

and so $\|P(\phi) - \phi_+\|_\infty \leq \eta$ and we have (1).
Now using the dual function estimate, we have

$$\begin{aligned} \|\phi^m\|^*_{\square^d_\mu} &\leq \big\|(\sum_{1\leq i\leq k} \lambda_i \mathcal{D}g_i)^m\big\|^*_{\square^d_\mu} \leq \sum_{1\leq i_1\leq ...\leq i_m\leq k} |\lambda_{i_1}...\lambda_{i_m}| \, \|\mathcal{D}g_{i_1}...\mathcal{D}g_{i_m}\|^*_{\square^d_\mu} \\ &\leq C(m) \sum_{1\leq i_1\leq ...\leq i_m\leq k} |\lambda_{i_1}...\lambda_{i_m}| \leq C(m)(\sum_{1\leq i\leq k} |\lambda_i|)^m \leq C(m)C^m \end{aligned}$$

Hence $\|P(\phi)\|^*_{\square^d_\mu} \leq \sum_{m=0}^{d} |a_m| C(m) C^m \leq \rho(C, T, \eta)$ $\square$

Now we are ready to prove the transference principle.

Theorem 3.4.6 (Transference Principle)**.** *Suppose ν gives an independent weight system. Let $f \in \mathcal{F}$ and $0 \leq f(\mathbf{x}_{[d]}) \leq \nu_{[d]}(\mathbf{x}_{[d]})$, let $\eta > 0$. Suppose $N \geq N(\eta, T)$ is large enough, then there are functions g, h on $X_1 \times ... \times X_d$ such that*

$$f = g + h \text{ on } \Omega(T), \ \ 0 \leq g \leq 2 \text{ on } \Omega(T), \quad \big\|h \cdot \mathbf{1}_{\Omega(T)}\big\|_{\square^d_\mu} \leq \eta$$

Since $h \cdot \mathbf{1}_{\Omega(T)} = f \cdot \mathbf{1}_{\Omega(T)} - g \cdot \mathbf{1}_{\Omega(T)}$, we have $-2 \leq h \cdot \mathbf{1}_{\Omega(T)} \leq \nu$ so $|h \cdot \mathbf{1}_{\Omega(T)}| \leq \nu + 2$ so $h \cdot \mathbf{1}_{\Omega(T)} \in \mathcal{S}_T$. Hence by the definifion of BAC-norm,

$$\eta \geq \|h \cdot \mathbf{1}_{\Omega(T)}\|_{\text{BAC}} \geq \langle h \cdot \mathbf{1}_{\Omega(T)}, \mathcal{D}(h \cdot \mathbf{1}_{\Omega(T)})\rangle_\nu = \|h \cdot \mathbf{1}_{\Omega(T)}\|^{2^d}_{\square^d_\mu}$$

To prove Theorem 3.4.6, it suffices to show $\left\|h \cdot \mathbf{1}_{\Omega(T)}\right\|_{\mathrm{BAC}} \leq \eta^{1/2^d}$ (WLOG, replace $\eta^{1/2^d}$ with η). Here the BAC-norm is the BAC-norm with respect to $\langle\ \cdot\ \rangle_\mu$

The following lemma will be used in the next proof.

Lemma 3.4.7 (Finite dimensional Hahn-Banach's Theorem). *Let $X = \mathbb{R}^d$ be a norm space and $f \in X, \|f\| \geq 1$. Then there is a vector $\phi \in \mathbb{R}^d$ such that $\langle f, \phi\rangle > 1$ and $|\langle g, \phi\rangle| \leq 1$ whenever $\|g\| \leq 1$. (i.e. $\|\phi\|^* \leq 1$)*

The following lemma is an easy corollary of the Hahn-Banach's theorem by considering the following norm with given $c_1, c_2 > 0$: $\|f\| = \inf\{c_1\|f_1\|_1 + c_2\|f_2\|_2; f = f_1 + f_2\}$ with its dual $\|f\|^* = \max\{c_1^{-1}\|f_1\|^*, c_2^{-1}\|f\|^*\}$.

Lemma 3.4.8 ([37], Cor. 3.2). *Let K_1, K_2 be closed convex subsets of $\mathbb{R}^d$, each containing 0 and suppose $f \in \mathbb{R}^d$ cannot be written as a sum $f_1 + f_2, f_1 \in c_1K_1, f_2 \in c_2K_2, c_1 > 0, c_2 > 0$. Then there is a linear functional ϕ such that $\langle f, \phi\rangle > 1$ and $\langle f_1, \phi\rangle \leq c_1^{-1}$ for all $f_1 \in K_1$, $\langle f_2, \phi\rangle \leq c_2^{-1}$ for all $f_2 \in K_2$.*

Proof of Theorem 3.4.6: Define

$$K_1 := \{g \in \mathcal{F} : 0 \leq g \leq 2 \text{ on } \Omega(T)\}, \quad K_2 := \{h \in \mathcal{F} : \|h\|_{\mathrm{BAC}} \leq \eta\}.$$

It is clear that K_1, K_2 are convex. (Also $0 \in K_1, 0 \in \mathrm{Int}(K_2)$). Assume that $f \notin K_1 + K_2$ on $\Omega(T)$ then by Lemma 3.4.8, there exists $\phi \in \mathcal{F}$ such that

$$\left\langle \phi, f \cdot \mathbf{1}_{\Omega(T)}\right\rangle_\mu > 1, \quad \langle \phi, g\rangle_\mu \leq 1 \ \ \forall g \in K_1, \quad \langle \phi, h\rangle_\mu \leq 1 \ \ \forall h \in K_2$$

First, we claim that $\phi \in \mathcal{F}_T$. To see this, suppose g is a function whose $\mathrm{supp}(g) \subseteq \Omega(T)^C$ (i.e. $g \equiv 0$ on $\Omega(T)$ so $g \in K_1$.) Since $g \in K_1, \langle \phi, g\rangle_\mu \leq 1$ but g could be chosen arbitrarily on $\Omega(T)^C$ so we must have $\phi\big|_{\Omega(T)^C} \equiv 0$ and hence $\phi \in \mathcal{F}_T$. Now let

$$g(\mathbf{x}_{[d]}) = \begin{cases} 2 & \text{if } \phi(\mathbf{x}_{[d]}) \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

then $g \in K_1$ and

$$\langle \phi, g\rangle_\mu = \langle \phi_+, 2\rangle_\mu = 2\,\langle \phi_+, 1\rangle_\mu \leq 1 \Rightarrow \langle \phi_+, 1\rangle_\mu \leq \frac{1}{2}$$

Now we want to show that $\|\phi\|^*_{BAC} <\leq \eta^{-1}$. Since $\phi \in \mathcal{F}_T, h \in K_2, \|h \cdot \mathbf{1}_{\Omega(T)^C}\|_{\mathrm{BAC}} \leq \eta$ then we have

$$\langle \phi, \eta h \cdot \mathbf{1}_{\Omega(T)^C}\rangle_\mu = \langle \phi, \eta h\rangle_\mu \leq 1.$$

Hence if $h' \in \mathcal{F}_T$ and $\|h'\|_{\text{BAC}} \leq 1$ then $\left\|h' \cdot \mathbf{1}_{\Omega(T)}\right\|_{\text{BAC}} = \|h'\|_{\text{BAC}} \leq 1$ so

$$\langle \phi, h' \rangle_\mu \leq \eta^{-1} \quad \forall h' \in \mathcal{F}_T, \left\|h'\right\|_{\text{BAC}} \leq 1$$

so $\|\phi\|^*_{\text{BAC}} \leq \eta^{-1}$ as $\|\cdot\|_{\text{BAC}}$ is a norm on $\mathcal{F}_T$.

Now we want to invoke positivity of ϕ_+. By the Lemma 3.4.5, there is a polynomial P such that

$$\|P(\phi) - \phi_+\|_\infty \leq \frac{1}{8} \quad \text{and} \quad \|P(\phi)\|^*_{\square^d_\mu} \leq \rho(C, T, \eta)$$

Then $\langle P(\phi), 1 \rangle_\mu \leq \langle P(\phi) - \phi_+, 1 \rangle_\mu + \langle \phi_+, 1 \rangle_\mu \leq \frac{1}{2} + \frac{1}{8}$ Also, from the definition of the weighted box norm and the linear form condition, we have

$$\|\nu_{[d]}(\mathbf{x}_{[d]}) - 1\|^{2^d}_{\square^d_\mu} = o_{N\to\infty}(1)$$

so suppose $N \geq N(T, \eta)$ then

$$\langle P(\phi), \nu \rangle_\mu = \langle P(\phi), 1 \rangle_\mu + \langle P(\phi), \nu - 1 \rangle_\mu \leq \frac{1}{2} + \frac{1}{8} + \|P(\phi)\|^*_{\square^d_\mu} \|\nu - 1\|_{\square^d_\mu} \leq \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$$

$$|\langle \nu, \phi_+ \rangle_\mu| = |\langle \nu, \phi_+ - P(\phi) \rangle_\mu| + |\langle \nu, P(\phi) \rangle_\mu| \leq \|\phi_+ - P(\phi)\|_\infty \langle \nu, 1 \rangle_\mu + \langle \nu, P(\phi) \rangle_\mu \leq \frac{1}{8} \cdot \frac{1}{2} + \frac{3}{4}$$

By positivity of ϕ_+ we have $\langle f, \phi_+ \rangle_\mu \leq \langle \nu, \phi_+ \rangle_\mu$. Hence

$$\left\langle f \cdot \mathbf{1}_{\Omega(T)}, \phi \right\rangle_\mu = \langle f, \phi \rangle_\mu \leq \langle f, \phi_+ \rangle_\mu \leq \langle \nu, \phi_+ \rangle_\mu \leq \frac{3}{4} + \frac{1}{10} < 1$$

which is a contradiction. Hence $f \in K_1 + K_2$ on $\Omega(T)$. $\square$

Now we can rephrase Theorem 3.4.6 as follow:

Theorem 3.4.9 (Transference Principle)**.** *Suppose ν is an independent weight system. Let $f \in \mathcal{F}, 0 \leq f \leq \nu$ and $0 < \eta < 1 \ll T$ then there exists $f_1, f_2, f_3 \in \mathcal{F}$ such that*

1. $f = f_1 + f_2 + f_3$
2. $0 \leq f_1 \leq 2$, *supp*$(f_1) \subseteq \Omega(T)$
3. $\|f_2\|_{\square^d_\mu} \leq \eta$, *supp*$(f_2) \subseteq \Omega(T)$
4. $0 \leq f_3 \leq \nu$, *supp*$(f_3) \subseteq \Omega(T)^C$, $\|f_3\|_{L^1_\mu} \lesssim \frac{1}{T}$.

Proof. Let g, h be as in Theorem 3.4.6. Take $f_1 = g \cdot \mathbf{1}_{\Omega_T}$, $f_2 = h \cdot \mathbf{1}_{\Omega_T}$ then $f \cdot \mathbf{1}_{\Omega_T} = f_1 + f_2$. Let

$f_3 = f \cdot \mathbf{1}_{\Omega_T^C}$. Now by the linear forms conditions,

$$\begin{aligned}\|f_3\|_{L^1_\mu} &\leq \frac{1}{T}\mathbb{E}_{\mathbf{x}_{[d+1]\setminus\{d\}}} f \cdot \prod_{I\subseteq[d+1]\setminus\{d\},|I|<d} \nu_I(\mathbf{x}_I)\cdot|\mathcal{D}f| \quad (\text{since } |\mathcal{D}f|>T) \\ &= \frac{1}{T}\mathbb{E}_{\mathbf{x}_{[d+1]\setminus\{d\}}}\mathbb{E}_{\mathbf{y}_{[d+1]\setminus\{d\}}} \prod_{I\subseteq[d+1]\setminus\{d\}} \nu_I(L^I(\mathbf{x}_I)) \prod_{I\subseteq[d+1]\setminus\{d\}} \prod_{\underline{\omega}_I\neq\underline{0}} \nu_I((P_{\underline{\omega}_I}(\mathbf{x}_I,\mathbf{y}_I))) \lesssim \frac{1}{T}.\end{aligned}$$

□

3.5 Relative Hypergraph Removal Lemma

First let us recall the statement of ordinary functional hypergraph removal lemma [104]. Recall the definition of Λ in equation (3.1.1).

Theorem 3.5.1. *Given probability measure spaces $(X_1,\mu_{X_1}),...,(X_{d+1},\mu_{X_{d+1}})$ and $f^{(i)}: X_I \to [0,1], I=[d+1]\setminus\{i\}$. Let $\epsilon>0$, suppose $|\Lambda_{d+1}(f^{(1)},...,f^{(d)},f^{(d+1)})|\leq\epsilon$. Then for $1\leq i\leq d$, there exists*

$$E_i \subseteq X_{[d+1]\setminus\{i\}}$$

such that $\prod_{1\leq j\leq d+1}\mathbf{1}_{E_j}\equiv 0$ and for $1\leq i\leq d+1$,

$$\int_{X_1}\cdots\int_{X_{d+1}} f^{(i)}\cdot\mathbf{1}_{E_i^C}d\mu_{X_1}\cdots d\mu_{X_d}d\mu_{X_{d+1}}\leq\delta(\epsilon)$$

where $\delta(\epsilon)\to 0$ as $\epsilon\to 0$.

Remark 3.5.2. *In fact the paper [104] proves this theorem only with the counting measure (with the notion of $e-$discrepancy in place of Box norm). But the proof also works for any finite measure that has direct product structure (with the notion of weighted Box Norm) as the energy increment as in [104] or [18] would run through in the same way. See also [100] for the case of probability measures in $d=2,3$. However we don't know how to genralize this argument to arbitrary measure on the product space. If we can prove this theorem for any measure $\mu_{X_1\times...\times X_d}$ then we would be able to prove multidimensional Green-Tao's Theorem.*

The proof of this removal lemma relies on the following regularity lemma.

Theorem 3.5.3 (Szemerédi's Regularity (Tao) Lemma [104])**.** *Let $(X_{[d+1]},\mu)$ be a weighted hypergraph system with pseudorandom weight attached only on each edge of size 1. Let $f: X_{[d+1]}\to[0,1]$ be measurable, let $\tau>0$ and $F:\mathbb{N}\to\mathbb{N}$ be arbitrary increasing functions (possibly depends on τ). Then there is an integer $M=O_{F,\tau}(1)$, factors $\mathcal{B}_I(I\subseteq[d+1],|I|=d)$ on X_I of complexity at most M such that $f=f_1+f_2+f_3$ where*

– $f_1 = \mathbb{E}(f|\bigvee_{I\subseteq[d+1],|I|=d}\mathcal{B}_I)$.

– $\|f_2\|_{L^2_\mu} \leq \tau$.

– $\|f_3\|_{\square^d_\mu} \leq F(M)^{-1}$.

– $f_1, f_1+f_2 \in [0,1]$.

Remark 3.5.4. *A consequence from this theorem that we will use later is the following: since f_1 is a constant on each atom of $\bigvee_{|I|=d}\mathcal{B}_I$, we can decompose f_1 as a finite sum with $O_M(1)$ terms of lower complexity functions i.e.a finite sum of product $\prod_{i=1}^{d+1} J_i$ where J_i is a function in $\mathbf{x}_{[d+1]\setminus\{i\}}$ variable and takes values in $[0,1]$.*

Theorem 3.5.5 (Weighted Simplex-Removal Lemma)**.** *Suppose $f^{(i)}(\mathbf{x}_{[d+1]\setminus\{i\}}) \leq \nu_{[d+1]\setminus\{i\}}(\mathbf{x}_{[d+1]\setminus\{i\}})$. Let $\epsilon > 0$, Suppose $|\Lambda| \leq \epsilon$ then there exist $E_i \subseteq \prod_{j\in[d+1]\setminus\{i\}} X_j$ such that for $1 \leq i \leq d+1$,*

– $\prod_{i\in[d+1]} \mathbf{1}_{E_i} \equiv 0$

– $\int_{X_1}\cdots\int_{X_{d+1}} f^{(i)}\mathbf{1}_{E_i^C}d\mu_{X_1}\cdots d\mu_{X_{d+1}} = \mathbb{E}_{\mathbf{x}_{[d+1]\setminus\{i\}}}\mathbf{1}_{E_i^C}f^{(i)}(\mathbf{x}_{[d+1]\setminus\{i\}})\prod_{J\subsetneq[d+1]\setminus\{i\}}\nu_J(\mathbf{x}_J) \leq \delta(\epsilon)$

where $\delta(\epsilon) \to 0$ as $\epsilon \to 0$.

Proof. Using the transference principle (Theorem 3.4.9) for $1 \leq i \leq d+1$, write $f^{(i)} = g^{(i)} + h^{(i)} + k^{(i)}$ where

1. $f^{(i)} = g^{(i)} + h^{(i)} + k^{(i)}$
2. $0 \leq g^{(i)} \leq 2,\ \operatorname{supp}(g^{(i)}) \subseteq \Omega^{(i)}(T)$
3. $\|h^{(i)}\|_{\square^d_\mu} \leq \eta,\ \operatorname{supp}(h^{(i)}) \subseteq \Omega^{(i)}(T)$
4. $k^{(i)} = f^{(i)} \cdot \mathbf{1}_{(\Omega^{(i)})^C(T)}$

where

$$\Omega^{(i)}(T) = \{\mathbf{x}_{[d+1]\setminus\{i\}} : |\mathcal{D}f^{(i)}| \leq T\},\ 1 \leq i \leq d$$

Step 1: We will show that if $T \geq T(\epsilon)$ is sufficiently large then[5]

$$\Lambda_{d+1,\mu}(g^{(1)} + h^{(1)}, ..., g^{(d+1)} + h^{(d+1)}) = \Lambda_{d+1}(f^{(1)} - k^{(1)}, ..., f^{(d+1)} - k^{(d+1)}) \lesssim \epsilon.$$

Proof of Step 1: For $I \subseteq [d+1]$, the term on LHS can be written as a sum of the following terms:

$$\Lambda_{d+1,I,\mu}(e^{(1)}, ..., e^{(d)}, e^{(d+1)}), \quad e^{(i)} = \begin{cases} -k^{(i)} & \text{if} \quad i \in I \\ f^{(i)} & \text{if} \quad i \notin I \end{cases}$$

[5] Λ_{d+1} is indeed defined on our weighted measures.

If $I = \emptyset$ then $\Lambda_{d+1,\mu}(f^{(1)}, ..., f^{(d)}, f^{(d+1)}) \leq \epsilon$ by the assumption. Suppose $I = \{i_1, ..., i_r\} \neq \emptyset$ then

$$
\begin{aligned}
|\Lambda_{d+1,I,\mu}(e^{(1)}, ..., e^{(d)}, f^{(d+1)})| &= \left| \int_{X_1} \cdots \int_{X_{d+1}} f^{(1)} \cdots f^{(d+1)} \cdot \prod_{i \in I} \mathbf{1}_{(\Omega^{(i)})^C} d\mu_{X_1} \cdots d\mu_{X_{d+1}} \right| \\
&\leq \mathbb{E}_{\mathbf{x}_{[d+1]}} \prod_{I \subseteq [d+1], |I| \leq d} \nu_I(\mathbf{x}_I) \mathbf{1}_{(\Omega^{(i_1)})^C} \quad \exists i_1 \in I \\
&\leq \frac{1}{T} \mathbb{E}_{\mathbf{x}_{[d+1]}} \mathbb{E}_{\mathbf{y}_{[d+1] \setminus \{i_1\}}} \prod_{I \subseteq [d+1], |I| \leq d} \nu_I(\mathbf{x}_I) \prod_{\substack{\underline{\omega}_I \neq \underline{0} \\ I \subseteq [d+1] \setminus \{i_1\}}} \nu_I(P_{\underline{\omega}_I}(\mathbf{x}_I, \mathbf{y}_I)) \\
&\lesssim \frac{1}{T} \leq \varepsilon
\end{aligned}
$$

by linear form condition.

Step 2 We will show $\Lambda_{d+1,\mu}(g^{(1)}, ..., g^{(d+1)}) \lesssim \epsilon$ if $\eta \leq \eta(\epsilon), N \geq N(\epsilon, \eta)$.

Proof of step 2: Write $g^{(i)} = g^{(i)} + h^{(i)} - h^{(i)} = f^{(i)} \cdot \mathbf{1}_{\Omega^{(i)}(T)} - h^{(i)}$ then we have

$$0 \leq f^{(i)} \cdot \mathbf{1}_{\Omega^{(i)}(T)} \leq \nu_i, \|h^{(i)}\|_{\square_\mu^d} \leq \eta$$

so by the weighted von-Neumann inequality and step 1 , we have

$$
\begin{aligned}
|\Lambda_{d+1,\mu}(g^{(1)}, ..., g^{(d+1)})| &= |\Lambda_{d+1,\mu}(g^{(1)} + h^{(1)}, ..., g^{(d+1)} + h^{(d+1)}) - \sum_{e^i = h^i, \exists i} \Lambda_{d+1,\mu}(e^{(1)}, .., e^{(d)}, e^{(d+1)})| \\
&\lesssim \epsilon + \eta + o_{N \to \infty}(1) \\
&\lesssim \epsilon
\end{aligned}
$$

if $1/T \leq \epsilon, \eta \leq \epsilon, N \geq N(\epsilon)$ and the proof of step 2 is completed.

Now since $0 \leq g^{(i)} \leq 2$ then (after normalizing) using the *ordinary hypergraph removal lemma*(Theorem 5.1), we have

$$F_i \subseteq X_{[d+1] \setminus \{i\}} \quad \text{such that} \quad \prod_{1 \leq k \leq d+1} \mathbf{1}_{F_k} \equiv 0 \quad \text{and}$$

$$\int_{X_1} \cdots \int_{X_{d+1}} g^{(i)} \cdot \mathbf{1}_{F_i^C} \, d\mu_{X_1} \cdots d\mu_{X_d+1} \lesssim \delta(\epsilon)$$

so

$$
\begin{aligned}
\int_{X_1} \cdots \int_{X_{d+1}} f^{(i)} \cdot \mathbf{1}_{F_i^C} \, d\mu_{X_1} \cdots d\mu_{X_{d+1}} \lesssim \delta(\epsilon) &+ \underbrace{\int_{X_1} \cdots \int_{X_{d+1}} h^{(i)} \cdot \mathbf{1}_{F_i^C} \, d\mu_{X_1} \cdots d\mu_{X_d} d\mu_{X_{d+1}}}_{(A)} + \\
&+ \underbrace{\int_{X_1} \cdots \int_{X_{d+1}} f^{(i)} \cdot \mathbf{1}_{\Omega_i^C(T)} \mathbf{1}_{F_i^C} \, d\mu_{X_1} \cdots d\mu_{X_{d+1}}}_{(B)}
\end{aligned}
$$

Now for our purpose, it suffices to show $(A), (B) \lesssim \epsilon$.

Estimate for (A) (error from uniformity function): By the assumption of complexity of $\sigma-$algebras, the function $\mathbf{1}_{F_i^C}$ could be written as a sum of $O_M(1)$ of functions of the form $\prod_{j\in[d+1]\setminus\{i\}} v_j^{(i)}$ where $v_j^{(i)}$ is a $[0,1]$- valued function in $\mathbf{x}_{[d+1]\setminus\{i,j\}}$. We could write estimate each term with $\prod_{j\in[d+1]\setminus\{i\}} v_j^{(i)}$ individually. Applying Cauchy-Schwartz's inequality d times to estimate the expression (A) (here let's assume $i<d$, the case $i=d$ is the same.) :

$$
\begin{aligned}
&\left(\int_{X_1}\cdots\int_{X_d}\int_{X_{d+1}} h^{(i)}\cdot \mathbf{1}_{F_i^C}\, d\mu_{X_1}\cdots d\mu_{X_d} d\mu_{X_{d+1}}\right)^{2^d}\\
&\lesssim \left[\left(\int_{X_1}\cdots\int_{X_d}\left(\int_{X_{d+1}} h^{(i)}\prod_{\substack{1\le j\le d\\ j\neq i}} u_j^{(i)} d\mu_{X_{d+1}}\right)u_{d+1}^i d\mu_{X_1}\cdots d\mu_{X_d} d\mu_{X_{d+1}}\right)^2\right]^{2^{d-1}}\\
&\le \Bigg[\int_{X_1}\cdots\int_{X_d}\left(\int_{X_{d+1}} h^{(i)}\prod_{\substack{1\le j\le d-1\\ j\neq i}} u_j^{(i)} d\mu_{X_{d+1}}\right)^2 d\mu_{X_1}\cdots d\mu_{X_d}\\
&\quad\times\int_{X_1}\cdots\int_{X_d}(u_{d+1}^i)^2 d\mu_{X_1}\cdots d\mu_{X_d}\Bigg]^{2^{d-1}}\\
&\lesssim \Bigg[\int_{X_1}\cdots\int_{X_d}\int_{X_{d+1}}\int_{Y_{d+1}} h^{(i)}(\mathbf{x}_{[d+1]\setminus\{i\}}, x_{d+1})h^{(i)}(\mathbf{x}_{[d]\setminus\{i\}}, y_{d+1})\\
&\prod_{\substack{1\le j\le d\\ j\neq i}} u_j^{(i)}(\mathbf{x}_{[d]\setminus\{i\}}, x_{d+1})u_j^{(i)}(\mathbf{x}_{[d]\setminus\{i\}}, y_{d+1}) d\mu_{X_1}\cdots d\mu_{X_{d+1}} d\mu_{Y_{d+1}}\Bigg]^{2^{d-1}}
\end{aligned}
$$

Continue applying Cauchy-Schwartz's inequality this way. After d application of Cauchy-Schwartz's inequality, the positive function $u_j^{(i)}$ eventually disappears and we have this bounded by $\|h^{(i)}\|_{\square_\mu}^{2^d}\le\epsilon$.

Estimate for (B): Next we estimate the expression in (B),

$$
\begin{aligned}
&\left|\int_{X_1}\cdots\int_{X_{d+1}} f^{(i)}\cdot\mathbf{1}_{(\Omega^{(i)}(T))^C}\cdot\mathbf{1}_{F_i^C} d\mu_{X_1}\cdots d\mu_{X_{d+1}}\right|\\
&\le \int_{X_1}\cdots\int_{X_{d+1}}(\nu_{[d+1]\setminus\{i\}})\cdot\mathbf{1}_{(\Omega^{(i)}(T))^C} d\mu_{X_1}\cdots d\mu_{X_{d+1}}\\
&\le \frac{1}{T}\mathbb{E}_{\mathbf{x}_{[d+1]}}\mathbb{E}_{\mathbf{y}_{[d+1]\setminus\{i\}}}\nu_{[d+1]\setminus\{i\}}(\mathbf{x}_{[d+1]\setminus\{i\}})\prod_{|I|\le d}\nu_I(\mathbf{x}_I)\prod_{\substack{\underline{\omega}_I\neq\underline{0}\\ I\subseteq[d+1]\setminus\{i\}}}\nu_I(P_{\underline{\omega}_I}(\mathbf{x}_I,\mathbf{y}_I))\\
&\lesssim \frac{1}{T},
\end{aligned}
$$

by the linear forms condition. Hence if we choose sufficiently large T then

$$\int_{X_1} \cdots \int_{X_{d+1}} f^{(i)} \cdot \mathbf{1}_{F_i^C} d\mu_{X_1} \cdots d\mu_{X_{d+1}} \lesssim \delta(\epsilon).$$

$\square$

3.6 Proof of the Main Result

3.6.1 From $\mathbb{Z}_N$ to $\mathbb{Z}$

First, recall that $\nu_{\varepsilon_1,\varepsilon_2}(n) \approx \frac{\phi(W)}{W} \log N, \varepsilon_1 N \le n \le \varepsilon_2 N, \varepsilon_1, \varepsilon_2 \in (0,1]$ for a sufficiently large prime N in the residue class $b \pmod W$. Also Lemma 4.7.1 allows us to work in $(\mathbb{Z}/N')^d$ for some big prime N'. By pigeonhole principle (see Lemma 4.7.2 in Chapter 4) we may choose $b \in (\mathbb{Z}/W)^d$ and small $\varepsilon_1, \varepsilon_2 > 0$ in the definition of ν and A' such that

$$|A'| := |\{n \in [1, N/W]^d;\ Wn + b \in A\} \cap [\varepsilon_1 N', \varepsilon_2 N']^d| \ge \frac{\alpha\, \varepsilon_2^d}{2} \frac{(N')^d\, W^d}{(\log N')^d\, \phi(W)^d}.$$

Here we choose N' so that $\varepsilon_2 N' = N/W(1 + o_{N,W\to\infty}(1))$.

3.6.2 Proof of the Main Theorem

To prove the theorem, suppose on the contrary that A' contains less than $\epsilon \frac{N'^{d+1}}{(\log N')^{2d}}$ corners.($\epsilon = c(\alpha)$) then

$$\begin{aligned}
&\Lambda_{d+1,\mu}(f^{(1)}, ..., f^{(d+1)}) \\
&= (N')^{-(d+1)} \sum_{\mathbf{x}_{[d+1]}} \prod_{1\le i\le d} \mathbf{1}_{A'}(x_1, ..., x_{i-1}, x_{d+1} - \sum_{\substack{1\le j\le d \\ j\ne i}}, x_{i+1}, ..., x_d)\nu_I \mathbf{1}_{A'}(x_1, ..., x_d) \cdot \nu(x_1)...\nu(x_d) \\
&\le \frac{1}{N'^{d+1}} \sum_{\substack{p_i \in A', 1\le i\le 2d \\ \text{that consitutes a corner}}} \prod_{1\le k\le d} 1_{A'}(p_1, ..., p_{k-1}, p_{d+k}, p_{k+1}, ..., p_d)\mathbf{1}_A(p_1, ..., p_d)\nu(p_1)...\nu(p_{2d}) \\
&\lesssim \frac{1}{N'^{d+1}} \left(\frac{\phi(W)\log N'}{W}\right)^{2d} \times (\text{The number of corners in } A') \\
&\le \epsilon
\end{aligned}$$

Now assume that $\Lambda_{d+1,\mu}(f^{(1)}, ..., f^{(d)}, f^{(d+1)}) \lesssim \epsilon$ then by the relative hypergraph removal lemma

$$\exists E_i, 1 \le i \le d+1, E_i \subseteq X_{[d+1]\setminus\{i\}} := \tilde{X}_i,$$

such that

$$\prod_{1\le i\le d+1} \mathbf{1}_{E_i} \equiv 0, \quad \int_{\tilde{X}_i} f^{(i)} \mathbf{1}_{E_i^C} d\mu_{\tilde{X}_i} \lesssim \delta(\epsilon)$$

where $\delta(\epsilon) \to 0$ as $\epsilon \to 0$. Let $A' = A \cap [\delta_1 N, \delta_2 N]^d$, $z = \sum_{1\le j\le d} x_j$, $g_{A'} := g \cdot \mathbf{1}_{A'}$ for any function g then

$$\begin{aligned}
\tilde{\Lambda} &:= N'^{-d} \sum_{(x_1,\dots,x_d)\in A'} f^{(1)}_{A'}(x_2,...,x_d,z) f^{(2)}_{A'}(x_1,x_3,...,x_d,z)...f^{(d)}_{A'}(x_1,x_2,...,x_{d-1},z) f^{(d+1)}_{A'}(x_1,...,x_d) \\
&\ge N'^{-d} \sum_{(x_1,\dots,x_d)\in A'} \nu(x_1)...\nu(x_d) \\
&\gtrsim (N')^{-d} \big(\frac{\phi(W)}{W} \log N'\big)^d \cdot \frac{\alpha \cdot (N'W)^d}{(\phi(W)\log N')^d} = \alpha.
\end{aligned}$$

for arbitrarily large N'. Now

$$\tilde{\Lambda} = \mathbb{E}_{\mathbf{x}_{[d]}}(f^{(1)}_{A'}\mathbf{1}_{E_1} + f^{(1)}_{A'}\mathbf{1}_{E_1^C})...(f^{(d+1)}_{A'}\mathbf{1}_{E_d+1} + f^{(d+1)}_{A'}\mathbf{1}_{E_{d+1}^C})$$

Now we have by the assumption $\mathbb{E}_{\mathbf{x}_{[d]}} f^{(1)}_{A'} \cdot \mathbf{1}_{E_1} ... f^{(d+1)}_{A'} \cdot \mathbf{1}_{E_{d+1}} \equiv 0$ so we just need to estimate each other term individually.
Consider $\mathbb{E}_{\mathbf{x}_{[d]}} f^{(1)}_{A'} \cdot \mathbf{1}_{E_1^C} f^{(2)}_{A'} \cdot \mathbf{1}_{E_2^\pm} ... f^{(d+1)}_{A'} \cdot \mathbf{1}_{E_{d+1}^\pm}$, where $F^\pm$ can be either F or F^C for any set F.
Now since

$$0 \le f^{(j)}_{A'}\mathbf{1}_{E_j^\pm} \le \nu(x_j), d \ge j \ge 2 \quad \text{and} \quad 0 \le f^{(d+1)}_{A} \le 1$$

We have

$$\mathbb{E}_{\mathbf{x}_{[d]}} f^{(1)}_{A'} \cdot \mathbf{1}_{E_1^C} f^{(2)}_{A'} \cdot \mathbf{1}_{E_1^\pm} ... f^{(d+1)}_{A'} \cdot \mathbf{1}_{E_{d+1}^\pm} \le \mathbb{E}_{\mathbf{x}_{[d]}} f^{(1)}_{A'} \cdot \mathbf{1}_{E_1^C} \nu(x_2)...\nu(x_d) = \int_{\tilde{X}_1} f^{(1)} \cdot \mathbf{1}_{E_1^C} d\mu_{X_2} \cdots d\mu_{X_{d+1}} \lesssim \delta(\epsilon).$$

In the same way, we have for any $1 \le i \le d+1$,

$$\mathbb{E}_{\mathbf{x}_{[d]}} f^{(i)}_{A'} \cdot \mathbf{1}_{E_i^C} \prod_{1\le j\le d+1, j\ne i} (f^{(j)} \cdot \mathbf{1}_{E_j^\pm}) \lesssim \delta(\epsilon)$$

So if $N' > N(\alpha)$ then

$$\mathbb{E}_{\mathbf{x}_{[d]}} f^{(1)}_{A'}(x_2,...,x_d,u) f^{(2)}_{A'}(x_1,x_3,...,x_d,u)...f^{(d)}_{A'}(x_1,...,x_{d-1},u) f^{(d+1)}_{A'}(x_1,...,x_d) \lesssim \delta(\epsilon) = o(\alpha)$$

This is a contradiction. Hence there are $\gtrsim \epsilon \frac{N'^{d+1}}{(\log N')^{2d}}$ corners in A. Note that the number of degenerated corners is at most $O(\frac{N'^d}{(\log N')^d})$ as the corner is degenerated (and will be degenerated into a single point) iff $z = \sum_{1\le j\le d} x_j$.

3.7 Further Remarks: Conlon-Fox-Zhao's Densification Trick

One important feature that make the transference principle work is that we are working on a $d-$regular hypergraph, this is also the same as in [52] , [16]. To prove a more general version of the theorem, we have to consider more general version of hypergraph removal lemma, this will be discussed in the next chapter.

A natural question in our method is that if the correlation condition is needed in our proof. In the setting of Gaussian prime, the correlation condition was needed to deal with the fact that if p is a Gaussian prime then its conjugate. $\overline{p}$ is also a Gaussian prime. Later, Conlon-Fox-Zhao [16] developed the "densification technique" to prove a relatively Szemerédi's Theorem like Theorem 1.2.1 but with ν satisfies only certain linear form conditions. Densification trick allows one to replace a sparse edge with a dense edge. No correlation conditions or bounded dual conditions are assumed as this technique allows most factors in correlation to be bounded. The question of simplifying the pseudorandom conditions is an interesting and active research questions. Gowers [37] asked if $\|\nu - 1\|_{U^s} = o(1)$ for some large $s = s(k)$ would allow us to deduce a relative $k - AP$ Szemereédi's theorem.

Definition 3.7.1 (H_r-linear forms condition)**.** *Consider a weight hypergraph system (J, V_j, H_r) with weight system $\nu = \{\nu_e\}_{e \in \mathcal{H}_r}$. We say that ν satisfies H_r-linear forms condition if*

$$\mathbb{E}_{x,y \in V_J} \prod_{e \in H_r} \prod_{\omega_e} \nu_e(P_{\omega_e}(x|_e, y|_e))^{n_{e,w}} = 1 + o(1)$$

Hence this linear forms condition is about counting the 2-blow up of $\mathcal{H}_r$ and any subgraph of this blow-up.

This linear forms condition is the same that is used in [102] (Def. 2.8). However in [102], one assumes the correlation condition and the bounded dual condition.

Theorem 3.7.2. *[16] If $S \subseteq \mathbb{Z}_N, \nu = \frac{N}{|S|}\mathbf{1}_S$ satisfies*

$$\mathbb{E}_{x,x',y,y',z,z'} \nu(x)\nu(x')\nu(z-x)\nu(z'-x)\nu(z-x')\nu(z'-x')\nu(y)\nu(y')\nu(z-y)\nu(z'-y)\nu(z-y')\nu(z'-y') = 1+o(1).$$

or if the condition holds if one of the ν above is replaced by 1. Then a corner-free subset of S has size $o(|S|^2)$

The densification trick allows them to prove the counting lemma with only such linear forms condition. This has application in simplifying many technical difficulties in previous results and one can obtain a better quantitative result such as primes with narrow polynomial progressions [112]: $a + P_1(r), a + P_2(r), \ldots, a + P_k(r)$ with $r \leq \log^L N$. We want to transfer the count in (pseudoransom) sparse setting to dense setting.

We may try to transfer some this kind of the theorems similar to corners (for example, we may model using $d-$regular hypergraph) we know in integer case to the prime case. An interesting problem would be to find an analogue of Shkredov's result [96] that is to obtain an exponential bound for corners in dense subsets of $\mathbb{P}[N]^2$.

Chapter 4

Weighted Simplices Removal Lemma and Multidimensional Szemerédi's Theorem in the Primes

4.1 Introduction

The main objective of this chapter is to prove the following generalization of the main result in chapter 3.

Theorem 4.1.1. *If A is a subset of $\mathcal{P}^d$ of positive upper relative density, then A contains infinitely many non-trivial affine copies of any finite set $F \subseteq \mathbb{Z}^d$.*

Note that it is enough to show that the set A contains at least one non-trivial affine copy of F, as deleting the set F from A will not affect its relative density. Also, replacing the set F by $F' = F \cup (-F)$ one can require that the dilation parameter t is positive.

By lifting the problem to a higher number of dimensions, it is easy to see that one can assume that F forms the vertices of a d-dimensional simplex[1] (which will be important as the linear forms appeared in the parametrization is pairwise linearly independent). Indeed, let $F = \{0, x_1, \dots, x_k\}$, choose a set of k linearly independent vectors $\{y_1, \dots, y_k\} \subseteq \mathbb{Z}^k$, and define the set $\Delta := \{0, (x_1, y_1), \dots, (x_k, y_k), z_{k+1}, \dots, z_{k+d}$ $\mathbb{Z}^{k+d}$ such that the vectors of $\Delta \backslash \{\mathbf{0}\}$ form a basis of $\mathbb{R}^{k+d}$. If the set $A' = A \times \mathcal{P}^k$ contains an affine copy of Δ then clearly A contains an affine copy of the set $\pi(\Delta) \supseteq F$, where $\pi : \mathbb{R}^d \times \mathbb{R}^k \to \mathbb{R}^d$ is the natural orthogonal projection.

[1]Unike integer case, it is not enough to lift to a corner as we don't know if the projection used there will project prime corners to prime points or not.

In the case when $\Delta = \{v_0, v_1, \ldots, v_d\} \subseteq \mathbb{Z}^d$ is a d-dimensional simplex, i.e. $v_1 - v_0, v_2 - v_0, \ldots, v_d - v_0$ are pairwise linearly independent, we prove a quantitative version of Theorem 4.1.1. To formulate it we define the quantity

$$l(\Delta) := \sum_{i=1}^{d} |\pi_i(\Delta)|, \tag{4.1.1}$$

$\pi_i : \mathbb{R}^d \to \mathbb{R}$ being the orthogonal projection to the i-th coordinate axis.

Theorem 4.1.2. *Let $\alpha > 0$ and let $\Delta \subseteq \mathbb{Z}^d$ be a d-dimensional simplex. There exists a constant $c(\alpha, \Delta) > 0$ such that for any $N > 1$ and any set $A \subseteq \mathcal{P}_N^d$ such that $|A| \geq \alpha |\mathcal{P}_N|^d$, the set A contains at least $c(\alpha, \Delta) N^{d+1} (\log\, N)^{-l(\Delta)}$ affine copies of the simplex Δ.*

The lower bound matches with the bound from the heuristic argument that primes is a random subset of $\mathbb{Z}^d$ with density $1/\log^d N$ in $[1, N]^d$: there are $\approx N^{d+1}$ affine copies $x + t\Delta$ of Δ in $[1, N]^d$, and for a fixed i the probability that all the i-th coordinates of an affine copy Δ are primes is roughly $(\log N)^{-|\pi_i(\Delta)|}$. Thus if the prime tuples behave randomly, the probability that $\Delta \subseteq \mathcal{P}^d$ is about $(\log N)^{-l(E)}$.

Note that in Theorem 4.1.2 we do not require the copies of Δ to be non-trivial, thus without loss of generality, N can be assumed to be sufficiently large with respect to α and Δ. It is clear that Theorem 4.1.2 implies Theorem 4.1.1 as the number of trivial copies of Δ in A (i.e. the one with $t = 0$) is at most $N^d (\log\, N)^{-d}$.

In the contrapositive, Theorem 4.1.2 states that if a set $A \subseteq \mathcal{P}_N^d$ contains at most $\delta N^{d+1} (\log N)^{-l(\Delta)}$ affine copies of Δ, then its relative density is at most ϵ, where $\epsilon = \epsilon(\delta)$ is a quantity such that $\epsilon(\delta) \to 0$ as $\delta \to 0$. As for a number of similar results on prime configurations [51], [102], [45].
Thus identifying $[1, N]$ with $\mathbb{Z}/N\mathbb{Z}$ it is easy to show that Theorem 4.1.2 follows from

Theorem 4.1.3. *Let $\Delta = \{v_0, \ldots, v_d\} \subseteq \mathbb{Z}^d$ be a d-dimensional simplex and let $\delta > 0$. Let N be a large prime and let $A \subseteq \mathbb{Z}_N^d$ such that*

$$\mathbb{E}_{x \in \mathbb{Z}_N^d, t \in \mathbb{Z}_N} \left(\prod_{i=0}^{d} \mathbf{1}_A(x + tv_i) \right) w(x + t\Delta) \leq \delta \tag{4.1.2}$$

then there exists $\epsilon = \epsilon(\delta)$ such that

$$\mathbb{E}_{x \in \mathbb{Z}_N^d} \mathbf{1}_A(x) w(x) \leq \epsilon(\delta) + o_{N, W \to \infty;\, \Delta}(1)$$

Moreover $\epsilon(\delta) \to 0$ as $\delta \to 0$.

The objective of this chapter is to prove this theorem. As in Chapter 3, our result would follow if we could prove the following version of simplices removal lemma (i.e. Lemma 4.1.7 below). Notice that the conclusion of the lemma does not hold in the same measure space but on a new one which is a small perturbation of the original measure space.

We define weighted system of hypergraph as in section 3.1 of Chapter 3 but now we will do energy increment so we will put sigma-algebras on our hypergraph. Hypergraph system can be considered an as analogue of measure preserving system in ergodic theory.

We will use the construction of a weighted hypergraph associated to a set $A \subseteq \mathbb{Z}_N^d$ and a simplex $\Delta = \{v_0, \dots, v_d\}$ given in the case of Gaussian Primes [102].

Definition 4.1.4. *[Hypergraph System] Let $J = \{0, 1, \dots, d\}, \overline{\mathcal{H}} := \{e : e \subseteq J\}$ be the set of all possible hyperedges, and for a set $e \in \overline{\mathcal{H}}$, let $V_e = \mathbb{Z}_N^e = \prod_{j\in e} \mathbb{Z}_N$. Identify V_e as the subspace of elements $x = (x_0, \dots, x_d) \in V_J$ such that $x_j = 0$ for all $j \notin e$ and let $\pi_e : V_J \to V_e$ denote the natural projection. For $e = \{j\}$ we write $V_j := V_{\{j\}}$ and for a given $\mathcal{H} \subseteq \overline{\mathcal{H}}$, we will call the quadruplet $(J, V_J, \mathcal{H})$ a **hypergraph system**.*
For each positive integers j denote $\mathcal{H}_j := \{e \in \mathcal{H}; |e| = j\}$. For $e \in \mathcal{H}, x_e = (x_j)_{j\in e}$.
For a given $e \subseteq J$ and a collection of sets (edges) on V_e, define $\mathcal{A}_e = \{\pi_e^{-1}(F) : F \subseteq V_e\}$ considered as corresponding sets on V_J.

Remark 4.1.5. *For convenience, we identify V_e as a subset of V_J as the set of points in V_J where the coordinates in $J\backslash e$ are allowed to be all possible values (i.e. no restrictions on $J\backslash e$). Hence we work on a single ambient space.*

Remark 4.1.6. *We can think of a point x_e, $e \in \mathcal{H}_d$ as a d-simplex with vertices $\{x_j : j \in e\}$. A set $G_e \subseteq V_e$ then may be viewed as a d-regular d-partite hypergraph with vertex sets V_j $(j \in e)$. Similarly a point $x \in V_J$ represents a $(d + 1)$-simplex with $d-$faces $x_e, e \in \mathcal{H}_d$.*

Theorem 4.1.7. *(Weighted Simplex Removal Lemma) Let $\{\nu_e\}_{e\subseteq J}, \{\mu_e\}_{e\subseteq J}$ be a system of weights and measures associated to a well-defined, pairwise linearly independent and symmetric family of linear forms $\mathcal{L}$ (as defined in (4.2.6)). Let $E_e \in \mathcal{A}_e, g_e : V_e \to [0, 1]$ be given for each $e \in \mathcal{H}_d$. Then for a given $\delta > 0$ there exists an $\epsilon = \epsilon(\delta) > 0$ such that the following holds: If*

$$\mathbb{E}_{x\in V_J} \prod_{e\in\mathcal{H}_d} \mathbf{1}_{E_e}(x)\mu_J(x) \leq \delta \tag{4.1.3}$$

then there exists a well-defined, symmetric family of linear forms $\tilde{L} = \{\tilde{L}_e^k; e \in \mathcal{H}_d, 1 \leq k \leq d\}$, such that the associated system of weights and measures $\{\tilde{\nu}_e\}_{e\subseteq J}, \{\tilde{\mu}_e\}_{e\subseteq J}$ satisfy

$$\mathbb{E}_{x\in V_J} \prod_{e\in\mathcal{H}_d} \mathbf{1}_{E_e}(x)\tilde{\mu}_J(x) = \mathbb{E}_{x\in V_J} \prod_{e\in\mathcal{H}_d} \mathbf{1}_{E_e}(x)\mu_J(x) + o_{N,W\to\infty}(1) \tag{4.1.4}$$

and for all $e \in \mathcal{H}_d$,

$$\mathbb{E}_{x\in V_e}\, g_e(x)\tilde{\mu}_e(x) = \mathbb{E}_{x\in V_e}\, g_e(x)\mu_e(x) + o_{N,W\to\infty}(1) \tag{4.1.5}$$

In addition there exist sets $E'_e \in \mathcal{A}_e$ such that

$$\bigcap_{e\in\mathcal{H}_d} (E_e \cap E'_e) = \emptyset \tag{4.1.6}$$

and for all $e \in \mathcal{H}_d$

$$\mathbb{E}_{x\in V_e} \mathbf{1}_{E_e\setminus E'_e}(x)\tilde{\mu}_e(x) \le \epsilon(\delta) + o_{N,W\to\infty}(1) \tag{4.1.7}$$

moreover

$$\epsilon(\delta) \to 0, \ \ \textit{as} \ \ \delta \to 0. \tag{4.1.8}$$

Roughly speaking, if $\mu(\cap_{e\in\mathcal{H}_d} E_e) < \delta$ is small then we can modify each E_e slightly in the magnitude of $\varepsilon(\delta)$ to obtain E'_e such that $\bigcap_{e\in\mathcal{H}_d} (E_e \cap E'_e) = \emptyset$ i.e. $\bigcap_{e\in\mathcal{H}_d} E_e \subseteq \bigcup_{e\in\mathcal{H}_d} E_e\setminus E'_e$ meaning $\bigcap_{e\in\mathcal{H}_d} E_e$ actually has a very smaller measure than expected.

4.1.1 Parametric Weight System

Recall the weight version of $\Box^2_\mu$ norm. Let $f : X\times Y \to \mathbb{R}$, $\nu_{12}(x_1,x_2) = \nu(x_1,x_2)$.

$$\begin{aligned}\|f\|^4_{\Box^2_\mu} &= \mathbb{E}_{p\in X\times Y}\mathbb{E}_{x\in X\times Y} f(x_1,x_2)f(p_1,x_2)f(x_1,p_2)f(p_1,p_2)\\ &\quad\times \nu(x_1,x_2)\nu(p_1,x_2)\nu(x_1,p_2)\nu(p_1,p_2)\nu(x_1)\nu(x_2)\nu(p_1)\nu(p_2)\end{aligned}$$

If it is possible, we would like to proceed as in previous chapter with transference principle; writing the weight appearing above in a form of direct product of measures, say

$$d\mu(x)d\mu(p)$$

with, say, $d\mu(x) = \nu(x_1,x_2)\nu(x_1)\nu(x_2)$. However this is not possible due to cross terms like $\nu(p_1,x_2), \nu(x_1,p_2)$. More importantly, in higher order box norm, we would have weight in intermediate order. Prohibiting us from working on regular hypergraph and application of transference principle as in chapter 3.

The way we get around these difficulties is that we will reprove the removal lemma in our setting, without using transference principle. To accomplish this, we need to introduce the parametric weight system. We could write

$$\begin{aligned}\|f\|_{\square^2_\mu}^4 &= \mathbb{E}_{p\in X\times Y}\mathbb{E}_{x\in X\times Y} f(x_1,x_2)f(p_1,x_2)f(x_1,p_2)f(p_1,p_2)\\ &\quad\times \nu(x_1,x_2)\nu(p_1,x_2)\nu(x_1,p_2)\nu(x_1)\nu(x_2)\nu(p_1,p_2)\nu(p_1)\nu(p_2)\\ &= \mathbb{E}_{p\in X\times Y}\mathbb{E}_{x\in X\times Y} f(x_1,x_2)f(p_1,x_2)f(x_1,p_2)f(p_1,p_2)d\mu_p(x_1,x_2)d\mu(p)\end{aligned}$$

Here for each fixed p, $d\mu_p(x_1,x_2)=\nu(x_1,x_2)\nu(p_1,x_2)\nu(x_1,p_2)\nu(x_1)\nu(x_2)$ is a measure in x_1,x_2-variables depending on p which we regard as parameters. Later, we will regard μ_p as a parametric extension of μ in the sense if we consider linear forms defining the measures as linear forms in p and x variables then linear forms defining μ_p is the same as those defining μ.

Example 4.1.8. *Consider the measure space* $\big(X_1\times X_2, d\mu(x_1,x_2)=\nu(x_1+2x_2)\nu(x_2)\big)$ *then*

$$\begin{aligned}\|f\|_{\square^2_\mu}^4 &= \mathbb{E}_{p\in X_1\times X_2}\mathbb{E}_{x\in X_1\times X_2} f(x_1,x_2)f(p_1,x_2)f(x_1,p_2)f(p_1,p_2)\\ &\quad\times \nu(x_1+2x_2)\nu(p_1+2x_2)\nu(x_1+2p_2)\nu(x_1)\nu(x_2)\nu(p_1+2p_2)\nu(p_1)\nu(p_2).\end{aligned}$$

Working on μ_p would be hard as linear forms conditions may not apply. But if we average $d\mu_p$ over all parameters p then linear form conditions do apply. In this regard, μ_p itself may not be pseudorandom but upon averaging, there should be many μ_p that are pseudorandom. Also we will prove that most μ_p is only a small perturbation of μ and still share many properties with μ.

4.1.2 Energy Increment in weighted setting.

Assume that there is an edge e, say $e=(1,2)$, so that the graph $G_e=\pi_e(E_e)$ is not ε-regular. This means

$$\|F\|_{\square_{\mu_e}} \geq \varepsilon, \tag{4.1.9}$$

where $F: V_1\times V_2\to\mathbb{R}, F=\mathbf{1}_{G_e}-\mu_e(G_e)\,\mathbf{1}_{V_e}$. In view of definition of the weight box norm, we may write

$$\|F\|_{\square_{\mu_e}}^4 = \int_{V_e}\int_{V_e} F(x)\,u_q^1(x_1)u_q^2(x_2)\nu_e(x_1,q_2)\nu_e(q_1,x_2)\,d\mu_e(x)\,d\mu_e(q)\geq\varepsilon^4 \tag{4.1.10}$$

where $x=(x_1,x_2)$, $q=(q_1,q_2)$, we set $u_q^1(x_1)=F(x_1,q_2)$, and $u_q^2(x_2)=F(q_1,x_2)F(q_1,q_2)$. If one defines the measures $\mu_{q,e}$, depending on the parameter q, by

$$\mu_{q,e}(x) := \nu_e(x_1,q_2)\nu_e(q_1,x_2)\mu_e(x),$$

then the inner expression in (4.1.10) can be viewed as the inner product

$$\Gamma(q) := \langle F, u_q^1 \cdot u_q^2 \rangle_{\mu_{q,e}} = \int_{V_e} F(x)\, u_q^1(x_1) u_q^2(x_2)\, d\mu_{q,e}(x), \tag{4.1.11}$$

on the Hilbert space $L^2(V_e, \mu_{q,e})$. Thus (4.1.10) translates to $\mathbb{E}_{q\in V_e}\, \Gamma(q)\, \mu_e(q) \geq \varepsilon^4$ while using the linear forms condition it is easy to see that $\mathbb{E}_{q\in V_e}\, \Gamma(q)^2\, \mu_e(q) \lesssim 1$ thus, by averaging[2],

$$\Gamma(q) \gtrsim \varepsilon^4, \;\; \textit{for} \;\; q \in \Omega, \tag{4.1.12}$$

for a set $\Omega \subseteq V_e$ of measure $\mu_e(\Omega) \gtrsim \varepsilon^8$. As the functions u_q^i are bounded, hence without loss of generality, using Fubini's Theorem[3], we may assume that they are indicator functions of sets $U_q^i \subseteq V_i$. Let $\mathcal{B}_q = \mathcal{B}_q^1 \vee \mathcal{B}_q^2$ denote the σ-algebra on V_e generated by the sets $\pi_i^{-1}(U_q^i)$ $(i = 1, 2)$, and let $\mathbb{E}_{\mu_{q,e}}(\mathbf{1}_{G_e}|\mathcal{B}_q)$ be the conditional expectation function of $\mathbf{1}_{G_e}$ with respect to this σ-algebra and the measure $\mu_{q,e}$. Then, as $u_q^1\, u_q^2$ is measurable with respect to $\mathcal{B}_q$, we have

$$\langle \mathbf{1}_{G_e} - \mathbb{E}_{\mu_{q,e}}(\mathbf{1}_{G_e}|\mathcal{B}_q)\,,\; u_q^1\, u_q^2\rangle_{\mu_{q,e}} = 0.$$

This together with (4.1.11) and (4.1.12) implies for $q \in \Omega$

$$\langle\, \mathbb{E}_{\mu_{q,e}}(\mathbf{1}_{G_e}|\mathcal{B}_q) - \mathbb{E}_{\mu_e}(\mathbf{1}_{G_e}|\mathcal{B}_0)\,,\; u_q^1\, u_q^2\,\rangle_{\mu_{q,e}} \gtrsim \varepsilon^4,$$

where $\mathcal{B}_0 = \{V_e, \emptyset\}$ is the trivial σ-algebra, and $\mathbb{E}_{\mu_e}(\mathbf{1}_{G_e}|\mathcal{B}_0) = \mu_e(G_e)\, \mathbf{1}_{V_e}$. Then by the Cauchy-Schwartz inequality, we have

$$\|\mathbb{E}_{\mu_{q,e}}(\mathbf{1}_{G_e}|\mathcal{B}_q) - \mathbb{E}_{\mu_e}(\mathbf{1}_{G_e}|\mathcal{B}_0)\|^2_{L^2(\mu_{q,e})} \gtrsim \varepsilon^8. \tag{4.1.13}$$

Notice that the condition expectations above are on different measure spaces. To overcome this "discrepancy", using the linear forms condition, we can show that for given $B \subseteq V_e$ one has

$$\mathbb{E}_{q\in V_e}\, |\mu_{q,e}(B) - \mu_e(B)|^2\, \mu_e(q) = o_{N,W\to\infty}(1).$$

This in turn implies that *for almost every*[4] q,

$$\|\mathbb{E}_{\mu_{q,e}}(\mathbf{1}_{G_e}|\mathcal{B}_0) - \mathbb{E}_{\mu_e}(\mathbf{1}_{G_e}|\mathcal{B}_0)\|_{L^2(\mu_{q,e})} = o_{N,W\to\infty}(1) \tag{4.1.14}$$

and

$$\|\mathbb{E}_{\mu_e}(\mathbf{1}_{G_e}|\mathcal{B}_0)\|_{L^2(\mu_e)} = \|\mathbb{E}_{\mu_{q,e}}(\mathbf{1}_{G_e}|\mathcal{B}_0)\|_{L^2(\mu_{q,e})} + o_{N,W\to\infty}(1). \tag{4.1.15}$$

[2]see e.g. arguments in the proof of Lemma 4.4.1.
[3]see e.g. calculations after (4.4.13).
[4]We will put a measure on the parametric space.

By (4.1.14) and triangle inequality, we have

$$\|\mathbb{E}_{\mu_{q,e}}(\mathbf{1}_{G_e}|\mathcal{B}_q) - \mathbb{E}_{\mu_e}(\mathbf{1}_{G_e}|\mathcal{B}_0)\|_{L^2(\mu_{q,e})} = \|\mathbb{E}_{\mu_{q,e}}(\mathbf{1}_{G_e}|\mathcal{B}_q) - \mathbb{E}_{\mu_{q,e}}(\mathbf{1}_{G_e}|\mathcal{B}_0)\|_{L^2(\mu_{q,e})} + o_{N,W\to\infty}(1)$$

Now by the Pythagoras theorem, one would obtain the "energy increment"

$$\|\mathbb{E}_{\mu_{q,e}}(\mathbf{1}_{G_e}|\mathcal{B}_q) - \mathbb{E}_{\mu_{q,e}}(\mathbf{1}_{G_e}|\mathcal{B}_0)\|^2_{L^2(\mu_{q,e})} = \|\mathbb{E}_{\mu_{q,e}}(\mathbf{1}_{G_e}|\mathcal{B}_q)\|^2_{L^2(\mu_{q,e})} - \|\mathbb{E}_{\mu_{q,e}}(\mathbf{1}_{G_e}|\mathcal{B}_0)\|^2_{L^2(\mu_{q,e})} \gtrsim \varepsilon^8. \tag{4.1.16}$$

(4.1.13), (4.1.15) and (4.1.16) give us that for almost every $q \in \Omega$, that

$$\|\mathbb{E}_{\mu_{q,e}}(\mathbf{1}_{G_e}|\mathcal{B}_q)\|^2_{L^2(\mu_{q,e})} \geq \|\mathbb{E}_{\mu_e}(\mathbf{1}_{G_e}|\mathcal{B}_0)\|^2_{L^2(\mu_e)} + c\,\varepsilon^8. \tag{4.1.17}$$

If $F : V \to \mathbb{R}$ is a function and $(V, \mathcal{B}, \mu)$ is a measure space, recall that the quantity $\|\mathbb{E}_\mu(F|\mathcal{B})\|^2_{L^2(\mu)}$ is referred to as the "energy" of the function F with respect to the measure space $(V, \mathcal{B}, \mu)$, so (4.1.17) is telling that if G_e is not ε-uniform with respect to the initial measure spaces $(V_e, \mathcal{B}_0, \mu_e)$ then its energy increases by *a fixed amount* when passing to the measure spaces $(V_e, \mathcal{B}_{q,e}, \mu_{q,e})$ for (almost) every $q \in \Omega$. One can iterate this argument to arrive to a family of measure spaces $(V_e, \mathcal{B}_{q,e}, \mu_{q,e})_{e\in\mathcal{H}_d,\, q\in\Omega}$ such that the atoms $G_{q,e} \in \mathcal{B}_{q,e}$ become sufficiently uniform, thus obtaining a parametric version of the so-called Koopman- von Neumann decomposition. This can be further iterated to eventually obtain a regularity lemma.

Remark 4.1.9. *The number of linear forms defining the measures $\mu_{q,e}$ is increasing at each step of the iteration, causing the linear forms condition to be used at a level depending eventually on the relative density of the set A and not just on the dimension d. (This can me made independent of α in dimension 1 in Colon-Fox-Zhao's arguments [16].)*

4.2 Weighted Hypergraph System

For a finite set $S \subseteq \mathbb{Z}^d$ we attach the weight

$$w(S) := \prod_{i=1}^{d} \prod_{y\in\pi_i(S)} \nu(y) \tag{4.2.1}$$

where $\pi_i(S)$ is the canonical projection of S to the i-th coordinate axis. If $S = \{x\}$ we write $w(x) := w(\{x\}) = \prod_{i=1}^{d} \nu(x_i)$. As in previous discussion, the weight ν is used to count configurations with prime coordinates. If $Wx + b \in \mathcal{P}_N^d$ (and $x \in [\varepsilon_1 N, \varepsilon_2 N]^d$), then

$$w(x) \approx_{d,W} (\log N)^d. \tag{4.2.2}$$

The implicit constant depends only on d and W which we will choose W sufficiently large but independent of N.

In particular, for $\Delta \subseteq [\varepsilon_1 N, \varepsilon_2 N]^d$ such that $W\Delta + b \subseteq A \subseteq \mathcal{P}_N^d$ one has

$$w(\Delta) \approx (\log N)^{l(\Delta)}. \tag{4.2.3}$$

Recall the definition of hypergraph system (Definition 4.1.4). For a given set $A \subseteq \mathbb{Z}_N^d$ and for $e = J\backslash\{j\}$, let

$$E_e = \{x \in V_J : \sum_{i=0}^{d} x_i(v_i - v_j) \in A\} \tag{4.2.4}$$

Note that $E_e \in \mathcal{A}_e$ for the collection $\mathcal{A}_e$ defined in Definition 4.1.4, as the expression in (4.2.4) is independent of the coordinate x_j. A point $x \in E_e \subseteq V_e$ represents a vertex of an affine copy in A of the simplex. A point $x \in \bigcap_{e\in\mathcal{H}_d} E_e$ represents an affine copy in A of the simplex.

Definition 4.2.1 (Weighted system)**.** *We will define now a family of functions $\nu_e : V_J \to \mathbb{R}_+, \mu_e : V_J \to \mathbb{R}_+$. For $e \in \mathcal{H}_d, e = J\backslash\{j\}$ and $1 \le k \le d$. Write each vertex as*

$$v_j = (v_j^1, v_j^2, \ldots, v_j^d)$$

where v_i^k denotes the $k^{th}-$coordinate of the vector v_i. Define

$$L_e^k(x) = \sum_{i=0}^{d} x_i(v_i^k - v_j^k) \tag{4.2.5}$$

We partition the family of forms

$$\mathcal{L} := \{L_e^k; |e| = d, 1 \le k \le d\} := \bigcup_{f\in\overline{\mathcal{H}}} \overline{\mathcal{L}}_f \tag{4.2.6}$$

according to which coordinates they depend on. Here we write

- *$\overline{\mathcal{L}}_f$ for the set all linear forms (in $x \in V_J$ variables) with variables depend exactly on x_f.*
- *$\mathcal{L}_f$ for the set of linear forms (in $x \in V_J$ variables) depending only on $x_g, g \subseteq f$.*

For this we define the support of a linear form $L(x) = \sum_{k=0}^{d} a_k x_k$ as $supp(L) = \{k : a_k \neq 0\}$. For a given $e \subseteq J$, define

$$\nu_e(x) = \prod_{L\in\mathcal{L}, supp(L)=e} \nu(L(x))\ , \quad \mu_e(x) = \prod_{L\in\mathcal{L}, supp(L)\subseteq e} \nu(L(x)), \tag{4.2.7}$$

with the convention that $\nu_e \equiv 1$ if $\{L; supp(L) = e\} = \emptyset$.

Note that if $\Delta = \{v_0, \ldots, v_d\}$ is in general position, that is if $v_i^k \neq v_j^k$ for all $i \neq j$ and k then supp$(L_e^k) = e$ for all $e \in \mathcal{H}_d$ hence

$$\mu_e(x) = \nu_e(x) = \prod_{k=1}^{d} \nu(L_e^k(x)).$$

Remark 4.2.2. *As mentioned before, it is sometimes more convenient to think of μ_e as a measure on V_J (rather than V_e) in the obvious way. In general for $x \in V_J$, we have $\mu_e(x) = \prod_{f \subseteq e} \nu_f(x)$ and also $\mu_e(x) = \mu_e(\pi_e(x))$, that is μ_e is constant along the fibers of the projection π_e hence we can think of μ_e as a function on V_e as well. We will refer the functions ν_e and μ_e as* weights *and* measures *respectively. To emphasize this point of view we will often use the integral notation and write*

$$\int_{V_J} F(x)\, d\mu_e(x) := \mathbb{E}_{x \in V_J} F(x)\, \mu_e(x), \ \ \textit{and} \ \ \int_{V_e} F_e(x)\, d\mu_e(x) := \mathbb{E}_{x \in V_e} F_e(x)\, \mu_e(x),$$

for functions $F : V_J \to \mathbb{R}$ and $F_e : V_e \to \mathbb{R}$. Thus we could think of μ_e as a measure on V_J or on the subspace V_e, the exact interpretation will be clear from the context. Note that for $F_e : V_e \to [-1, 1]$,

$$\int_{V_J} F_e(\pi_e(x)) d\mu_e(x) = \int_{V_e} F_e(x) d\mu_e(x)$$

and it follows easily from the linear forms condition (see Lemma 4.2.3 below) that

$$\int_{V_J} F_e(\pi_e(x)) d\mu_J(x) = \int_{V_e} F_e(x) d\mu_e(x) + o_{N,W \to \infty}(1).$$

Now we prove in Lemma 4.2.3 below that measure μ_e and μ_J are *essentially* probability measures and in fact *essentially* the same measure and this supports the idea of identifying functions or sets on V_e with functions or sets on V_J in the obvious way: consider sets $G_e \subseteq V_e$ as sets $\overline{G}_e = \pi_e^{-1}(G_e) \subseteq V_J$, changing their measure only by a negligible amount

$$\mu_J(\overline{G}_e) = \mu_e(G_e) + o(1) \tag{4.2.8}$$

The proof is a prototype of the arguments that are based on the Linear Forms Condition. Here, as in previous chapter, we don't need to analyze the structure of linear forms, only inspect that each linear forms depend on different sets of variables and hence linearly independent.

Note that for any $|e| = d$ we have that $\{L_e^k, 1 \leq k \leq d\}$ are linearly independent which may be inspected from (4.2.5). Alternately, since each linear form of this set represents a coordinate of a vertex, the set of all normal vectors to each of face of the simplex (choose one vector from each face) called $\{n_1, \ldots, n_{d+1}\}$. Any d vectors chosen from this set are linearly independent. Recall that we can parametrize affine copies of a simplex in $\mathbb{Z}_N^d$ by $\mathbb{Z}_N^{d+1}$: a point $(x_1, \ldots, x_{d+1}) \in \mathbb{Z}_N^{d+1}$ represents an affine copy of the simplex where the equation of the $d+1$ planes that constitute the simplex are $p \cdot n_i = x_i, 1 \leq i \leq d+1$. Each vertex $p = (p_1, \ldots, p_d)$ of this affine simplex is obtained by solving

$$p \cdot m_1 = y_1$$
$$\vdots$$
$$p \cdot m_d = y_d$$

here $m_1, \ldots, m_d$ are d vectors chosen from $\{n_1, \ldots n_{d+1}\}$ and $y_1, \ldots y_d$ are chosen from $\{x_1, \ldots x_{d+1}\}$, corresponding to choices of $m_1, \ldots, m_d$. This means we solve $Ap = (y_1, ..., y_d)$ for some invertible matrix A hence $p = A^{-1}(y_1, \ldots, y_d)$ so each p_i is represented by a linear form in variables $y_1, \ldots y_d$ and since A is invertible, these k linear forms are in fact linearly independent.

Lemma 4.2.3. *For all $e \in \mathcal{H}$ we have that*

$$\mu_e(V_e) = 1 + o(1), \tag{4.2.9}$$

moreover if $g : V_e \to [-1, 1]$,

$$\mathbb{E}_{x_e \in V_e}\, g(x_e)\, \mu_e(x_e) = \mathbb{E}_{x \in V_J}\, g(\pi_e(x))\, \mu_J(x) + o(1),$$

or equivalently

$$\int_{V_e} g\, d\mu_e = \int_{V_J} (g \circ \pi_e)\, d\mu_J + o(1). \tag{4.2.10}$$

Proof. Note that the linear forms appearing on the right side of

$$\mu_e(V_e) = \mathbb{E}_{x \in V_e} \prod_{supp(L) \subseteq e} \nu(L(x))$$

are pairwise linearly independent, and as they are supported on e they remain pairwise independent when restricted to V_e. Thus (4.2.9) follows from the linear forms condition.
To show (4.2.10), let $e' = J \backslash e$ and write $x = (x_e, x_{e'})$ with $x_e = \pi_e(x),\ x_{e'} = \pi_{e'}(x)$. Then

$$E := \mathbb{E}_{x \in V_J} (g \circ \pi_e)(x) \mu_J(x) - \mathbb{E}_{x_e \in V_e}\, g(x_e)\, \mu_e(x_e) = \mathbb{E}_{x_e \in V_e}\, g(x_e)\, \mu_e(x_e)\, \mathbb{E}_{x_{e'} \in V_{e'}} (\mathrm{w}(x_e, x_{e'}) - 1),$$

where $\mathrm{w}(x_e, x_{e'}) = \prod_{f \nsubseteq e} \nu_f(x_{e \cap f}, x_{e' \cap f})$.

Now we consider E^2 to get rid of g (using $|g| \leq 1$). By (4.2.9) we have that $\mu_e(V_e) \lesssim 1$, and then apply the Cauchy-Schwartz inequality in x_e variables,

$$|E|^2 = \mathbb{E}_{x_e \in V_e} g(x_e) \mu_e(x_e)^{1/2} \times \mathbb{E}_{x_{e'} \in V_{e'}} \mu_e(x_e)^{1/2} (\mathrm{w}(x_e, x_{e'}) - 1)$$

$$\lesssim \mathbb{E}_{x_e \in V_e} \mathbb{E}_{x_{e'}, y_{e'} \in V_{e'}} \left(\mathrm{w}(x_e, x_{e'}) - 1\right)\left(\mathrm{w}(x_e, y_{e'}) - 1\right) \mu_e(x_e).$$

The right hand side of this expression is a combination of four terms and (4.2.10) follows from the fact that each term is $1 + o(1)$. Indeed the linear forms appearing in the definition of the function $\mu_e(x_e)$ depend only on the variables x_j for $j \in e$ and are pairwise linearly independent. All linear forms involved in $\mathrm{w}(x_e, x_{e'})$ depend also on some of the variables in x_j, $j \in e'$, while the ones in $\mathrm{w}(x_e, y_{e'})$ depend on the variables in y_j, $j \in e'$, hence these forms depend on different sets of variables. Thus the forms appearing in the expression $\mu_e(x_e)\mathrm{w}(x_e, x_{e'})\mathrm{w}(x_e, y_{e'})$ are pairwise linearly independent and (4.2.10) follows from the linear forms condition:

$$|E|^2 \lesssim \big(1 + o(1)\big) - \big(1 + o(1)\big) - \big(1 + o(1)\big) + \big(1 + o(1)\big) = o(1)$$

Note that the estimate is independent on the function g. $\square$

Counting Prime Simplices.

To see how to use weighted hypergraph $\{\nu_e\}_{e \in \overline{\mathcal{H}}}$ to count prime simplices we follow [102] to parameterize affine copies of Δ. Define the map $\Phi : \mathbb{Z}_N^{d+1} \to \mathbb{Z}_N^{d+1}$ by

$$\Phi(x) = \Big(\sum_{i=0}^{d} x_i v_i, -\sum_{i=0}^{d} x_i\Big) := (y, t) \tag{4.2.11}$$

By (4.2.4) and (4.2.11) we have that $x \in E_e$ for $e = J \backslash \{j\}$ if and only if $y + tv_j \in A$ thus $x \in \bigcap_{e \in \mathcal{H}_d} E_e$ exactly when $y + t\Delta \subseteq A$. Since $\{v_1 - v_0, \ldots, v_d - v_0\}$ is a linearly independent family of vectors, we have that Φ is one to one. Hence, this gives a parametrization of all affine copies of Δ contained in $A \pmod N$. Also for $e = J \backslash \{j\}$,

$$L_e^k(x) = \sum_{i=0}^{d} x_i (v_i^k - v_j^k) = \pi_k(y + tv_j) \tag{4.2.12}$$

where π_k is the orthogonal projection to the k^{th} coordinate axis. This implies that

$$\mu_e(x) = \prod_{\mathrm{supp}(L) \subseteq e} \nu(L(x)) = \prod_{k=1}^{d} \nu(L_e^k(x)) = w(y + tv_j), \tag{4.2.13}$$

and also

$$\mu_J(x) = \prod_{L \in \mathcal{L}} \nu(L(x)) = w(y + t\Delta). \tag{4.2.14}$$

In particular $\mu_J(\bigcap_{e \in \mathcal{H}_d} E_e)$ counts the number of prime affine copies of Δ.

Next we observe that our linear forms do satisfy some useful properties which we will refer to later:

Theorem 4.2.4 (Properties of a family of linear forms)**.** *Consider a family of linear forms* $\mathcal{L} = \{L_e^k; e \in \mathcal{H}_d, 1 \leq k \leq d\}$ *associated with hypergraph. Our system of linear forms satisfies the following properties.*

- *If* $e = J\backslash\{j\}$, $e' = J\backslash\{j'\}$ *then supp*$(L_{e'}^k) \subseteq e$ *if and only if* $v_j^k = v_{j'}^k$ *(i.e. the* k^{th} *coordinate of* $v_j, v_{j'}$ *are the same). This is equivalent to* $L_{e'}^k = L_e^k$*. We call such a family* $\mathcal{L}$ **well-defined***.*
- *Since for a given* $e \in \mathcal{H}_d$*, the forms* $\mathcal{L}_e = \{L_e^k, 1 \leq k \leq d\}$ *are linearly independent. Any two distinct forms of the family* $\mathcal{L}$ *are linearly independent. We will refer to such families of forms as being* **pairwise linearly independent***.*
- *Let* $M = \{x \in V_J : x_0 + \ldots + x_d = 0\}$*. Then for any* $x \in M$*, from definition of the linear forms, we have* $L_e^k(x) = L_{e'}^k(x)$ *for all* $e, e' \in \mathcal{H}_d$ *and* k*. We call a family of linear forms* $\mathcal{L}$ *satisfying this property* **symmetric***.*[5]

Example 4.2.5 (Corners in $\mathbb{Z}^2$)**.** $\Phi(x_0, x_1, x_2) = (x_1, x_2, -x_0 - x_1 - x_2); y - (x_1, x_2), t = -x_0 - x_1 - x_2$.

$$y + tv_1 = (x_1, x_2) + (-x_0 - x_1 - x_2)(0, 0) = (x_1, x_2)$$
$$y + tv_2 = (x_1, x_2) + (-x_0 - x_1 - x_2)(1, 0) = (-x_0 - x_2, x_2)$$
$$y + tv_3 = (x_1, x_2) + (-x_0 - x_1 - x_2)(0, 1) = (x_1, -x_0 - x_1)$$

- *Linear forms are* $x_1, x_2, -x_0 - x_2, -x_0 - x_1$.
- *Linear forms associated with* $(0, 1)$ *are* $L_{(0,1)}^1(x_0, x_1) = -x_0 - x_1, L_{(0,1)}^2(x_0, x_1) = x_1$*. Linear forms associated with* $(0, 2)$ *are* $L_{(0,2)}^1(x_0, x_2) = -x_2 - x_0, L_{(0,2)}^2(x_0, x_2) = x_2$*. Linear forms associated with* $(1, 2)$ *are* $L_{(1,2)}^1(x_1, x_2) = x_1, L_{(1,2)}^2(x_1, x_2) = x_2$*.*
- *Examples of symmetric property: Let* $(x_0, x_1, x_2) = M = \{(x_0, x_1, x_2) : x_0 + x_1 + x_2 = 0\}$ *then*

$$-x_0 - x_1 = L_{(0,1)}^1(x_0, x_1, x_2) = L_{(0,1)}^1(x_0, x_1) = L_{(0,2)}^1(x_0, -x_2 - x_0) = -x_0 - (-x_2 - x_0) = x_2 = L_{(0,2)}^2(x_0, x$$

$$x_1 = L_{(0,1)}^2(x_0, x_1, x_2) = L_{(0,2)}^2(x_0, x_2 - x_0) = -x_2 - x_0 = L_{(0,2)}^1(x_0, x_1, x_2).$$

[5]The set M correspond the degenerated copy $x + t\Delta$ with $t = 0$, saying that it should be degenerated to a single point. Given $f \in \mathcal{H}$ with a set of linear forms $\mathcal{L}_f$, we can have a process of symmetrization to obtain a system of linear forms defined on all hyperedges which is symmetric such that the linear forms that only depend on variables in f is $\mathcal{L}_f$. See section 4.3.2.

4.3 Parametric Weight Systems: Extensions, Stability, and Symmetrization.

Definition 4.3.1 (weight systems and associated families of measures depending on parameters.)**.** *Let*

$$\mathcal{L}_q := (L^1(q,x), ..., L^s(q,x))$$

be a family of linear forms with integer coefficients depending on the parameters $q \in \mathbb{Z}^R$ and the variables $x \in \mathbb{Z}^D$. We call the family **pairwise linearly independent** *if no two forms in the family are rational multiples of each other (considered as forms over q and x). If N is a sufficiently large prime with respect to the coefficients of the linear forms $L^i(q,x)$, then the forms remain pairwise linearly independent when considered as forms over $Z \times V$, $Z = \mathbb{Z}_N^R$, $V = \mathbb{Z}_N^D$. We refer to the set $Z = \mathbb{Z}_N^R$ as the* **parameter space** *of the family $\mathcal{L}_q$. We call the family of parametric forms $\mathcal{L}_q$* **well-defined** *if all forms $L^i(q,x)$ depend on some of the x-variables and there is measure on Z of the form*

$$\int_Z g(q)\, d\psi(q) = \mathbb{E}_{q\in Z}\; g(q)\,\psi(q), \quad \psi(q) = \prod_{i=1}^{t} \nu(Y_i(q)), \tag{4.3.1}$$

for a family of pairwise linearly independent linear forms Y_i defined over Z.
*If $V = V_J$ then we define an associated system of weights $\{\nu_{q,e}\}_{q\in Z, e\in\mathcal{H}}$ and measures $\{\mu_{q,e}\}_{q\in Z, e\in\mathcal{H}}$ as follows: For a form $L^k(q,x) = \sum_i b_i q_i + \sum_j a_j x_j$ define its x-**support** as $supp_x(L) = \{j \in J;\ a_j \neq 0\}$. For $e \subseteq J$ and $q \in Z$, let*

$$\nu_{q,e}(x) := \prod_{\substack{L\in\mathcal{L}_q \\ supp_x(L)=e}} \nu(L(q,x)), \qquad \mu_{q,e}(x) := \prod_{\substack{L\in\mathcal{L}_q \\ supp_x(L)\subseteq e}} \nu(L(q,x)) \tag{4.3.2}$$

We use the convention that $\nu_{q,e} \equiv 1$ if there is no form $L \subseteq \mathcal{L}_q$ such that $supp_x(L) = e$. Note that the x-support partitions the family of forms $\mathcal{L}_q$ is independent of the parameters q, thus for given $e \in \mathcal{H}$

$$\mu_{q,e}(x) = \prod_{f\subseteq e} \nu_{q,f}(x), \quad \text{for all } \ q \in Z.$$

A crucial observation is that many of the properties of the measure system $\{\mu_e\}$ still hold for well-defined measure systems $\{\mu_{q,f}\}$ for *almost every* value of the parameter $q \in Z$. In order to formulate such statements, we give the following definition.

Definition 4.3.2. *We define the* **dimension** *of the space Z, the number of linear forms $L^j(q,x)$, $Y_l(q)$. We say that the family $\mathcal{L}_q$ has* **complexity** *at most K if the dimension of the space Z and the magnitude of their coefficients are all bounded by K.*

Remark 4.3.3. *The error terms in applications of the linear forms conditions will depend on quantity*

K.

We have the analogue of Lemma 4.2.3 for parametric weight system.

Lemma 4.3.4. *Let $\{\mu_{q,e}\}_{e\in\mathcal{H},q\in Z}$ be a well-defined parametric measure system of complexity at most K.*
For every $e \in \mathcal{H}$ there is a set $\mathcal{E}_e \subseteq Z$ such that $\psi(\mathcal{E}_e) = o_K(1)$, and for every $q \notin \mathcal{E}_e$

$$\mu_{q,e}(V_e) = 1 + o_K(1). \tag{4.3.3}$$

Moreover for every $e \in \mathcal{H}$ there is a set $\mathcal{E}_e \subseteq Z$ of measure $\psi(\mathcal{E}_e) = o(1)$, such the following holds. For any function $g : Z \times V_e \to [-1, 1]$ and for every $q \notin \mathcal{E}_e$ one has the estimate

$$\int_{V_e} g(q, x_e)\, d\mu_{q,e}(x_e) = \int_{V_J} g(q, \pi_e(x))\, d\mu_{q,J}(x) + o_K(1). \tag{4.3.4}$$

Proof. To prove (4.3.3), consider the quantity

$$\begin{aligned}\Lambda_e &:= \int_Z |\mu_{q,e}(V_e) - 1|^2\, d\psi(q)\\ &= \mathbb{E}_{q\in Z}\mathbb{E}_{x_e,y_e}\big(\prod_{supp_x(L)\subseteq e} \nu(L(q, x_e)) - 1\big)\big(\prod_{supp_x(L)\subseteq e} \nu(L(q, y_e)) - 1\big)\, d\psi(q).\end{aligned}$$

The above expression is a combination of four terms and note that the family of linear forms

$$\{Y_k(q), L^i(q, x_e), L^j(q, y_e)\}$$

is pairwise linearly independent in the (q, x_e, y_e) variables by our assumption on the linear forms. Applying the linear forms condition gives that each term is $1 + o_K(1)$ and so $\Lambda_e = o_K(1)$ and (4.3.3) follows.

Now let $e' = J\backslash e$, write $x = (x_e, x_{e'})$ and arguing as in Lemma 4.2.3 we consider the difference in (4.3.4),

$$\begin{aligned}\Lambda(q, e, g) &:= \Big| \int_{V_J} g(q, \pi_e(x)) d\mu_{q,J}(x) - \int_{V_e} g(q, x_e) d\mu_{q,e}(x_e)\Big|\\ &= |\, \mathbb{E}_{x\in V_J}\, g(q, \pi_e(x))\, \mu_{q,J}(x) - \mathbb{E}_{x_e\in V_e}\, g(q, x_e)\, \mu_{q,e}(x_e)|\\ &= |\mathbb{E}_{x_e\in V_e}\, g(q, x_e)\, \mu_{q,e}(x_e)\, \mathbb{E}_{x_{e'}\in V_{e'}}\, (\mathrm{w}_q(x_e, x_{e'}) - 1)|\\ &\le \mathbb{E}_{x_e\in V_e}\, \mu_{q,e}(x_e)\, |\mathbb{E}_{x_{e'}\in V_{e'}}\, (\mathrm{w}_q(x_e, x_{e'}) - 1)|,\end{aligned}$$

where $\mathrm{w}_q(x_e, x_{e'}) = \prod_{f \not\subseteq e} \nu_{q,f}(x_{e\cap f}, x_{e'\cap f})$.

Notice that the right hand side of the above inequality is independent of the function g; if we denote it by $\Lambda(q,e)$ then (4.3.4) (holds for almost every q) would follow from the estimate $\mathbb{E}_{q\in Z}\, \Lambda(q,e)\, d\psi(q) = o_K(1)$. By the linear forms condition $\mathbb{E}_{q,x_e}\, d\psi(q)\, d\mu_{q,e}(x_e) = 1 + o_K(1) \leq 2$, for N sufficiently large with respect to K. Then by the Cauchy-Schwartz inequality one has

$$(\mathbb{E}_{q\in Z}\, \Lambda(q,e)\, d\psi(q))^2 \lesssim \mathbb{E}_{q\in Z}\mathbb{E}_{x_e\in V_e}\, \mu_{q,e}(x_e)|\mathbb{E}_{x_{e'}\in V_{e'}}\, (\mathrm{w}_q(x_e,x_{e'})-1)|^2\, d\psi(q)$$
$$\lesssim \mathbb{E}_{q\in Z,\, x_e\in V_e}\, \mathbb{E}_{x_{e'},y_{e'}\in V_{e'}}\, (\mathrm{w}_q(x_e,x_{e'})-1)(\mathrm{w}_q(x_e,y_{e'})-1)\, d\mu_{q,e}(x_e)\, d\psi(q).$$

This is a combination of four terms, however each term again is $1+o_K(1)$ as the linear forms defining ψ depend on the variables q while the ones defining $\mu_{q,e}$ depend also on the x_e variables. On the other hand all linear forms appearing in the weight functions $\mathrm{w}_q(x_e,x_{e'})$ (respectively, $\mathrm{w}_q(x_e,y_{e'})$) depend on the $x_{e'}$ (respectively, $y_{e'}$) variables as well. Thus the family of all linear forms in the above expressions is pairwise linearly independent in the $(q,x_e,x_{e'},y_{e'})$ variables. The result follows from linear forms condition. $\square$

4.3.1 Extension of Parametric Weight System, Stability and Symmetrization.

Energy increment argument involves iterations. In our setting, it turns out that when we do an iteration, due to our averaging argument, we end up with a new parametric system of measure which is an extension of the original measure system (in the sense that the weights in the definition of the original measure are included in the definition of the new measure). The fact we will prove is that most of these extensions are just a small perturbation of the original measure and still shares many important properties with the original measure.

Definition 4.3.5 (Parametric Extension)**.** *Let*

$$\mathcal{L}^1_{q_1} = \{L^1_1(q_1,x), ..., L^{s_1}_1(q_1,x)\},\;\; \mathcal{L}^2_{q_2} = \{L^1_2(q_2,x), ..., L^{s_2}_2(q_2,x)\}$$

be two pairwise linearly independent families of linear forms defined on the parameter spaces $Z_1 = \mathbb{Z}_N^{k_1}$ and $Z_2 = \mathbb{Z}_N^{k_2}$. Let ψ_1 and ψ_2 be measures on Z_1 and Z_2 defined by the families of linear forms $\{Y^1_1(q_1), \dots Y^1_{s_1}(q_1)\}$ and $\{Y^2_1(q_2), \dots Y^2_{s_2}(q_2)\}$.
We say that the family $\mathcal{L}^2_{q_2}$ is an ***extension*** *of the family $\mathcal{L}^1_{q_1}$ if $Z_1 \leq Z_2$ (Z_1 may be empty) and the following holds: The family of forms $L^i_2(q_2,x), Y^2_j(q_2)$ which depend only on the variables $q_1 = \pi(q_2)$ is exactly the family of forms $L^i_1(q_1,x)$, $Y^1_j(q_1)$, where $\pi : Z_2 \to Z_1$ is the natural orthogonal projection.*
If $V = V_J$ let $\mu^1 := \{\mu_{q_1,e}\}_{q_1\in Z_1, e\in\mathcal{H}}$ and $\mu^2 := \{\mu_{q_2,f}\}_{q_2\in Z_2, f\in\mathcal{H}}$ be the associated measure systems as defined in (4.3.2). We say that the measure system μ^2 is an **extension** *of the system μ^1.*

Remark 4.3.6. *Writing $Z_2 = Z_1 \times Z$, $Z = \mathbb{Z}_N^r$ and $q_2 = (q_1, q)$, we have*

$$\psi_2(q_1, q) = \psi_1(q_1) \cdot \varphi(q_1, q) \tag{4.3.5}$$

where $\varphi(q, q_1) = \prod_{i=1}^{t} \nu(Y_i(q_1, q))$. The linear forms $Y_i(q_1, q)$ defining $\varphi(q_1, q)$ depend on some of the variables of $q = (q_i)_{1\leq i\leq k}$ and are pairwise linearly independent. Similarly one may write for any $e \in \mathcal{H}$

$$\mu^2_{(q_1,q),e}(x_e) = \mu^1_{q_1,e}(x_e)\mathrm{w}_e(q_1, q, x_e) \tag{4.3.6}$$

where the linear forms $L_2^j(q_1, q, x_e)$ defining the function $\mathrm{w}_e(q, q_1, x_e)$ depend on (some of) the variables q as well as on all of the variables x_e.

Next lemma, we prove ***Stability Property*** of a parametric extension. Saying that **most** extensions μ_{q_2} of μ_{q_1} is just a small perturbation of μ_{q_1}, by quantities that is independent of q_2.

Lemma 4.3.7 (Stability property of measure)**.** *Let $\{\mu_f\}_{f\in\mathcal{H}}$ be a well defined measure system, and let $\{\mu_{q,f}\}_{q\in Z, f\in\mathcal{H}}$ be a well-defined parametric extension of $\{\mu_f\}_{f\in\mathcal{H}}$ of complexity at most K. Then for any $f \in \mathcal{H}$ and for any function $g : V_f \to [-1, 1]$ there is a set $\mathcal{E}_{g,f} \subseteq Z$ of measure $\psi(\mathcal{E}_{g,f}) = o_K(1)$, so that for all $q \notin \mathcal{E}_{g,f}$*

$$\int_{V_f} g \, d\mu_{q,f} - \int_{V_f} g \, d\mu_f = o_K(1). \tag{4.3.7}$$

Similarly if $\{\mu_{q_1,f}\}_{f\in\mathcal{H}, q_1\in Z_1}$ is a well-defined parametric system and if $\{\mu_{q_2,f}\}_{f\in\mathcal{H}, q_2\in Z_2}$ is an extension of complexity at most K_2, then to any function $g : Z_1 \times V_f \to [-1, 1]$ there exists a set $\mathcal{E}_{g,f} \subseteq Z_2$ of measure $\psi_2(\mathcal{E}_{g,f}) = o_{K_2}(1)$, such that for all $q_2 = (q_1, q) \notin \mathcal{E}_{g,f}$

$$\int_{V_f} g(q_1, x) \, d\mu_{q_2,f}(x) - \int_{V_f} g(q_1, x) \, d\mu_{q_1,f}(x) = o_{K_2}(1). \tag{4.3.8}$$

Proof. As $\mu_{q,f} = \mu_f(x_f)\mathrm{w}_f(q, x_f)$, the left side of (4.3.7) may be written as

$$\Lambda_{f,g}(q) := \int_{V_f} g(x)(\mathrm{w}_f(q, x) - 1) \, d\mu_f(x).$$

Consider the average over q,

$$\Lambda_{f,g} := \int_Z |\Lambda_{f,g}(q)|^2 \, d\psi(q).$$

which is non-negative. By expanding, we have

$$\begin{aligned}
\Lambda_{f,g} &= \int_Z \int_{V_f} \int_{V_f} (\mathrm{w}_f(q, x) - 1)(\mathrm{w}_f(q, y) - 1) g(x)g(y) \, d\mu_f(x)d\mu_f(y)d\psi(q) \\
&\leq \int_{V_f} \int_{V_f} \left| \int_Z (\mathrm{w}_f(q, x) - 1)(\mathrm{w}_f(q, y) - 1) d\psi(q) \right| \, d\mu_f(x)d\mu_f(y).
\end{aligned}$$

Now the Cauchy-Schwartz inequality in $V_f \times V_f$ variables and (4.2.9) gives

$$|\Lambda_{f,g}|^2 \lesssim \int_{V_f}\int_{V_f}\int_Z\int_Z (\mathrm{w}_f(q,x)-1)(\mathrm{w}_f(q,y)-1)\times$$
$$\times\ (\mathrm{w}_f(p,x)-1)(\mathrm{w}_f(p,y)-1)\ d\mu_f(x)d\mu_f(y)d\psi(q)d\psi(p).$$

This last expression is a combination of 16 terms where each term is $1+o_K(1)$ by the linear form conditions and their total contribution is $o(1)$. Indeed the linear forms which can appear in any of these terms are $Y_{i_1}(q)$,$Y_{i_2}(p)$,$L^{i_3}(x)$,$L^{i_4}(y)$, $L^{i_5}(q,x)$, $L^{i_6}(q,y)$, $L^{i_7}(p,x)$, $L^{i_8}(p,y)$. Note that the last 4 terms depend on both sets of variables (for example $L^i(q,x)$ depends both on $q \in Z$ and on $x \in V_f$), and hence the family of these forms are pairwise linearly independent in the (q,p,x,y) variables. This Proves (4.3.7).

The proof of (4.3.8) is essentially the same. Set

$$\Lambda_{f,g}(q_2) := \int_{V_f} g(q_1,x)d\mu_{q_2,f}(x) - \int_{V_f} g(q_1,x)d\mu_{q_1,f}(x)$$

and

$$\Lambda_{f,g} := \int_{Z_2} |\Lambda_{f,g}(q_2)|^2\, d\psi_2(q_2).$$

where we write $Z_2 = Z_1 \times Z$, $Z = \mathbb{Z}_N^k$, and $q_2 = (q_1, q)$ for $q_2 \in Z_2$. By (4.3.5) we estimate as above

$$\Lambda_{f,g} \lesssim \int_{V_f}\int_{V_f}\int_{Z_1} d\psi_1(q_1)d\mu_{q_1,f}(x)d\mu_{q_1,f}(y)\ \left|\mathbb{E}_{q\in Z}\,(\mathrm{w}_f(q_1,q,x)-1)(\mathrm{w}_f(q_1,q,y)-1)\varphi(q_1,q)\right|.$$

The linear forms condition gives

$$\int_{V_f}\int_{V_f}\int_{Z_1} d\psi_1(q_1)d\mu_{q_1,f}(x)d\mu_{q_1,f}(y) = 1 + o_{K_2}(1),$$

so by Cauchy-Schwartz's inequality, we have

$$|\Lambda_{f,g}|^2 \lesssim \int_{V_f}\int_{V_f}\int_{Z_1} \mathbb{E}_{p,q\in Z}\,(\mathrm{w}_f(q_1,q,x)-1)(\mathrm{w}_f(q_1,q,y)-1)\times$$
$$\times\,(\mathrm{w}_f(q_1,p,x)-1)(\mathrm{w}_f(q_1,p,y)-1)\ \varphi(q_1,q)\varphi(q_1,p)\ d\psi_1(q_1)d\mu_{q_1,f}(x)d\mu_{q_1,f}(y).$$

Now any linear form $L_f^i(q_1,q,x)$ depends both on the variables q and x. Thus again the left side is a combination of 16 terms, each being $1+o_{K_2}(1)$ by the linear forms condition as all the linear forms involved in any of these expressions are pairwise linearly independent in the (x,y,q_1,q,p) variables. $\square$

Next we will prove stability property of box norm with respect to an extension. Let $g : Z_1 \times V_e \to \mathbb{R}$ be a function and let $e \in \mathcal{H}$, $|e| = d'$. For a given $q_1 \in Z_1$ recall the box norm of $g_{q_1}(x) = g(q_1, x)$

$$\|g_{q_1}\|^{2^{d'}}_{\Box_{\mu_{q_1,e}}} = \mathbb{E}_{p,x\in V_e} \prod_{\omega_e\in\{0,1\}^e} g(q_1, \omega_e(p,x)) \prod_{f\subseteq e} \prod_{\omega_f\in\{0,1\}^f} \nu_{q,f}(\omega_f(p_f, x_f)), \tag{4.3.9}$$

where $x_f = \pi_f(x)$, $p_f = \pi_f(p)$, $\pi_f : V_e \to V_f$ being the natural projection. Here we have a parametric linear forms $\mathcal{L}_{q_1,e}$ on $Z_1 \times V_e$ as in Definition 4.3.1. The inner product on the right side of (4.3.9) is defined by the parametric family of forms (in $(p,x)-$variables)

$$\tilde{\mathcal{L}}_{q_1,e} = \bigcup_{f\subseteq e} \{L(q_1, \omega_f(p_f, x_f));\ L \in \mathcal{L}_{q_1},\ supp_x(L) = f,\ \omega_f \in \{0,1\}^f\}. \tag{4.3.10}$$

Claim. *$\tilde{\mathcal{L}}_{q_1}$ is a pairwise linearly independent family of forms defined over $Z_1 \times V$ ($V = V_e \times V_e$).*

Proof of Claim. Suppose we would have that

$$L'(q_1, \omega'_{f'}(p_{f'}, x_{f'})) = \lambda L(q_1, \omega_f(p_f, x_f)), \tag{4.3.11}$$

then restriction both forms to the subspace $\{(p,x) \in V_f \times V_f : p = x\}$ would imply that $L'(q_1, x_{f'}) = \lambda L(q_1, x_f)$ and hence $f' = supp_x(L') = supp_x(L) = f$. Then, as L and L' depend exactly variables $x_j, j \in f$. For the equation (4.3.11) to hold, we should have $\omega'_f = \omega_f$ and $L = L'$. □

Lemma 4.3.8 (Stability Property of Box Norm)**.** *Let $\{\nu_{q_1,f}\}_{f\in\mathcal{H},q_1\in Z_1}$ be a parametric weight system with a well-defined extension $\{\nu_{q_2,f}\}_{f\in\mathcal{H},q_2\in Z_2}$ of complexity at most K_2. Then to any $e \in \mathcal{H}$ and to any function $g : Z_1 \times V_e \to [-1,1]$ there exists a set $\mathcal{E} = \mathcal{E}(g,e) \in Z_2$ of measure $\psi_2(\mathcal{E}) = o_{K_2}(1)$ such that for all $q_2 = (q_1, p) \notin \mathcal{E}$*

$$\|g_{q_1}\|_{\Box_{\mu_{q_2,e}}} = \|g_{q_1}\|_{\Box_{\mu_{q_1,e}}} + o_{K_2}(1). \tag{4.3.12}$$

Proof. Let

$$G_{q_1}(p,x) := \prod_{\omega_e\in\{0,1\}^e} g(q_1, \omega_e(p,x)), \tag{4.3.13}$$

and let $\{\tilde{\mu}_{q_1,e}\}_{q_1\in Z_1}$ denotes the associated system of measures on $Z_1 \times V_e$, then for given $q_1 \in Z_1$ (according to definition in the equation (4.3.9)), we can write

$$\|g_{q_1}\|^{2^{d'}}_{\Box_{\mu_{q_1,e}}} = \mathbb{E}_{p,x\in V_e} G_{q_1}(p,x)\, \tilde{\mu}_{q_1,e}(p,x). \tag{4.3.14}$$

Now, if $\mathcal{L}_{q_2}$ is a well-defined parametric extension of $\mathcal{L}_{q_1}$ then (4.3.10) yields to a well-defined parametric extension $\tilde{\mathcal{L}}_{q_2}$ of the family $\tilde{\mathcal{L}}_{q_1}$. Then by Lemma 4.3.7, and the simple observation that $|a^{2^{d'}} - b^{2^{d'}}| \leq \varepsilon$ implies[6] $|a-b| \leq \varepsilon^{2^{-d'}}$ for $a, b \geq 0$, we are done.

$\square$

Finally we prove prove the stability of the conditional expectation, both with respect to L^2 norm and Box norm. Recall that if $(V, \mathcal{B}, \mu)$ and $g : V \to \mathbb{R}$, we defined

$$\mathbb{E}_\mu(g|\mathcal{B})(x) = \frac{1}{\mu(B(x))} \int_{B(x)} g(y) d\mu(y) := \frac{1}{\mu(B(x))} \mathbb{E}_{y \in V} \mathbf{1}_{B(x)}(y) g(y) \mu(y),$$

where $B(x) \in \mathcal{B}$ is the atom containing x. If $\mu(B(x)) = 0$ then we set $\mathbb{E}_\mu(g|\mathcal{B})(x) = 1$.

Lemma 4.3.9. *Let $(\mu_{q_1,f})_{q_1 \in Z_1, f \in \mathcal{H}}$ be a well-defined parametric measure system with a well-defined extension $(\mu_{q_2,f})_{q_2 \in Z_2, f \in \mathcal{H}}$ of complexity at most K_2. For $q_1 \in Z_1$ and $e \in \mathcal{H}$, let $\mathcal{B}_{q_1,e}$ be a $\sigma-$algebra on V_e such that compl($\mathcal{B}_{q_1,e}$)$\leq M$ for some fixed number M. For any function $g_{q_1} : Z_1 \times V_e \to [-1, 1]$ there exists a set $\mathcal{E} = \mathcal{E}(\mathcal{B}_{q_1,e}, g) \subseteq Z_2$ of measure $\psi_2(\mathcal{E}) = o_{M,K_2}(1)$ such that for any $q_2 = (q_1, q) \notin \mathcal{E}$. We have*

1.
$$\left\| \mathbb{E}_{\mu_{q_2,e}}(g_{q_1}|\mathcal{B}_{q_1,e}) - \mathbb{E}_{\mu_{q_1,e}}(g_{q_1}|\mathcal{B}_{q_1,e}) \right\|^2_{L^2(\mu_{q_2,e})} = o_{M,K_2}(1) \tag{4.3.15}$$

2.
$$\left\| \mathbb{E}_{\mu_{q_2,e}}(g_{q_1}|\mathcal{B}_{q_1,e}) \right\|^2_{L^2(\mu_{q_2,e})} = \left\| \mathbb{E}_{\mu_{q_1,e}}(g_{q_1}|\mathcal{B}_{q_1,e}) \right\|^2_{L^2(\mu_{q_1,e})} + o_{M,K_2}(1). \tag{4.3.16}$$

Proof. Let $m = 2^M$ and enumerate the atoms of $\mathcal{B}_{q_1,e}$ as $B^1_{q_1}, ..., B^m_{q_1}$, allowing some of them to possibly be empty. For a fixed $1 \leq i \leq m$ define the functions

$$b_i(q_1, x) := \mathbf{1}_{B^i_{q_1}(x)} = \begin{cases} 1 & \text{if} \quad x \in B^i_{q_1} \\ 0 & \text{otherwise} \end{cases}$$

and for $q_2 = (q_1, q) \in Z_2$ define the quantities

$$\mu_i(q_2, g) := \int_{V_e} g(q_1, x) b_i(q_1, x) d\mu_{q_2,e}(x), \qquad \mu_i(q_2) := \mu_i(q_2, 1) = \mu_{q_2,e}(B^i_{q_1}),$$

$$\mu_i(q_1, g) := \int_{V_e} g(q_1, x) b_i(q_1, x) d\mu_{q_1,e}(x), \qquad \mu_i(q_1) := \mu_i(q_1, 1) = \mu_{q_1,e}(B^i_{q_1})$$

[6]Indeed $|a^{2^{d'-1}} - b^{2^{d'-1}}|^2 \leq |a^{2^{d'-1}} - b^{2^{d'-1}}||a^{2^{d'-1}} + b^{2^{d'-1}}| = |a^{2^{d'}} - b^{2^{d'}}| \leq \varepsilon$. Then we may argue by induction.

By Lemma 4.3.7 we have that

$$\mu_i(q_2,g)=\mu_i(q_1,g)+o_{K_2}(1),\quad \mu_i(q_2)=\mu_i(q_1)+o_{K_2}(1) \tag{4.3.17}$$

for all $q_2\notin \mathcal{E}_i$ where $\mathcal{E}_i\subseteq Z_2$ is a set of ψ_2- measure $o_{K_2}(1)$. Let $\mathcal{E}=\bigcup_{i=1}^m \mathcal{E}_i$ then $\psi_2(\mathcal{E})=o_{K_2,M}(1)$. By definition, the left hand side of (4.3.15) takes the form

$$\sum_{i=1}^m\left(\frac{\mu_i(q_2,g)}{\mu_i(q_2)}-\frac{\mu_i(q_1,g)}{\mu_i(q_1)}\right)^2\mu_i(q_2), \tag{4.3.18}$$

with the convention that if $\mu_i(q_1)=0$ or $\mu_i(q_2)=0$ then $\mu_i(q_1,g)/\mu_i(q_1):=1$ or $\mu_i(q_2,g)/\mu_i(q_2):=1$.

If $q_2=(q_1,q)\notin\mathcal{E}$ then by (4.3.17)

$$\varepsilon=\varepsilon(N):=\sum_{i=1}^m\Big(|\mu_i(q_2,g)-\mu_i(q_1,g)|+|\mu_i(q_2)-\mu_i(q_1)|\Big)=o_{K_2,M}(1) \tag{4.3.19}$$

We split the sum in (4.3.18) in i into 2 parts:

- If $\mu_i(q_1)\le 2\varepsilon^{1/4}$ then $\mu_i(q_2)\le 3\varepsilon^{1/4}$ by (4.3.17) and we have the trivial bound

  $$\left(\frac{\mu_i(q_2,g)}{\mu_i(q_2)}-\frac{\mu_i(q_1,g)}{\mu_i(q_1)}\right)^2\le 2^2.$$

  Hence the total contribution of such terms is bounded by $12m\,\varepsilon^{1/4}=o_{K_2,M}(1)$.
- If $\mu_i(q_1)\ge 2\varepsilon^{1/4}$ then $\mu_i(q_2)\ge\varepsilon^{1/4}$, we have the estimate

  $$\begin{aligned}\left|\frac{\mu_i(q_2,g)}{\mu_i(q_2)}-\frac{\mu_i(q_1,g)}{\mu_i(q_1)}\right| &= \left|\frac{(\mu_i(q_2,g)-\mu_i(q_1,g))\mu_i(q_1)-\mu_i(q_1,g)(\mu_i(q_1)-\mu_i(q_2))}{\mu_i(q_1)\mu_i(q_2)}\right|\\ &\le\frac{\epsilon\cdot 2+2\varepsilon}{\mu_i(q_1)\mu_i(q_2)}\le\frac{4\varepsilon}{2\varepsilon^{1/2}}=o_{K_2,M}(1).\end{aligned}$$

This proves (4.3.15). The proof of inequality (4.3.16) proceeds the same way, here one needs to estimate the quantity

$$\sum_{i=1}^m\left|\frac{\mu_i(q_2,g)^2}{\mu_i(q_2)}-\frac{\mu_i(q_1,g)^2}{\mu_i(q_1)}\right|=\sum_{i=1}^m\left|\left(\frac{\mu_i(q_2,g)}{\mu_i(q_2)}\right)^2\mu_i(q_2)-\left(\frac{\mu_i(q_1,g)}{\mu_i(q_1)}\right)^2\mu_i(q_1)\right| \tag{4.3.20}$$

- If $\mu_i(q_1)\le 2\varepsilon^{1/4}$ then $\mu_i(q_2)\le 3\varepsilon^{1/4}$ for $q_2=(q_1,q)\notin\mathcal{E}$. Since we have the trivial bounds $(\mu_i(q_j,g)/\mu_i(q_j))^2\le 1$ for $j=1,2$, the contribution of such terms to the right side of (4.3.20) is trivially estimated by

$$5m\,\varepsilon^{1/4} = o_{M,K_2}(1)$$

– If $\mu_i(q_1) \geq 2\varepsilon^{1/4}$ then $\mu_i(q_2) \geq \varepsilon^{1/4}$, using

$$\left|\left(\frac{\mu_i(q_2,g)}{\mu_i(q_2)}\right)^2 \mu_i(q_2) - \left(\frac{\mu_i(q_1,g)}{\mu_i(q_1)}\right)^2 \mu_i(q_1)\right| = \left|\frac{(\mu_i(q_2,g)^2 - \mu_i(q_1,g)^2)\mu_i(q_1) - \mu_i(q_1,g)^2(\mu_i(q_1) - \mu_i(q_2))}{\mu_i(q_1)\mu_i(q_2)}\right|$$

then proceed as in the proof of (4.3.15),using $|\mu_i(q_2,g)^2 - \mu_i(q_1,g)^2| \leq 2|\mu_i(q_1,g) - \mu_i(q_2,g)|$ we have that these remaining terms are bounded by $8\,\varepsilon^{1/2}$ and (4.3.16) follows.

$\square$

Finally, we need an analogue of the above result when the $\|\cdot\|_{L^2(\mu_{q,e})}$ norm is replaced by the more complicated $\|\cdot\|_{\square_{\mu_{q,e}}}$ norms.

Lemma 4.3.10. *Let $\{\nu_{q_2,f}\}_{q_2 \in Z_2, f \in \mathcal{H}}$ be a well-defined extension of the parametric weight system $\{\nu_{q_1,f}\}_{q_1 \in Z_1, f \in \mathcal{H}}$, of complexity at most K_2. For $q_1 \in Z_1$ and $e \in \mathcal{H}$, let $\mathcal{B}_{q_1,e}$ be a σ-algebra of complexity at most M, for some fixed constant $M > 0$. Then*

$$\|\mathbb{E}_{\mu_{q_2,e}}(g_{q_1}|\mathcal{B}_{q_1,e}) - \mathbb{E}_{\mu_{q_1,e}}(g_{q_1}|\mathcal{B}_{q_1,e})\|_{\square_{\mu_{q_2,e}}} = o_{M,K}(1), \tag{4.3.21}$$

for all $q_2 = (q_1, q) \notin \mathcal{E}$, where $\mathcal{E} = \mathcal{E}(g, \mathcal{B}) \subseteq Z_2$ is a set of measure $\psi_2(\mathcal{E}) = o_{M,K_2}(1)$.

Proof. First we show that for any family of sets $\mathcal{A} = (A_{q_1})_{q_1 \in Z_1}$, $A_{q_1} \subseteq V_e$ there is a set $\mathcal{E}_1 = \mathcal{E}_1(g, \mathcal{A})$ of measure $\psi_2(\mathcal{E}_1) = o_{K_2}(1)$ such that for all $q_2 = (q_1, q) \notin \mathcal{E}_1$ we have

$$\|\mathbf{1}_{A_{q_1}}\|_{\square_{\mu_{q_2,e}}}^{2^{|e|}} \leq \mu_{q_2,e}(A_{q_1}) + o_{K_2}(1). \tag{4.3.22}$$

To see this, first note that for $q_2 = (q_1, q) \in Z_2$ one has

$$\begin{aligned}\|\mathbf{1}_{A_{q_1}}\|_{\square_{\mu_{q_2,e}}}^{2^{|e|}} &\leq \mathbb{E}_{x,p \in V_e} \mathbf{1}_{A_{q_1}}(x)\, \mu_{q_2,e}(x) \prod_{f \subseteq e} \prod_{\omega_f \neq 0} \nu_{q_2,f}(\omega_f(p_f, x_f)) \\ &= \mu_{q_2,e}(A_{q_1}) + E(q_2),\end{aligned}$$

with

$$E(q_2) \leq \mathbb{E}_{x \in V_e} \mu_{q_2,e}(x) |\mathbb{E}_{p \in V_e}(\mathrm{w}(q_2, p, x) - 1)|,$$

where

$$\mathrm{w}(q_2, p, x) = \prod_{f \subseteq e} \prod_{\omega_f \neq 0} \nu_{q_2,f}(\omega_f(p_f, x_f)).$$

Arguing as in the proof of Lemma 4.3.7, we see that

$$\mathbb{E}_{q_2\in Z_2}\mathbb{E}_{x,p,p'\in V_e}\,\psi_2(q_2)d\mu_{q_2,e}(x)\;(\mathrm{w}(q_2,p,x)-1)(\mathrm{w}(q_2,p',x)-1)=o_{M,K_2}(1)$$

and (4.3.22) follows.

Now let $\{B^i_{q_1}\}_{i=1}^m$ $(m=2^M)$ be the atoms of $\mathcal{B}_{q_1,e}$ and define the quantities $\mu_i(q_2,g),\mu_i(q_2),\mu_i(q_1,g),\mu_i(q_1)$ as in Lemma 4.3.9. Using the facts that $\mu_i(q_2,g)=\mu_i(q_1,g)+o_{K_2}(1)$ and $\mu_i(q_2)=\mu_i(q_1)+o_{K_2}(1)$ outside a set of measure $o_{M,K_2}(1)$. Arguing as in Lemma 4.3.9, we obtain

$$\left|\frac{\mu_i(q_2,g)}{\mu_i(q_2)}-\frac{\mu_i(q_1,g)}{\mu_i(q_1)}\right|=o_{M,K_2}(1). \tag{4.3.23}$$

The expression in (4.3.21) is then estimated :

$$\begin{aligned}\left\|\sum_{i=1}^m\left(\frac{\mu_i(q_2,g)}{\mu_i(q_2)}-\frac{\mu_i(q_1,g)}{\mu_i(q_1)}\right)\mathbf{1}_{B^i_{q_1}}\right\|_{\square_{\mu_{q_2,e}}} &\leq\sum_{i=1}^m\left|\frac{\mu_i(q_2,g)}{\mu_i(q_2)}-\frac{\mu_i(q_1,g)}{\mu_i(q_1)}\right|\|\mathbf{1}_{B^i_{q_1}}\|_{\square_{\mu_{q_2,e}}}\\ &\lesssim\sum_{i=1}^m\left|\frac{\mu_i(q_2,g)}{\mu_i(q_2)}-\frac{\mu_i(q_1,g)}{\mu_i(q_1)}\right|\mu_{q_2,e}(B^i_q)^{2-d}+o_{M,K_2}(1),\end{aligned}$$

for $q_2=(q_1,q)\notin\mathcal{E}_1$, where $\mathcal{E}_1=\mathcal{E}_1(\mathcal{B}_{q_1,e},g)$ is a set of measure $o_{M,K_2}(1)$.
Now

$$\sum_{i=1}^m\mu_{q_2,e}(B^i_{q_1})=\mu_{q_2,e}(V_e)=1+o_{K_2}(1),$$

In particular $\mu_{q_2,e}(B^i_{q_1})^{2^{-|e|}}=O_{M,K_2}(1)$ and it follows from (4.3.23) that the above expression (4.3.21) is $o_{M,K_2}(1)$. This completes the proof. □

4.3.2 Symmetrization of Parametric Weight System

In this subsection we prove symmetric property of the parametric weight system.

Definition 4.3.11 (Symmetric of parametric weight system)**.** *For each $e\in\mathcal{H}$, let*

$$\mathcal{L}_{q,e}=\{L^1_e(q,x),...,L^s_e(q,x)\}$$

be a pairwise linearly independent family of linear forms in x variables defined on $V=V_J$, depending on parameters $q\in Z$, such that $supp_x(L^j_e)\subseteq e$. We denote $\mathcal{L}_q=\bigcup_{f\in\mathcal{H}}\mathcal{L}_{q,f}=\bigcup_{e\in\mathcal{H}_d}\mathcal{L}_{q,e}$
We say that the family of forms $\mathcal{L}_q$ is **symmetric** *if $L^j_e(q,x)=L^j_{e'}(q,x)$ for all $q\in Z$, $x\in M=\{x: x_0+\cdots+x_d=0\}$, $e,\,e'\in\mathcal{H}_d$ and $1\leq j\leq s$.*
For any given $e\in\mathcal{H}$ we call $\mathcal{L}_q$ the **symmetrization** *of the family $\mathcal{L}_{q,e}$. The reason and validity of*

this later definition is the content of theorem 4.3.13 below.

Remark 4.3.12. *Recall that our initial family of forms defined in (4.2.5) has this property.*

Theorem 4.3.13. *For a fixed $e \in \mathcal{H}, |e| \leq d$ and pairwise linearly independent family $\mathcal{L}_{q,e}$. There is a unique symmetric family of linear forms $\mathcal{L}_q$ defined above such that $\mathcal{L}_{q,e} = \{L \in \mathcal{L}_q;\ supp_x(L) \subseteq e\}$. Also $\mathcal{L}_{q,e}$ is a pairwise linearly independent*

Proof. First assume $|e| = d$ and suppose we have a family of pairwise linearly independent linear forms $\mathcal{L}_{q,e}$. Let $M = \{x : x_0 + \cdots + x_d = 0\}$ (which is isomorphic to V_e for any $|e| = d$) and $\phi_e : V_e \to M$ be the inverse of the projection π_e restricted to M. Now for any $e' \in \mathcal{H}_d, q \in Z, x \in V_J$, define

$$L^j_{e'}(q, x) := L^j_e(q, \phi_{e'} \circ \pi_{e'}(x)) \tag{4.3.24}$$

Note that $\phi_{e'} \circ \pi_{e'}(x)$ is an isomorphism (the identity map) between M and V_e. Hence $supp_x L^j_{e'} \subseteq e'$. Now if $x \in M$ then $x = \phi_{e'} \circ \pi_{e'}(x)$ hence $L^j_{e'}(q, x) = L^j_e(q, x)$, this shows symmetry of $\mathcal{L}_q$.

Indeed, $\mathcal{L}_{q,e} \subseteq \{L \in \mathcal{L}_q;\ supp_x(L) \subseteq e\}$. Next, we show $\mathcal{L}_{q,e} \supseteq \{L \in \mathcal{L}_q;\ supp_x(L) \subseteq e\}$. Suppose $supp_x L^j_{e'} \subseteq e$ (so $L^j_{e'}(q, \phi_e \circ \pi_e(x))$ is defined) then for all $q \in Z_1, x \in V_J$, then by symmetry property

$$L^j_{e'}(q, x) = L^j_{e'}(q, \phi_e \circ \pi_e(x)) = L^j_e(q, \phi_e \circ \pi_e(x)) = L^j_e(q, x) \tag{4.3.25}$$

Finally, we verify that $\mathcal{L}_q$ is a pairwise linearly independent family by considering the set of variables they depend on. Now all forms in $\mathcal{L}_q$ are constructed via (4.3.24), any two of them are either of the form $L^j_e(q, x_e)$ or depends on different sets of variables, hence must be pairwise linearly independent.

Now suppose $|f| < d$ and we have a family of linear forms $\mathcal{L}_{q,f}$. We choose $|e| = d$ with $f \subseteq e$ and we consider $\mathcal{L}_{q,f}$ as family of forms on V_e. This is independent of the choice of e since (by well-definedness of the system) if $f \subseteq e'$ as well then $L^j_e = L^j_{e'}$ for all $1 \leq j \leq s(f)$ and we can do the symmetrization as above. □

4.4 Regularity Lemma for Parametric Weight Hypergraph

In this section we will prove a decomposition theorem for the functions on our hypergraph where we will exploit the machinery of parametric weight system we developed.

4.4.1 A Koopman-von Neumann Type Decomposition for Parametric Weight System

Let $e \subseteq J$ and let $\mathcal{B}_f$ be a σ-algebra on V_f for $f \in \partial e$, where $\partial e = \{f \subseteq e;\ |f| = |e| - 1\}$ denotes the boundary of the edge e. Let $\mathcal{B} := \bigvee_{f \subseteq \partial e} \mathcal{B}_f$ be the σ-algebra generated by the sets $\pi^{-1}_{ef}(\mathcal{B}_f)$ where

$\pi_{ef} : V_e \to V_f$ is the canonical projection. The atoms of $\mathcal{B}$ are the sets $G = \bigcap_{f \subseteq \partial e} \pi_{ef}^{-1}(G_f)$ with G_f being an atom of $\mathcal{B}_f$. We may interpret G the collection of simplices $x \in V_e$ whose faces x_f are in G_f for all $f \in \partial e$.

The first lemma we will prove says that if there is a large "bad" set Ω of parameters q for which the set $G_{q,e}$ is not sufficiently uniform with respect to the σ-algebra $\bigvee_{f \in \partial e} \mathcal{B}_{q,f}$, then the energy of the set (with respect to the sigma-algebra) will increase by a fixed amount when passing to a well defined extension $\{\mathcal{B}_{q',f}\}$, each has complexity increased at most 1 and $\{\mu_{q',e}\}$ with complexity $O(K)$. This holds for all $q' = (q, p) \in \Omega'$ with positive measure.

Lemma 4.4.1 (Large Box Norm implies Structure and Energy Increment)**.** *For given $e \subseteq J$, $|e| = d'$, let $\{\mu_{q,f}\}_{q \in Z, f \subseteq e}$ be a well-defined family of measures of complexity at most K. For $q \in Z$ let $G_{q,e} \subseteq V_e$ and $\{\mathcal{B}_{q,f}\}_{f \in \partial e}$ be a σ-algebra on V_f.*

Assume that for all $q \in \Omega$, where $\Omega \subseteq Z$ is a set of measure $\psi(\Omega) \geq c_0 > 0$, we have

$$\big\|\mathbf{1}_{G_{q,e}} - \mathbb{E}_{\mu_{q,e}}(\mathbf{1}_{G_{q,e}} | \bigvee_{f \in \partial e} \mathcal{B}_{q,f})\big\|_{\Box_{\mu_{q,e}}}^{2^{d'}} \geq \eta, \tag{4.4.1}$$

for some $\eta > 0$.
Then for N, W sufficiently large with respect to the parameters c_0, η, there exists a well-defined extension $\{\mu_{q',f}\}_{q' \in Z', f \subseteq e}$ of the system $\{\mu_{q,f}\}$ of complexity $K' = O(K)$, and a set $\Omega' \subseteq \Omega \times V_e \subseteq Z' = Z \times V_e$ such that all of the following hold.

1. *(positive measure of parameter) We have*

$$\psi'(\Omega') \geq 2^{-4} c_0^2 \eta^2, \tag{4.4.2}$$

   where ψ' is the measure on the parameter space Z'.

2. *(complexity control) For all $q' = (q, p) \in Z'$ and $f \in \partial e$ there is a $\sigma-$algebra $\mathcal{B}_{q',f} \supseteq \mathcal{B}_{q,f}$ of complexity*

$$\textit{compl}(\mathcal{B}_{q',f}) \leq \textit{compl}(\mathcal{B}_{q,f}) + 1. \tag{4.4.3}$$

3. *(energy increment) For all $q' = (q, p) \in \Omega'$, one has*

$$\big\|\mathbb{E}_{\mu_{q',e}}(\mathbf{1}_{G_{q,e}} | \bigvee_{f \in \partial e} \mathcal{B}_{q',f})\big\|_{L^2(\mu_{q',e})}^2 \;\geq\; \big\|\mathbb{E}_{\mu_{q,e}}(\mathbf{1}_{G_{q,e}} | \bigvee_{f \in \partial e} \mathcal{B}_{q,f})\big\|_{L^2(\mu_{q,e})}^2 + 2^{-2d-5}\, \eta^2, \tag{4.4.4}$$

4. *(probability measure)*

$$\mu_{q',e}(V_e) \leq 2. \tag{4.4.5}$$

Proof. Let

$$g_q := \mathbf{1}_{G_{q,e}} - \mathbb{E}_{\mu_{q,e}}(\mathbf{1}_{G_{q,e}}| \bigvee_{f\in\partial e} \mathcal{B}_{q,f}). \tag{4.4.6}$$

Then by definition in equation (4.3.9) we have for each $q \in \Omega$,

$$\|g_q\|_{\square_{\mu_{q,e}}}^{2^{d'}} = \int_{V_e} \langle g_q, \prod_{f\in\partial e} u_{q,p,f}\rangle_{\mu_{(q,p),e}} d\mu_{q,e}(p) \geq \eta, \tag{4.4.7}$$

where $u_{q,p,f} : V_e \to [-1,1]$ are functions in $x-$variables and $d\mu_{q,e}(p)$ is from terms without x variables. $\{\mu_{(q,p),e}\}_{(q,p)\in Z'}$ is the family of measures defined by

$$\mu_{(q,p),e}(x) = \prod_{f\subseteq e} \prod_{\substack{\omega_f\in\{0,1\}^f \\ \omega_f\neq 0}} \nu_{q,f}(\omega_f(p_f, x_f)).$$

As explained after (4.3.2) the measures $\mu_{(q,p),e}$ are defined by a pairwise independent family of forms $\mathcal{L}_{(q,p),e}$ depending on the parameters $(q,p) \in Z \times V_e$, which is a well-defined extension of the family $\mathcal{L}_{q,e}$ defining the measures $\mu_{q,e}$. It is clear from (4.4.7) that the measure ψ' on Z' has the form $\psi'(q,p) = \mu_{q,e}(p)\,\psi(q)$ where $\psi(q)$ is the product of terms without p variables.

For $q' = (q,p)$, let

$$\Gamma(q,p) := \langle g_q, \prod_{f\in\partial e} u_{q,p,f}\rangle_{\mu_{q,p,f}} \tag{4.4.8}$$

First, we show that there is a set $\Omega_1' \subseteq \Omega \times V_e$ of measure

$$\psi'(\Omega_1') \geq 2^{-3} c_0^2 \eta^2, \tag{4.4.9}$$

such that for every $(q,p) \in \Omega_1'$ one has

$$\Gamma(q,p) \geq \frac{\eta}{4}. \tag{4.4.10}$$

Indeed, by Lemma 4.3.4 we have that $\mu_{q,e}(V_e) = 1 + o_K(1) \leq 2$ for $q \notin \mathcal{E}_1$ where $\mathcal{E}_1 \subseteq \Omega$ is a set of measure $\psi(\mathcal{E}_1) = o_K(1)$. Thus for $q \in \Omega \backslash \mathcal{E}_1 = \Omega_1$ we have by (4.4.7) that

$$\int_{V_e} \mathbf{1}_{\{\Gamma(q,p)\geq \eta/4\}}\Gamma(q,p)d\mu_{q,e}(p) \geq \eta - \int_{V_e} \mathbf{1}_{\{\Gamma(q,p)<\eta/4\}}\Gamma(q,p)d\mu_{q,e}(p) \geq \eta - \frac{\eta}{4}(1+o(1)) \geq \frac{\eta}{2}. \tag{4.4.11}$$

Now we use this fact to bound the L^2-moment of $\Gamma(p,q)$. By (4.4.7) and (4.4.8) we have

$$\Gamma(q,p) = \int_{V_e} g_q(x)\Big(\prod_{f\in\partial e} u_{q,f}\Big)\mathrm{w}_{q,p}(x)d\mu_{q,e}(x)$$

The function $\mathrm{w}_{q,p}(x)$ is the product of weight functions of the form $\nu(L(q,p,x))$ depending on both p and x. Thus, using the bounds $|g_q|\leq 1, |u_{q,p,f}|\leq 1$, one has

$$\int_Z\int_{V_e} |\Gamma(q,p)|^2 d\mu_{q,e}(p)d\psi(q) \leq \int_Z\int_{V_e}\int_{V_e}\int_{V_e} \mathrm{w}_{q,p}(x)\mathrm{w}_{q,p}(x')d\mu_{q,e}(x)d\mu_{q,e}(x')d\mu_{q,e}(p)d\psi(q) \tag{4.4.12}$$

$$= 1 + o_K(1) \leq 2$$

by the linear forms condition as the factors in the product depend on different sets of variables. Let $\Omega_1' := \{(q,p)\in\Omega_1\times V_e;\ \Gamma(q,p)\geq \eta/4\}$. Thus by (4.4.11) and (4.4.12) over Ω_1 with the Cauchy-Schwartz inequality,

$$\frac{c_0^2\eta^2}{4} \leq \frac{\psi(\Omega_1)^2\eta^2}{4} \leq \left(\int_{\Omega_1'} \Gamma(q,p)d\mu_{q,e}(p)\,d\psi(q)\right)^2 \leq \int_{\Omega_1'} \Gamma(q,p)^2 d\mu_{q,e}(p)\,d\psi(q)\,\psi'(\Omega_1') \leq 2\,\psi'(\Omega_1').$$

This shows $\psi'(\Omega_1')\geq 2^{-3}c_0^2\eta^2$ as claimed.

Since $|u_{q',f}|\leq 1$, decomposing of each function $u_{q',f}$ into its positive and negative parts in (4.4.7) yields that

$$\int_{V_e}\langle g_q, \prod_{f\in\partial e} v_{q',f}\rangle_{\mu_{q',e}} d\mu_{q,e}(p) \geq 2^{-d'}\eta \geq 2^{-d}\eta$$

i.e.

$$\langle g_q, \prod_{f\in\partial e} v_{q',f}\rangle_{\mu_{q',e}} \geq 2^{-d'}\eta \geq 2^{-d}\eta \tag{4.4.13}$$

for some q' and functions[7] $v_{q',f}: V_f \to [0,1]$. Now we obtained ***correlation with structures*** and we

[7]For each fixed f, take $v_{q',f}(x) = \max u^+_{q',f}$ where $u^+_{q',f}$ denotes nonnegative terms in the decomposition and the maximum is taken over these nonnegative terms.

will pass from structural functions to sets. For a given $f \in \partial e$ and some $0 \le t_f \le 1$, let

$$\mathcal{U}_{q',t_f} := \{x_f \in V_f : v_{q',f}(x_f) \ge t_f\}$$

be the level set of the functions $v_{q',f}$. That is, for each fixed x_f,

$$F(t_f) := \mathbf{1}_{\mathcal{U}_{q',t_f}}(x_f) \Rightarrow F(t_f) = \begin{cases} 1 & \text{if} \quad t_f \le v_{q',f}(x_f) \\ 0 & \text{if} \quad t_f > v_{q',f}(x_f) \end{cases}$$

Then $v_{q',f}(x_f) = \int_0^1 F(t_f)dt_f = \int_0^1 \mathbf{1}_{\mathcal{U}_{q',t_f}}(x_f)dt_f$, and for each term in (4.4.13) we have by swapping the integrals,

$$\int_0^1 \cdots \int_0^1 \langle g_q, \prod_{f\in\partial e} \mathbf{1}_{\mathcal{U}_{q',t_f}} \rangle_{\mu_{q',e}} dt \ge 2^{-d}\eta,$$

where $t = (t_f)_{f\in\partial e}$. By pigeonhole principle the integrand must be at least $2^{-d}\eta$ for some value of the parameter t.

Fix such a $t = (t_f) \in [0,1]^d$ and write $\mathcal{U}_{q',f}$ for $\mathcal{U}_{q',t_f}$ for simplicity of notation. For $q' = (q,p) \in \Omega_1'$, define $\mathcal{B}_{q',f}$ to be the $\sigma-$algebra generated by $\mathcal{B}_{q,f}$, and the $\mathcal{U}_{q',f}$. For $q' \notin \Omega_1'$, set $\mathcal{B}_{q',f} = \mathcal{B}_{q,f}$. The function $\prod_{f\in\partial e} \mathbf{1}_{\mathcal{U}_{q',f}}$ is constant on the atoms of the $\sigma-$algebra $\bigvee_{f\in\partial e} \mathcal{B}_{q',f}$, and therefore for $q' \in \Omega_1'$

$$\big\langle \mathbf{1}_{G_{q,e}} - \mathbb{E}_{\mu_{q',e}}(\mathbf{1}_{G_{q,e}} | \bigvee_{f\in\partial e} \mathcal{B}_{q',f}), \prod_{f\in\partial e} \mathbf{1}_{\mathcal{U}_{q',f}} \big\rangle_{\mu_{q',e}} = 0$$

for $q' \in \Omega_1'$. Hence, by (4.4.6) and (4.4.13) it follows that

$$\langle \mathbb{E}_{\mu_{q',e}}(\mathbf{1}_{G_{q,e}} | \bigvee_{f\in\partial e} \mathcal{B}_{q',f}) - \mathbb{E}_{\mu_{q,e}}(\mathbf{1}_{G_{q,e}} | \bigvee_{f\in\partial e} \mathcal{B}_{q,f}), \prod_{f\in\partial e} \mathbf{1}_{\mathcal{U}_{q',f}} \rangle_{\mu_{q',e}} \ge 2^{-d}\eta \tag{4.4.14}$$

By Lemma 4.3.4 there is a set $\mathcal{E}_1 \subseteq Z'$ such that $\psi'(\mathcal{E}_1) = o_K(1)$ and

$$\| \prod_{f\in\partial e} \mathbf{1}_{\mathcal{U}_{q',f}} \|_{L^2(\mu_{q',e})} \le \mu_{q',e}(V_e)^{1/2} = 1 + o_K(1) \le 2$$

for $q' \in \Omega_1' \backslash \mathcal{E}_1 =: \Omega_2'$. Then apply Cauchy-Schwartz inequality to (4.4.14),

$$\|\mathbb{E}_{\mu_{q',e}}(\mathbf{1}_{G_{q,e}} | \bigvee_{f\in\partial e} \mathcal{B}_{q',f}) - \mathbb{E}_{\mu_{q,e}}(\mathbf{1}_{G_{q,e}} | \bigvee_{f\in\partial e} \mathcal{B}_{q,f})\|_{L^2(\mu_{q',e})} \ge 2^{-d-1}\eta,$$

for $q' \in \Omega_2'$. By (4.3.15) in Lemma 4.3.9 there is an exceptional set $\mathcal{E}_2 \subseteq Z'$ of measure $\psi'(\mathcal{E}_2) = o_{K,M}(1)$ such that for $q' = (q,p) \in \Omega_3' := \Omega_2' \backslash \mathcal{E}_2$ we have

$$\big\|\mathbb{E}_{\mu_{q',e}}(\mathbf{1}_{G_{q,e}}| \bigvee_{f\in\partial e} \mathcal{B}_{q',f}) - \mathbb{E}_{\mu_{q',e}}(\mathbf{1}_{G_{q,e}}| \bigvee_{f\in\partial e} \mathcal{B}_{q,f})\big\|_{L^2(\mu_{q',e})} \geq 2^{-d-1}\eta - o_{K,M}(1) \geq 2^{-d-2}\eta. \tag{4.4.15}$$

Since $\mathcal{B}_{q,f} \subseteq \mathcal{B}_{q',f}$, for $q' = (q,p)$, applying Pythagorus's Theorem, (4.4.15) is equivalent to

$$\big\|\mathbb{E}_{\mu_{q',e}}(\mathbf{1}_{G_{q,e}}| \bigvee_{f\in\partial e} \mathcal{B}_{q',f})\big\|^2_{L^2(\mu_{q',e})} - \big\|\mathbb{E}_{\mu_{q',e}}(\mathbf{1}_{G_{q,e}}| \bigvee_{f\in\partial e} \mathcal{B}_{q,f})\big\|^2_{L^2(\mu_{q',e})} \geq 2^{-2d-4}\eta^2. \tag{4.4.16}$$

Finally, an invocation of (4.3.16) in Lemma 4.3.9 there is a set $\mathcal{E}_3 \subseteq Z'$ of measure $\psi'(\mathcal{E}_3) = o_{K,M}(1)$ such that for $q' \in \Omega_4' := \Omega_3' \backslash \mathcal{E}_3$ we have (for N, W sufficiently large)

$$\big\|\mathbb{E}_{\mu_{q',e}}(\mathbf{1}_{G_{q,e}}| \bigvee_{f\in\partial e} \mathcal{B}_{q',f})\big\|^2_{L^2(\mu_{q',e})} - \big\|\mathbb{E}_{\mu_{q,e}}(\mathbf{1}_{G_{q,e}}| \bigvee_{f\in\partial e} \mathcal{B}_{q,f})\big\|^2_{L^2(\mu_{q,e})} \geq 2^{-2d-5}\eta^2. \tag{4.4.17}$$

This proves the lemma choosing $\Omega' = \Omega_4'$. □

Now for any given $e \in \mathcal{H}$, we shall prove a Koopman-von Neumann type decomposition for $\mathbf{1}_{G_e}$ for any $G_e \in \mathcal{B}_{q,e}$. The will be done via iterations argument; repeated applications of lemma 4.4.1 and boundedness of the total energy of the hypergraph system. The *total energy of the family $\{\mathcal{B}_{q,e}\}_{e\in\mathcal{H}_{d'}}$ with respect to a family of lower order σ-algebras $\{\mathcal{B}_{q,f}\}_{f\in\mathcal{H}_{d'-1}}$ and a family of measures $\{\mu_{q,e}\}_{e\in\mathcal{H}_{d'}}$* is the quantity

$$\mathcal{E}_{d'}(\{\mathcal{B}_{q,f}\}_{f\in\mathcal{H}_{d'-1}}) = \sum_{e\in\mathcal{H}_{d'},G_e\in\mathcal{B}_e} \big\|\mathbb{E}_{\mu_e}(\mathbf{1}_{G_e}| \bigvee_{f\in\partial e} \mathcal{B}_{q,f})\big\|^2_{L^2(\mu_{q,e})} \leq 2\binom{d}{d'}2^{2^{M_{d'}}}. \tag{4.4.18}$$

And the total energy of the hypergraph system is

$$\mathcal{E}(\{\mathcal{B}_{q,f}\}_{f\in\mathcal{H}_{d'-1}}) := \sum_{1\leq d'\leq d} \mathcal{E}_{d'}(\{\mathcal{B}_{q,f}\}_{f\in\mathcal{H}_{d'-1}}) \tag{4.4.19}$$

Assuming the measures μ_e are normalized i.e. $\mu_e(V_e) = 1 + o(1) \leq 2$, a crude upper bound for the $\mathcal{E}$ are $2^{d+1}2^{2^M} = O_M(1)$ is a universal bound, where M is the complexity of the σ-algebras $\bigvee_{f\in\mathcal{H}} \mathcal{B}_{q,f}$.

Lemma 4.4.2 (Koopman-von Neumann decomposition for Parametric Weight System)**.** *Let $\{\mu_{q,f}\}_{q\in Z, f\in\mathcal{H}}$ be a well-defined, symmetric family of measures of complexity at most K. Let $1 \le d' \le d$, and let $\{\mathcal{B}_{q,e}\}_{q\in Z, e\in\mathcal{H}_{d'}}$ and $\{\mathcal{B}_{q,f}\}_{q\in Z, f\in\mathcal{H}_{d'-1}}$ be families of σ-algebras of complexity at most $M_{d'}$ and $M_{d'-1}$. Finally let $\Omega \subseteq Z$ with $\psi(\Omega) \ge c_0 > 0$, and let $\delta > 0$ be a constant. (In the iteration process, these quantities in the assumption are obtained from the previous step of the iteration.)*

Then for N, W sufficiently large with respect to the constants $\delta, c_0, M_{d'}, M_{d'-1}$ and K, $Z' = Z \times V$, there exists a well-defined, symmetric extension $\{\mu_{q',f}\}_{q'\in Z', f\in\mathcal{H}}$ of the system $\{\mu_{q,f}\}_{q\in Z, f\in\mathcal{H}}$ of complexity at most $K' = O_{M_{d'},K,\delta}(1)$ and a family of σ-algebras $\{\mathcal{B}_{q',f}\}_{q'\in Z', f\in\mathcal{H}_{d'-1}}$ such that the following hold.

1. *For all $q' = (q,p) \in Z'$ and $f \in \mathcal{H}_{d'-1}$ we have*

$$\mathcal{B}_{q,f} \subseteq \mathcal{B}_{q',f}, \quad compl(\mathcal{B}_{q',f}) \le compl(\mathcal{B}_{q,f}) + O_{M_{d'},\delta}(1). \tag{4.4.20}$$

2. *There exists a set $\Omega' \subseteq \Omega \times V \subseteq Z'$ of measure $\psi'(\Omega') \ge c(c_0, \delta, M_{d'}) > 0$ such that for all $q' = (q,p) \in \Omega'$ and for all $e \in \mathcal{H}_{d'}$, for all $G_{q,e} \in \mathcal{B}_{q,e}$ one has*

$$\big\|\mathbf{1}_{G_{q,e}} - \mathbb{E}_{\mu_{q',e}}(\mathbf{1}_{G_{q,e}} | \bigvee_{f\in\partial e} \mathcal{B}_{q',f})\big\|_{\square_{\mu_{q',e}}} \le \delta. \tag{4.4.21}$$

   and the stability property

$$\big\|\mathbb{E}_{\mu_{q',e}}(\mathbf{1}_{G_{q,e}} | \bigvee_{f\in\partial e} \mathcal{B}_{q,f})\big\|^2_{L^2(\mu_{q',e})} = \big\|\mathbb{E}_{\mu_{q,e}}(\mathbf{1}_{G_{q,e}} | \bigvee_{f\in\partial e} \mathcal{B}_{q,f})\big\|^2_{L^2(\mu_{q,e})} + o_{M_{d'},K,\delta}(1), \tag{4.4.22}$$

Proof. Initially set $Z' = Z$, then (4.4.20) and (4.4.22) trivially holds for $q' = q$. If there is a set $\Omega_1 \subseteq \Omega$ of measure $\psi(\Omega_1) \ge \frac{c_0}{2}$ such that inequality (4.4.21) holds for all $q \in \Omega_1$ and $G_{q,e} \in \mathcal{B}_{q,e}$ then the conclusions of the lemma hold for the initial system of measures and σ-algebras

$$(\{\mu_{q,f}\}_{q\in Z, f\in\mathcal{H}}, \{\mathcal{B}_{q,e}\}_{q\in Z, e\in\mathcal{H}_{d'}}, \{\mathcal{B}_{q,f}\}_{,q\in Z f\in\mathcal{H}_{d'-1}})$$

and the set Ω_1.

Otherwise, for all sets $\Omega_2 \subseteq \Omega$ of measure $\psi(\Omega_2) \ge \frac{c_0}{2}$ such that for each $q \in \Omega_2$ there is an $e \in \mathcal{H}_{d'}$ and a set $G_{q,e} \in \mathcal{B}_{q,e}$ for which the inequality (4.4.21) fails. Fix one of such Ω_2. By the pigeonholing, up to a factor of $\binom{d}{d'}$, we may assume that there is an $e \in \mathcal{H}_{d'}$ that (4.4.21) fails for all q. Then by Lemma 4.4.1, with $\eta := \delta^{2^{d'}}$, there is a well-defined extension $\{\mu_{q',f}\}_{q'\in Z', f\subseteq e}$, a family of σ-algebras $\{\mathcal{B}_{q',f}\}_{q'\in Z', f\in\partial e}$ and a set $\Omega' \subseteq \Omega_2$ of positive measure for which (4.4.2)-(4.4.4) hold.

Let $\{\mu_{q',f}\}_{q'\in Z', f\in\mathcal{H}}$ be the *symmetrization* of the system $\{\mu_{q',f}\}_{q'\in Z', f\subseteq e}$ as described in section

4.3.2, and set $\mathcal{B}_{q',f} := \mathcal{B}_{q,f}$ for $q' \notin \Omega'$ or $f \in \mathcal{H}_{d'-1}$, $f \nsubseteq e$ or $f \in \mathcal{H}_{d'}$. By Lemma 4.3.9 and Lemma 4.4.1 one may remove a set $\mathcal{E}$ of measure $\psi'(\mathcal{E}) = o_{M_{d'},K}(1)$ such that for all $q' \in \Omega' \backslash \mathcal{E}$, (4.4.20) and (4.4.22) hold for the extended system $(\{\mu_{q',f}\}_{q' \in Z', f \in \mathcal{H}}, \{\mathcal{B}_{q',e}\}_{q' \in Z', e \in \mathcal{H}_{d'}}, \{\mathcal{B}_{q',f}\}_{,q' \in Z', f \in \mathcal{H}_{d'-1}})$, whose total energy is at least $2^{-2d-5}\delta^{2^{d'}}$ larger than that of the initial system.

Based on the above argument we perform the following iteration. Let $\{\mu_{q',f}\}_{q' \in Z', f \in \mathcal{H}}$ be a well-defined, symmetric extension of the initial system $\{\mu_{q,f}\}_{q \in Z, f \in \mathcal{H}}$. Let $\{\mathcal{B}_{q',f}\}_{q' \in Z', f \in \mathcal{H}_{d'-1}}$ be a family of σ-algebras and let $\Omega' \subseteq \Omega \times V' \subseteq Z'$ for which (4.4.20) and (4.4.22) hold. If there is a set $\Omega'_1 \subseteq \Omega'$ of measure $\psi'(\Omega'_1) \geq \psi(\Omega')/2$ such that for all $q \in \Omega'_1$, $e \in \mathcal{H}_{d'}$ and $G_{q,e} \in \mathcal{B}_{q,e}$ inequality (4.4.21) holds, then the system $(\{\mu_{q',f}\}_{q' \in Z', f \in \mathcal{H}}, \{\mathcal{B}_{q',e}\}_{q' \in Z', e \in \mathcal{H}_{d'}}, \{\mathcal{B}_{q',f}\}_{,q' \in Z', f \in \mathcal{H}_{d'-1}})$ together with the set Ω'_1 satisfies the conclusions of the lemma. (Note that the family of sigma-algebras $\mathcal{B}_{q',e}$ is unchanged.)

Otherwise there is a well-defined, symmetric extension $\{\mu_{q'',f}\}_{q'' \in Z'', f \in \mathcal{H}}$ together with a family of σ-algebras $\{\mathcal{B}_{q'',f}\}_{q'' \in Z'', f \in \mathcal{H}_{d'-1}}$ and a set $\Omega'' \subseteq \Omega' \times \mathbb{Z}_N^{d'}$ such that for all $q'' \in \Omega''$ inequalities (4.4.20) and (4.4.22) hold, and total energy of the system $(\mu_{q'',e}, \mathcal{B}_{q'',e}, \mathcal{B}_{q'',f})$ is at least $2^{-2d-6}\delta^{2^{d'}}$ larger than that of the system $(\mu_{q',e}, \mathcal{B}_{q',e}, \mathcal{B}_{q',f})$. Set $Z' := Z$, $\mu_{q',e} := \mu_{q'',e}$ and $\mathcal{B}_{q',f} := \mathcal{B}_{q'',f}$. then return to previous step.

As (4.4.18) is bounded by an absolute constant (as $\mathcal{B}_{q,e}, e \in \mathcal{H}_d$ is never changed in the iteration) the iteration process must stop in $O_{M_{d'},\delta}(1)$ steps and the system obtained from the last step satisfies (4.4.20)-(4.4.22).

□

4.4.2 Regularity Lemma

The shortcoming of Lemma 4.4.2 is that the complexity of the σ-algebras $\mathcal{B}_{q,f}$ might be very large with respect to the parameter δ, which measures the uniformity of the graphs $G_{q,e}$. Hence it is not a good tool to describe the structure. This issue can be taken care of with an iteration process using Lemma 4.4.2 repeatedly, along the lines it was done in [104]. In the weighted settings we have to pass to a new system of weights and measures at each iteration and have to exploit the stability properties of well-defined extensions to show that the iteration process terminates.

In the first step, we will prove *Preliminary Regularity Lemma* which regularize only graphs on $\mathcal{H}_{d'}$ for a fixed d', as in Lemma 4.4.2. Then we prove a full regularity lemma which regularizes simultaneously all elements of the hypergraph.

Remark 4.4.3. *In our energy increment process (last subsection), $\mathcal{B}_{q,e}$ is not changed in each step of the iteration. So it is okay to write system as e.g.$(\mu_{q',f}, \mathcal{B}_{q,e}, \mathcal{B}_{q',f})$.*

Remark 4.4.4. *In regularity lemma below, to obtain the extreme uniformity like (4.4.28), we would need a pair of sigma algebras $\mathcal{B}_{\overline{q},f}, \mathcal{B}'_{\overline{q},f}$ which are close in L^2 norm in our decomposition. $\mathcal{B}'_{\overline{q},f}$ itself would not be able to play a role as the structure part since it has complexity $O_{M_{d'-1},\delta}(1) = O_{M_{d'-1},F}(1)$, due to choice of δ which we have no control as F can be chosen to be arbitrarily fast growing.*

This lemma, as a regularity lemma, is more widely applicable than Lemma 4.4.2 as the uniformity of the hypergraphs $G_{q,e}$ with respect to the (fine) $\sigma-$algebras $B'_{\overline{q},f}$ can be chosen to be arbitrarily small with respect to the complexity of the (coarse) $\sigma-$algebras $\mathcal{B}_{\overline{q},f}$, while the approximations $\mathbb{E}_{\mu_{\overline{q},e}}(\mathbf{1}_{G_{q,e}} | \bigvee \mathcal{B}'_{\overline{q},f})$ and $\mathbb{E}_{\mu_{\overline{q},e}}(\mathbf{1}_{G_{q,e}} | \bigvee \mathcal{B}_{\overline{q},f})$ stay very close in $L^2(\mu_{\overline{q},e})$. First, we start by regularizing hyperedges in a given $\mathcal{H}_{d'}$ for some $1 \leq d' \leq d$.

Lemma 4.4.5 (Preliminary regularity lemma.)**.** *Let $1 \leq d' \leq d$ and $M_{d'} > 0$ be a constant. Let $\{\mu_{q,f}\}_{q \in Z, f \in \mathcal{H}}$ be a well-defined, symmetric family of measures of complexity at most K, and $1 \leq d' \leq d$ and $\{\mathcal{B}_{q,e}\}_{q \in Z, e \in \mathcal{H}_{d'}}$ be a family of $\sigma-$algebras on V_e so that for all $q \in Z,\ e \in \mathcal{H}_{d'}$*

$$compl\,(\mathcal{B}_{q,e}) \leq M_{d'}. \tag{4.4.23}$$

Let $\varepsilon > 0$ and $F : \mathbb{R}_+ \to \mathbb{R}_+$ be a non-negative, increasing function, possibly depending on ε and $\Omega \subseteq Z$ be a set of measure $\psi(\Omega) \geq c_0 > 0$.

If N, W is sufficiently large with respect to the parameters $\varepsilon, c_0, M_{d'}, K$, and F, then there exists a well-defined, symmetric extension $\{\mu_{\overline{q},f}\}_{\overline{q} \in \overline{Z}, f \in \mathcal{H}}$ of complexity at most $O_{K,M_{d'},F,\varepsilon}(1)$, and families of σ-algebras $\mathcal{B}_{\overline{q},f}, \mathcal{B}'_{\overline{q},f}$, $\mathcal{B}_{\overline{q},f} \subseteq \mathcal{B}'_{\overline{q},f}$ defined for $\overline{q} \in \overline{Z},\ f \in \mathcal{H}_{d-1}$ and a set $\overline{\Omega} \subseteq \overline{Z}$ such that the following holds.

1. *We have that $\overline{\Omega} \subseteq \Omega \times \overline{V} \subseteq \overline{Z} = Z \times \overline{V}$ where $\overline{V} = \mathbb{Z}_N^k$ of dimension $k = O_{M_{d'},F,\varepsilon}(1)$. Moreover*
$$\overline{\psi}(\overline{\Omega}) \geq c(c_0, F, M_{d'}, \varepsilon) > 0. \tag{4.4.24}$$

2. *There is a constant $M_{d'-1}$ such that*
$$F(M_{d'}) \leq M_{d'-1} = O_{M_{d'},F,\varepsilon}(1) \tag{4.4.25}$$
and for all $\overline{q} \in \overline{Z}$ and $f \in \mathcal{H}_{d'-1}$ we have
$$compl(\mathcal{B}'_{\overline{q},f}) \leq M_{d'-1}. \tag{4.4.26}$$

3. *For all $\overline{q} = (q, p) \in \overline{\Omega},\ e \in \mathcal{H}_{d'}$ and $G_{q,e} \in \mathcal{B}_{q,e}$, we have*

$$\big\|\mathbb{E}_{\mu_{\overline{q},e}}(\mathbf{1}_{G_{q,e}}|\bigvee_{f\in\partial e}\mathcal{B}'_{\overline{q},f})-\mathbb{E}_{\mu_{\overline{q},e}}(\mathbf{1}_{G_{q,e}}|\bigvee_{f\in\partial e}\mathcal{B}_{\overline{q},f})\big\|_{L^2(\mu_{\overline{q},e})}\le\varepsilon \tag{4.4.27}$$

and

$$\big\|\mathbf{1}_{G_{q,e}}-\mathbb{E}_{\mu_{\overline{q},e}}(\mathbf{1}_{G_{q,e}}|\bigvee_{f\in\partial e}\mathcal{B}'_{\overline{q},f})\big\|_{\square_{\mu_{\overline{q},e}}}\le\frac{1}{F(M_{d'-1})}. \tag{4.4.28}$$

Proof. Let $\{\mu_{q',f}\}_{q'\in Z',\,f\in\mathcal{H}}$ be a well-defined, symmetric extension of the initial system $\{\mu_{q,f}\}$ defined on a parameter space $Z'=Z\times V'$ of complexity at most K'. We start by putting trivial sigma-algebra $\mathcal{B}_{q',f}=\{\emptyset,V_J\}$ on each $f\in\mathcal{H}_{d'-1}$. Set

$$M_{d'-1}:=\max\{F(M_{d'}),\sup_{f\in\partial\mathcal{H}_d}compl(\mathcal{B}_{q',f})\}=O_{K,M_{d'},F}(1),\quad \delta:=\frac{1}{F(M_{d'-1})} \tag{4.4.29}$$

Indeed the point here is that $M_{d'-1}$ (and later $M_{d'-1}$) is $O_{K,M_{d'},F}(1)$. Set $\mathcal{B}_{q',e}:=\mathcal{B}_{q,e}$ for $q'=(q,p)\in Z'$, $e\in\mathcal{H}_{d'}$, and apply Lemma 4.4.2 to the system $(\mu_{q',e},\,\mathcal{B}_{q',e},\,\mathcal{B}_{q',f})$, with $\delta=F(M_{d'-1})^{-1}$. This generates a well-defined, symmetric extension $\{\mu_{\overline{q},f}\}_{\overline{q}\in\overline{Z},f\in\mathcal{H}}$ and a family of σ-algebras $\{\mathcal{B}'_{\overline{q},f}\}_{\overline{q}\in\overline{Z},f\in\mathcal{H}_{d'-1}}$ and a set $\overline{\Omega}\subseteq\overline{Z}$ satisfying the conclusion of that Lemma.

The new system $(\mu_{\overline{q},f},\mathcal{B}_{\overline{q},e},\mathcal{B}'_{\overline{q},f})$ satisfies (4.4.24)-(4.4.26) and (4.4.28). Note that the parameters K', $M_{d'-1}$ are of magnitude $O_{K,M_{d'},F,\varepsilon}(1)$. Set $\mathcal{B}_{\overline{q},f}:=\mathcal{B}_{q',f}$ for $\overline{q}=(q',p)\in\overline{Z}$, $f\in\mathcal{H}_{d'-1}$.

To ensure L^2-closeness property, we run the energy increment. There are two possibilities.

- *Case 1:* There exists a set $\overline{\Omega}_1\subseteq\overline{\Omega}$ of measure $\overline{\psi}(\overline{\Omega}_1)\ge\overline{\psi}(\overline{\Omega})/2$ such that (4.4.27) holds for all $\overline{q}\in\overline{\Omega}_1$. In this case the conclusions of the lemma hold for the system $(\mu_{\overline{q},e},\mathcal{B}_{\overline{q},e},\mathcal{B}'_{\overline{q},f})$ and the set $\overline{\Omega}_1$.
- *Case 2:* For every $\overline{\Omega}_1\subseteq\overline{\Omega},\overline{\psi}(\overline{\Omega}_1)\ge\overline{\psi}(\overline{\Omega})/2$, we have (4.4.27) fails for some $\overline{q}\in\overline{\Omega}_1$. Let $\overline{\Omega}_2:=\{\overline{q}\in\overline{\Omega}:(4.4.27)\text{ fails}\}$. Then $\overline{\Omega}_2\subseteq\overline{\Omega}$ is of measure $\overline{\psi}(\overline{\Omega}_2)\ge\frac{1}{2}\overline{\psi}(\overline{\Omega})$. Now, thanks to the stability condition (4.4.22) and the fact that $\mathcal{B}_{q',f}=\mathcal{B}_{\overline{q},f}\subseteq\mathcal{B}'_{\overline{q},f}$, we have for $\overline{q}\in\overline{\Omega}_2$, $q'=\pi'(\overline{q})$, and $q=\pi(\overline{q})$ where $\pi:\overline{Z}\to Z,\pi':\overline{Z}\to Z'$ are projections, we have that

$$\begin{aligned}\mathcal{E}_{d'}(\mathcal{B}'_{\overline{q},f})-\mathcal{E}_{d'}(\mathcal{B}_{\overline{q},f})&=\sum_{e,G_{q,e}}\big\|\mathbb{E}_{\mu_{\overline{q},e}}(\mathbf{1}_{G_{q,e}}|\bigvee_{f\in\partial e}\mathcal{B}'_{\overline{q},f})\big\|^2_{L^2_{\mu_{\overline{q},e}}}-\sum_{e,G_{q,e}}\big\|\mathbb{E}_{\mu_{q',e}}(\mathbf{1}_{G_{q,e}}|\bigvee_{f\in\partial e}\mathcal{B}_{q',f})\big\|^2_{L^2_{\mu_{q',e}}}\\&\ge\sum_{e,G_{q,e}}\Big(\big\|\mathbb{E}_{\mu_{\overline{q},e}}(\mathbf{1}_{G_{q,e}}|\bigvee_{f\in\partial e}\mathcal{B}'_{\overline{q},f})\big\|^2_{L^2_{\mu_{\overline{q},e}}}-\big\|\mathbb{E}_{\mu_{\overline{q},e}}(\mathbf{1}_{G_{q,e}}|\bigvee_{f\in\partial e}\mathcal{B}_{q',f})\big\|^2_{L^2_{\mu_{\overline{q},e}}}\Big)-o_{M_{d'},K',F}(1)\end{aligned}$$

$$= \sum_{e,G_{q,e}} \|\mathbb{E}_{\mu_{\overline{q},e}}(\mathbf{1}_{G_{q,e}}| \bigvee_{f\in\partial e} \mathcal{B}'_{\overline{q},f}) - \mathbb{E}_{\mu_{\overline{q},e}}(\mathbf{1}_{G_{q,e}}| \bigvee_{f\in\partial e} \mathcal{B}_{\overline{q},f})\|^2_{L^2_{\mu_{\overline{q},e}}} - o_{M_{d'},K',F}(1)$$

$$\geq \varepsilon^2 - o_{M_{d'},K',F}(1), \tag{4.4.30}$$

where the summation is taken over all $e \in \mathcal{H}_{d'}$ and $G_{q,e} \in \mathcal{B}_{q,e}$.

Thus, for sufficiently large N, W, we have for all $\overline{q} = (q,p) \in \overline{\Omega}_2$ that the total energy of the system $(\mu_{\overline{q},f},\ \mathcal{B}_{\overline{q},e},\ \mathcal{B}'_{\overline{q},f})$ is at least $\frac{\varepsilon^2}{2}$ larger than that of the system $(\mu_{q',f},\ \mathcal{B}_{q',e},\ \mathcal{B}_{q',f})$. In this case, set $Z' := \overline{Z}$, $\Omega' := \overline{\Omega}_3$, $\mu_{q',f} := \mu_{\overline{q},f}$, and $\mathcal{B}_{q',f} := \mathcal{B}'_{\overline{q},f}$ and repeat the above argument. The iteration process must stop in at most $\varepsilon^{-2}2^{2^{(M_{d'})+1}}2^{d+1} = O_{M_{d'},\varepsilon}(1)$ steps, generating a system $(\mu_{\overline{q},f}, \mathcal{B}_{\overline{q},e}, \mathcal{B}'_{\overline{q},f})$ which satisfies the conclusions of the lemma. □

In order to obtain a counting and a removal lemma starting from a given measure system $\{\mu_{q,e}\}$ and σ-algebras $\{\mathcal{B}_{q,e}\}$ we need to regularize the elements of the σ-algebras $\mathcal{B}_{\overline{q},e}$ for *all* $e \in \mathcal{H}$ with respect to its lower order σ-algebras $\bigvee_{f\in\partial e} \mathcal{B}_{q,f}$. This is done by applying Lemma 4.4.5 inductively, and provides the final form of the regularity lemma we need. Let us call a function $F : \mathbb{R}_+ \to \mathbb{R}_+$ a *growth function* if it is continuous, increasing, and satisfies[8] $F(x) \geq 1 + x$ for $x \geq 0$.

Theorem 4.4.6. *[Full Regularity lemma.] Let $1 \leq d' \leq d$ and $M_{d'} > 0$ be a constant. Let $\{\mu_{q,f}\}_{q\in Z, f\in\mathcal{H}}$ be a well-defined, symmetric family of measures of complexity at most K, and $\{\mathcal{B}_{q,e}\}_{q\in Z, e\in\mathcal{H}_{d'}}$ be a family of $\sigma-$algebras on V_e so that for all $q \in Z,\ e \in \mathcal{H}_{d'}$*

$$\textit{compl}\,(\mathcal{B}_{q,e}) \leq M_{d'}. \tag{4.4.31}$$

Let $F : \mathbb{R}_+ \to \mathbb{R}_+$ be a growth function, and $\Omega \subseteq Z$ be a set of measure $\psi(\Omega) \geq c_0 > 0$.

If N, W is sufficiently large with respect to the parameters $c_0, M_{d'}, K$, and F, then there exists a well-defined, symmetric extension $\{\mu_{\overline{q},f}\}_{\overline{q}\in\overline{Z}, f\in\mathcal{H}}$ of complexity at most $O_{K,M_{d'},F}(1)$ on a parametric space $\overline{Z}$, and families of σ-algebras $\mathcal{B}_{\overline{q},f} \subseteq \mathcal{B}'_{\overline{q},f}$ defined for $\overline{q} \in \overline{Z},\ f \in \mathcal{H}_{d'-1}$ and a set $\overline{\Omega} \subseteq \overline{Z}$ such that the following holds.

1. *We have that $\overline{\Omega} \subseteq \Omega \times \overline{V} \subseteq \overline{Z} = Z \times \overline{V}$ where $\overline{V} = \mathbb{Z}_N^k$ of dimension $k = O_{M_{d'},F}(1)$. Moreover*

$$\overline{\psi}(\overline{\Omega}) \geq c(c_0, F, M_{d'}) > 0. \tag{4.4.32}$$

2. *There exist numbers*

$$M_{d'} < F(M_{d'}) \leq M_{d'-1} < F(M_{d'-1}) \leq \cdots \leq M_1 < F(M_1) \leq M_0 = O_{M_{d'},F}(1) \tag{4.4.33}$$

such that for all $1 \leq j < d',\ f \in \mathcal{H}_j$, and $\overline{q} \in \overline{Z}$,

[8] This condition is just for ensuring that $M_d, \ldots, M_0$ in (4.4.33) is a strictly increasing sequence of integers.

$$compl(\mathcal{B}'_{\overline{q},f}) \leq M_j. \tag{4.4.34}$$

3. *For all* $1 \leq j \leq d'$, $e \in \mathcal{H}_j$, $\overline{q} = (q,p) \in \overline{\Omega}$, *and* $G_{\overline{q},e} \in \mathcal{B}_{\overline{q},e}$ *(with* $\mathcal{B}_{\overline{q},e} := \mathcal{B}_{q,e}$*, if* $j = d'$*), one has*

$$\|\mathbb{E}_{\mu_{\overline{q},e}}(\mathbf{1}_{G_{\overline{q},e}} | \bigvee_{f \in \partial e} \mathcal{B}'_{\overline{q},f}) - \mathbb{E}_{\mu_{\overline{q},e}}(\mathbf{1}_{G_{\overline{q},e}} | \bigvee_{f \in \partial e} \mathcal{B}_{\overline{q},f})\|_{L^2(\mu_{\overline{q},e})} \leq \frac{1}{F(M_j)} \tag{4.4.35}$$

and

$$\|\mathbf{1}_{G_{\overline{q},e}} - \mathbb{E}_{\mu_{\overline{q},e}}(\mathbf{1}_{G_{\overline{q},e}} | \bigvee_{f \in \partial e} \mathcal{B}'_{\overline{q},f})\|_{\Box_{\mu_{\overline{q}.e}}} \leq \frac{1}{F(M_1)}. \tag{4.4.36}$$

Proof. We proceed by an induction on d'. If $d' = 1$ the statement follows from Preliminary Regularity Lemma 4.4.5 with $\varepsilon = \frac{1}{F(M_1)}$, so assume that $d' \geq 2$ and the theorem holds for all $j \leq d' - 1$. Apply Lemma 4.4.5 on $\mathcal{H}_{d'}$ with a very fast growing growth function $F^* \geq F$ (to be specified later [9]) and with $\varepsilon = \frac{1}{2F^*(M_{d'})}$. This gives a well-defined, symmetric extension $\{\mu_{q',f}\}_{f \in \mathcal{H}}$ and a family of σ-algebras $\mathcal{B}_{q',f} \subseteq \mathcal{B}'_{q',f}$, $f \in \mathcal{H}_{d'-1}$ defined on a parameter space $Z' = Z \times V$, such that (recall the definition of M_{d-1} in (4.4.29))

$$F(M_{d'}) \leq F^*(M_{d'}) \leq M_{d'-1} \leq O_{K,M_{d'},F^*}(1) \tag{4.4.37}$$

$$\|\mathbb{E}_{\mu_{q',e}}(\mathbf{1}_{G_{q',e}} | \bigvee_{f \in \partial e} \mathcal{B}'_{q',f}) - \mathbb{E}_{\mu_{q',e}}(\mathbf{1}_{G_{q',e}} | \bigvee_{f \in \partial e} \mathcal{B}_{q',f})\|_{L^2(\mu_{q',e})} \leq \frac{1}{2F^*(M_{d'})} \tag{4.4.38}$$

and

$$\|\mathbf{1}_{G_{q',e}} - \mathbb{E}_{\mu_{q',e}}(\mathbf{1}_{G_{q',e}} | \bigvee_{f \in \partial e} \mathcal{B}'_{q',f})\|_{\Box_{\mu_{q',e}}} \leq \frac{1}{F^*(M_{d'-1})}, \tag{4.4.39}$$

hold for all $q' = (q,p) \in \Omega'$, $e \in \mathcal{H}_{d'}$, and $G_{q',e} \in \mathcal{B}_{q',e} = \mathcal{B}_{q,e}$, where $\Omega' \subseteq \Omega \times V \subseteq Z'$ is a set of measure $\psi'(\Omega') \geq c(c_0, F, M_{d'}) > 0$. With this system, apply the induction hypothesis to the system $\{\mu_{q',f}\}_{q' \in Z', f \in \mathcal{H}}$, $\{\mathcal{B}_{q',f}\}_{q' \in Z', f \in \mathcal{H}_{d'-1}}$,$M_{d'-1}$ the growth function F, and the set Ω', one obtains an extension $\{\mu_{\overline{q},f}\}_{\overline{q} \in \overline{Z}, f \in \mathcal{H}}$ and families of $\sigma-$algebras $\{\mathcal{B}_{\overline{q},f} \subseteq \mathcal{B}'_{\overline{q},f}\}_{\overline{q} \in \overline{Z},\, f \in \mathcal{H}_j}$ such that (4.4.34) - (4.4.36) hold for $j < d' - 1$, with constants

$$M_{d'-1} < F(M_{d'-1}) \leq \cdots \leq M_1 < F(M_1) = O_{M_{d'-1},F}(1). \tag{4.4.40}$$

[9] F^* will be chosen depending on F and grows much faster than F so that $F(M_{d'}) < F^*(M_{d'}) < M_0 = \frac{1}{2}F^*(M_{d'-1})$ so F^* controls the size of M_0 in terms of M_d, F. F^* is also used in inductive argument to conclude that $F(M_d) < M_{d-1}$. We apply preliminary regularity lemma to F^*.

For $\overline{q}=(q',p)\in\overline{Z}$, $f\in\mathcal{H}_{d'-1}$ set

$$\mathcal{B}_{\overline{q},f}:=\mathcal{B}_{q',f},\quad \text{and}\quad \mathcal{B}'_{\overline{q},f}:=\mathcal{B}'_{q',f}. \tag{4.4.41}$$

We show that inequalities (4.4.35) and (4.4.36) hold for $j=d'$. Indeed, by the stability property (4.3.16), one has

$$\begin{aligned}
&\big\|\mathbb{E}_{\mu_{\overline{q},e}}(\mathbf{1}_{G_{q',e}}|\bigvee_{f\in\partial e}\mathcal{B}'_{q',f})-\mathbb{E}_{\mu_{\overline{q},e}}(\mathbf{1}_{G_{q',e}}|\bigvee_{f\in\partial e}\mathcal{B}_{q',f})\big\|_{L^2(\mu_{\overline{q},e})}\\
&=\big\|\mathbb{E}_{\mu_{q',e}}(\mathbf{1}_{G_{q',e}}|\bigvee_{f\in\partial e}\mathcal{B}'_{q',f})-\mathbb{E}_{\mu_{q',e}}(\mathbf{1}_{G_{q',e}}|\bigvee_{f\in\partial e}\mathcal{B}_{q',f})\big\|_{L^2(\mu_{q',e})}+o_{K,M_{d'},F,F^*}(1)\\
&\le\frac{1}{2F^*(M_{d'})}+o_{K,M_{d'},F,F^*}(1),
\end{aligned}\tag{4.4.42}$$

for all $\overline{q}=(q',p)\in\overline{\Omega}\backslash\overline{\mathcal{E}}_1$, $e\in\mathcal{H}_{d'}$, and $G_{q',e}\in\mathcal{B}_{q',e}$. Here $\overline{\mathcal{E}}_1\subseteq\overline{\Omega}$ is a set of measure $\overline{\psi}(\overline{\mathcal{E}}_1)=o_{K,M_{d'},F,F^*}(1)$.

Similarly using the stability properties (4.3.12) and (4.3.21) of the box norms (and also (4.4.41)), we have

$$\begin{aligned}
&\big\|\mathbf{1}_{G_{q',e}}-\mathbb{E}_{\mu_{\overline{q},e}}(\mathbf{1}_{G_{q',e}}|\bigvee_{f\in\partial e}\mathcal{B}'_{q',f})\big\|_{\Box_{\mu_{\overline{q},e}}}=\big\|\mathbf{1}_{G_{q',e}}-\mathbb{E}_{\mu_{q',e}}(\mathbf{1}_{G_{q',e}}|\bigvee_{f\in\partial e}\mathcal{B}'_{q',f})\big\|_{\Box_{\mu_{\overline{q},e}}}+o_{K,M_{d'},F,F^*}(1)\\
&=\big\|\mathbf{1}_{G_{q',e}}-\mathbb{E}_{\mu_{q',e}}(\mathbf{1}_{G_{q,e}}|\bigvee_{f\in\partial e}\overline{\mathcal{B}}_{q',f})\big\|_{\Box_{\mu_{q',e}}}+o_{K,M_{d'},F,F^*}(1)\le\frac{1}{2F^*(M_{d'-1})}+o_{K,M_{d'},F,F^*}(1),
\end{aligned}\tag{4.4.43}$$

for all $\overline{q}=(q',p)\in\overline{\Omega}\backslash\overline{\mathcal{E}}_2$, $e\in\mathcal{H}_{d'}$ and $A_{q',e}\in\mathcal{B}_{q',e}=\mathcal{B}_{\overline{q},e}$, where $\overline{\mathcal{E}}_2\subseteq\overline{\Omega}$ is a set of measure $\overline{\psi}(\overline{\mathcal{E}}_2)=o_{K,M_{d'},F,F^*}(1)$.

With $F(M_1)=O_{M_{d'-1},F}(1)$. Now we link $M_{d'}$ with (4.4.40). Choose (modify) the function $F^*=F^*_{M_{d'},M_{d'-1},F}$ so that it grows fast enough that

- $F^*(M_{d'})<M_{d'-1}$.
- $F(M_1)<c(M_{d'},F)<\frac{F^*(M_{d'-1})}{2}$.

Then we have from (4.4.37) that

$$M_{d'}<F(M_{d'})\le F^*(M_{d'})\le M_{d'-1}<F(M_{d'-1})\le\cdots\le M_1<F(M_1)\le M_0=O_{M_{d'},F}(1):=\frac{1}{2}F^*(M_{d'-1}) \tag{4.4.44}$$

Assuming $N,\ W$ are sufficiently large with respect to $M_{d'}$ and K, inequalities (4.4.35), (4.4.36) for $j = d'$ and $\overline{q} \in \overline{\Omega} \backslash (\overline{\mathcal{E}}_1 \cup \overline{\mathcal{E}}_2)$ follow from (4.4.38) and (4.4.39) and (4.4.44). The rest of the conclusions of the theorem (4.4.32), (4.4.33), (4.4.34) are clear from the construction. $\square$

4.5 Counting Lemma

In this section we formulate a so-called counting lemma and show how it implies Theorem 4.1.7. Our arguments will closely follow and are straightforward adaptations of those in [106] to the weighted settings.

For $e \in \mathcal{H}_d$ let $G_e \subseteq V_e$ be a hypergraph, and let $\mathcal{B}_e = \{G_e, G_e^C, \emptyset, V_e\}$ be the $\sigma-$algebra generated by G_e. Let $\{\nu_e\}_{e\in\mathcal{H}}$ and $\{\mu_e\}_{e\in\mathcal{H}}$ be the weights and measures associated to a well-defined, symmetric family forms $\mathcal{L} = \{L_e^k;\ e \in \mathcal{H}_d,\ 1 \leq k \leq d\}$. Take $M_d > 0$, $F : \mathbb{R}_+ \to \mathbb{R}_+$ be a growth function (F to be determined later) and apply Theorem 4.4.6 with $d' = d$ to obtain a well-defined, symmetric parametric extension $\{\mu_{q,f}\}_{q\in Z, f\in\mathcal{H}}$ together with σ-algebras $\mathcal{B}_{q,e}, \mathcal{B}'_{q,e}, \mathcal{B}_{q,e} \subseteq \mathcal{B}'_{q,e}$ and a set $\Omega \subseteq Z$ such that (4.4.32)-(4.4.36) hold.[10] Note that the complexity of the system as well as the σ-algebras is $O_{M_d,F}(1)$. We consider the system of measures $\mu_{q,f}$ and $\mathcal{B}_{q,f},\ \mathcal{B}'_{q,f}, f \in \mathcal{H}$ fixed for the rest of this section.

It will be convenient to define all our σ-algebras on the same space V_J and eventually replace the ensemble of measures $\{\mu_{q,e}\}_{e\in\mathcal{H}}$ with the measure $\mu_q := \mu_{q,J} = \prod_{f\in\mathcal{H}} \nu_{q,f}$. Thanks to the stability conditions (4.3.3)-(4.3.4) this can be done at essentially no cost: Indeed for any $e \in \mathcal{H}$ there is an exceptional set $\mathcal{E}_e \subseteq \Omega$ of measure $\psi(\mathcal{E}_e) = o_{M_d,F}(1)$, such that for any family of sets $G_{q,e} \subseteq V_e$ we have that

$$\mu_q(\pi_e^{-1}(G_{q,e})) = \mu_{q,e}(G_{q,e}) + o_{M_d,F}(1), \tag{4.5.1}$$

uniformly for $q \in \Omega \backslash \mathcal{E}_e$. Let $\mathcal{E} = \bigcup_{e\in\mathcal{H}} \mathcal{E}_e$, $\Omega' := \Omega \backslash \mathcal{E}$, then (4.5.1) means that for any set $A_{q,e} \in \mathcal{A}_e$ one has that $\mu_q(A_{q,e}) = \mu_{q,e}(\pi_e(A_{q,e})) + o_{M_d,F}(1)$ uniformly for $q \in \Omega'$. We will write

$$\mu_{q,e}(\overline{A}_{q,e}) = \mu_{q,e}(\pi_e(\overline{A}_{q,e}))$$

for simplicity of notations.

Define the σ-algebras $\overline{\mathcal{B}}_{q,e} := \pi_e^{-1}(\mathcal{B}_{q,e})$, $\overline{\mathcal{B}}'_{q,e} := \pi_e^{-1}(\mathcal{B}'_{q,e})$ on V_J, and note that $\overline{\mathcal{B}}_{q,e} = \overline{\mathcal{B}}_e$ for $e \in \mathcal{H}_d$ as the initial σ-algebras $\mathcal{B}_e$ are not altered in Theorem 4.4.6. Let $\overline{\mathcal{B}}_q := \bigvee_{e\in\mathcal{H}} \overline{\mathcal{B}}_{q,e}$ be the σ-algebra generated by the algebras $\overline{\mathcal{B}}_{q,e}$, and define similarly the σ-algebra $\overline{\mathcal{B}}'_q$. The atoms of $\overline{\mathcal{B}}_q$ are of the form $A_q = \bigcap_{e\in\mathcal{H}} A_{q,e}$ where $A_{q,e}$ is an atom of $\overline{\mathcal{B}}_{q,e}$. In particular if $E_e \in \overline{\mathcal{B}}_e$ then $\bigcap_{e\in\mathcal{H}_d} E_e$

[10] The family $\{\nu_e\}$ can be considered as a parametric family of weights in a trivial way, setting $Z = \Omega = \{0\}$, and $\psi(0) = 1$.

is the union[11] of the atoms of $\overline{\mathcal{B}}_q$.

Basically, the counting lemma says that as we decompose

$$\mathbf{1}_{A_{q,e}} = \mathbb{E}_{\mu_{q,e}}(\mathbf{1}_{A_{q,e}}| \bigvee_{f\in\partial e} \overline{\mathcal{B}}_{q,f}) + b_{q,e} + c_{q,e}$$

where $b_{q,e}$ is small in L^2 norm and $c_{q,e}$ is small is box norm. Our counting lemma says that for *most atoms*, when we calculate the measure $\mu_q(A_q) = \mu_q(\cap_{f\in\mathcal{H}} A_{q,f})$ we have

$$\mu_q(\cap_{f\in\mathcal{H}} A_{q,f}) = \prod_{e\in\mathcal{H}} \mathbb{E}_{\mu_{q,e}}(\mathbf{1}_{A_{q,e}}| \bigvee_{f\in\partial e} \overline{\mathcal{B}}_{q,f}) + \text{small error} = \prod_{e\in\mathcal{H}} \frac{\mu_{q,e}(A_{q,e}\cap\cap_{f\in\partial e} A_{q,f})}{\mu_{q,e}(\cap_{f\in\partial e} A_{q,f})} + \text{small error}$$

That is most atoms can be approximated by its *relative density with respect to one lower order atoms in* $\bigvee_{f\in\partial e\overline{\mathcal{B}}_{q,f}}$ which comes from the main term of the decomposition. The terms $b_{q,e}, c_{q,e}$ only contribute to small error terms. To get rid of $c_{q,e}$, we only need the usual *Generalized von Neumann Inequality* argument. A bit more work will be needed to get rid of $b_{q,e}$.

A consequence of the counting lemma is that one can show the measures of an atom that we are interested is bounded below by a positive constant depending only on the initial data F and M_d. If, as in Theorem 4.1.7, one assumes that the measure of $\bigcap_{e\in\mathcal{H}_d} E_e$ is sufficiently small then it cannot contain most of the atoms (will be named *regular atoms*) thus removing the exceptional atoms from the sets E_e, the intersection of the remaining sets becomes empty, leading to a proof of Theorem 4.1.7.

To make this heuristic precise let us start by defining the *relative density* $\delta_{q,e}(A|B) := \mu_{q,e}(A\cap B)/\mu_{q,e}(B)$ for $A, B \in \overline{\mathcal{B}}_{q,e}$, with the convention that $\delta_{q,e}(A|B) := 1$ if $\mu_{q,e}(B) = 0$.

Definition 4.5.1. *Let $A_q = \cap_{e\in\mathcal{H}} A_{q,e}$ be an atom of $\mathcal{B}_q$.For $e \in \mathcal{H}_j, 1 \le j \le d$. We say that the atom A_q is* regular *if the following hold.*

1. *For all atoms $A_{q,e}$ the relative density is not too small*[12]*:*

$$\delta_{q,e}(A_{q,e}| \bigcap_{f\in\partial e} A'_{q,f}) \ge \frac{1}{\log F(M_j)}, \tag{4.5.2}$$

2. *It satisfies an regularity condition*[13]*:*

$$\int_{V_e} |\mathbb{E}_{\mu_q}(\mathbf{1}_{A_{q,e}}| \bigvee_{f\in\partial e} \overline{\mathcal{B}}'_{q,f}) - \mathbb{E}_{\mu_q}(\mathbf{1}_{A_{q,e}}| \bigvee_{f\in\partial e} \overline{\mathcal{B}}_{q,f})|^2 \prod_{f\subsetneq e} \mathbf{1}_{A'_{q,f}}\, d\mu_{q,e} \le \frac{1}{F(M_j)} \int_{V_e} \prod_{f\subsetneq e} \mathbf{1}_{A'_{q,f}}\, d\mu_{q,e}. \tag{4.5.3}$$

[11] Indeed, $\bigvee_{e\in\mathcal{H}_d} \overline{\mathcal{B}}_{q,e} \subseteq \bigvee_{f\in\mathcal{H}} \overline{\mathcal{B}}_{q,f}$

[12] Don't take the log function too seriously.

[13] As mentioned in chapter 1, this is related to Box norm. The notation of regular atoms has some relations to (hyper-)graph regularity

This roughly means that all atoms $A_{q,e}$ are both somewhat large and regular on the intersection of the lower order atoms $A'_{q,f}$, $(f \in \partial e)$. Note that if $|e| = 1$ then $\partial e = \emptyset$ and by convention we define $\bigcap_{f\in\partial e} A'_{q,f} = V_J$, and the left side of (4.5.2) becomes $\mu_{q,e}(A_{q,e})$.

Now we state the counting lemma

Proposition 4.5.2. *[Counting lemma] There is a set $\mathcal{E} \subseteq \Omega$ of measure $\psi(\mathcal{E}) = o_{N,W\to\infty;M_d,F}(1)$ such that if $q \in \Omega \backslash \mathcal{E}$ and if $A_q = \bigcap_{e\in\mathcal{H}} A_{q,e} \in \bigvee_{e\in\mathcal{H}} \overline{\mathcal{B}}_{q,e}$ is a regular atom, then*

$$\mu_q(A_q) = (1 + o_{M_d\to\infty}(1)) \prod_{e\in\mathcal{H}} \delta_{q,e}(A_{q,e}| \bigcap_{f\in\partial e} A_{q,f}) + O_{M_1}\left(\frac{1}{F(M_1)}\right) + o_{N,W\to\infty;M_d,F}(1). \tag{4.5.4}$$

An important corollary of the counting lemma is that each of the regular atoms is not too small in measure and the total measures of all irregular atoms is small, if we assume F is sufficiently fast growing, of exponential type.

Lemma 4.5.3 (Regular atoms)**.** *For $F(M) \geq 2^{2^{2^{M+3}}}$ and sufficiently large M_d,*

1. *(Total measure of Irregular atoms is small) For each $A_{q,e} \in \overline{\mathcal{B}}_{q,e}$, define the set*

   $B_{q,e,A_{q,e}}$ be the union of all sets of the form $\bigcap_{f\subsetneq e} A'_{q,f}$ for which (4.5.2) or (4.5.3) fails.

   Note that if an atom $A_q = \bigcap_{e\in\mathcal{H}} A_{q,e}$ is irregular then $A_q \subseteq A_{q,e} \cap B_{q,e,A_{q,e}}$ for some $e \in \mathcal{H}$. Then for $q \notin \mathcal{E}_1$, where $\mathcal{E}_1 \subseteq \Omega$ is a set of measure $\psi(\mathcal{E}_1) = o_{M_d,F}(1)$. We have

$$\mu_q(A_{q,e} \cap B_{q,e,A_{q,e}}) \lesssim \frac{1}{\log F(M_j)} \tag{4.5.5}$$

2. *(A regular atom is large) For $q \in \Omega$ and a regular atom $A_q = \cap_{f\in\mathcal{H}} A_{q,f}$,*

$$\mu_q(A_q) \geq \frac{1}{F(M_1)} > 0, \tag{4.5.6}$$

Proof. First we show (4.5.5). Note that the measure μ_q can be replaced by the measure $\mu_{q,e}$ as they differ by a negligible quantity on sets which belong to $\mathcal{A}_e$. We estimate first the contribution of those sets $\bigcap_{f\subsetneq e} A_{q,f}$ to the left side of (4.5.5) for which (4.5.2) fails. This quantity is bounded by

$$\sum_{\{A_{q,f}\}_{f\subsetneq e},\ (4.5.2)\text{ fails}} \mu_{q,e}(A_{q,e} \cap \bigcap_{f\in\partial e} A_{q,f}) \lesssim_d \frac{1}{\log F(M_j)} \sum_{\{A_{q,f}\}_{f\in\partial e}} \mu_{q,e}(\bigcap_{f\in\partial e} A_{q,f})$$

$$\leq \frac{1}{\log F(M_j)}\,\mu_{q,e}(V_e) \lesssim \frac{1}{\log F(M_j)},$$

as the summation is taken over the disjoint atoms of the σ-algebra $\bigvee_{f\in\partial e}\overline{\mathcal{B}}_{q,f}$.

Similarly, one estimates the total contribution of the disjoint atoms $\bigcap_{f\subsetneq e} A_{q,f}$ for which (4.5.3) fails as follows.

$$\begin{aligned}
&\sum_{\{A_{q,f}\}_{f\subsetneq e},\ (4.5.3)\text{ fails}} \mu_{q,e}\Big(\bigcap_{f\subsetneq e} A_{q,f}\Big) \\
&\leq F(M_j) \sum_{\{A_{q,f}\}_{f\subsetneq e},\ (4.5.3)\text{ fails}} \int_{V_e} \mathbb{E}_{\mu_{q,e}}(\mathbf{1}_{A_{q,e}}|\bigvee_{f\in\partial e}\overline{\mathcal{B}}'_{q,f}) - \mathbb{E}_{\mu_{q,e}}(\mathbf{1}_{A_{q,e}}|\bigvee_{f\in\partial e}\overline{\mathcal{B}}_{q,f})|^2 \prod_{f\subsetneq e}\mathbf{1}_{A_{q,f}}\,d\mu_{q,e} \\
&\leq F(M_j) \int_{V_e} |\mathbb{E}_{\mu_{q,e}}(\mathbf{1}_{A_{q,e}}|\bigvee_{f\in\partial e}\overline{\mathcal{B}}'_{q,f}) - \mathbb{E}_{\mu_{q,e}}(\mathbf{1}_{A_{q,e}}|\bigvee_{f\in\partial e}\overline{\mathcal{B}}_{q,f})|^2\,d\mu_{q,e} \\
&\leq F(M_j)\,\frac{1}{F(M_j)^2} = \frac{1}{F(M_j)}.
\end{aligned}$$

Since the sets $A_{q,e}\cap B_{q,e,A_{q,e}}$ contain all irregular atoms, and for given $e\in\mathcal{H}_j$ the number of all atoms of the σ-algebra $\overline{\mathcal{B}}_{q,e}$ is at most $2^{2^{M_j}}$, one estimates the total measure of all irregular atoms as

$$\sum_{j=1}^{d}\sum_{e\in\mathcal{H}_j}\sum_{A_{q,e\in\mathcal{B}_{q,e}}} \mu_q(A_{q,e}\cap\mathcal{B}_{q,e,A_{q,e}}) \leq \sum_{j=1}^{d}\binom{d}{j}2^{2^{M_j}}\frac{1}{\log F(M_j)} \leq \sum_{j=1}^{d}\frac{2^{2^{M_j}+d}}{\log F(M_j)} \leq \frac{1}{\sqrt{\log F(M_d)}} \leq 2^{-2^{M_d}} \tag{4.5.7}$$

Here the two last inequalities will follow if we choose M_d sufficiently large and F sufficiently fast growing: choose M_d so that $d2^d2^{2^{M_j}} \leq 2^{2^{M_j+1}}$ and $2^{2^{2^{M_j+3}}} \geq e^{2^{2^{M_j+2}}}$ for all j ($M_d\geq 1$ should suffice here). Now choose

$$F(M) \geq 2^{2^{2^{M+3}}} \tag{4.5.8}$$

Then

$$\sum_{j=1}^{d}\frac{2^{2^{M_j}+d}}{\log F(M_j)} \leq \sum_{j=1}^{d}\frac{2^{2^{M_j}+d}}{\sqrt{2^{2^{M_j+2}}}}\frac{1}{\sqrt{\log F(M_j)}} = \sum_{j=1}^{d}\frac{2^{2^{M_j}+d}}{2^{2^{M_j+1}}}\frac{1}{\sqrt{\log F(M_j)}} \leq \sum_{j=1}^{d}\frac{1}{d}\frac{1}{\sqrt{\log F(M_j)}} \leq \frac{1}{\sqrt{\log F(M_d)}}$$

So (4.5.7) follows.

Now we use the counting lemma to show (4.5.6). Indeed by (4.5.2), (4.5.4), we have that for $q\in\Omega$ and a regular atom $A_q=\cap_{f\in\mathcal{H}}A_{q,f}$,

$$
\begin{aligned}
\mu_q(A_q) &\geq \prod_{j\leq d}\prod_{e\in\mathcal{H}_j}\frac{1}{F(M_j)^{1/10}} - O_{d,M_1}\left(\frac{1}{F(M_1)}\right) + o_{M_d,F}(1) \\
&\geq \frac{1}{F(M_1)^{1/10}}\frac{1}{M_1^{c(d)}} - O_{d,M_1}\left(\frac{1}{F(M_1)}\right) + o_{M_d,F}(1) \geq \frac{1}{F(M_1)} \geq \frac{1}{F^*(M_{d'-1})} = c(M_d,F) > 0,
\end{aligned}
\tag{4.5.9}
$$

as long as F is sufficiently rapid growing and M_d is sufficiently large with respect to d, here we apply (4.4.44) and (4.4.29). □

4.5.1 Proof of the Counting Lemma.

We will in fact prove a stronger version of counting lemma for hypergraph bundle for which proposition 4.5.2 is a special case. The reason is that when we try to eliminate the error term $b_{q,e}$ we will apply Cauchy-Schwartz's inequality to lower order graph, causing the double vertices which could be described as lower order hypergraph bundle, allowing us to apply induction hypothesis from the statement of counting lemma for hypergraph bundle.

Definition 4.5.4 (Weighted hypergraph bundles over $\mathcal{H}$)**.** *Let K be a finite set together with a map $\pi : K \to J$, called the projection map of the bundle to the index set J. Let $\mathcal{G}_K$ be the set of edges $g \subseteq K$ such that π is injective on g and $\pi(g) \in \mathcal{H}$.*

For any $g \in \mathcal{G}_K$, write

$$V_g := V_{\pi(g)} = \prod_{k\in g} V_{\pi(k)},$$

and define the weights and measures $\overline{\nu}_{q,g}$, $\overline{\mu}_{q,g} : V_g \to \mathbb{R}_+$ as

$$\overline{\nu}_{q,g}(x_g) := \nu_{q,\pi(g)}(x_g),\ \ \overline{\mu}_{q,g}(x_g) = \prod_{g'\subseteq g}\overline{\nu}_{q,g'}(x_{g'}).$$

The total measure measure $\overline{\mu}_{q,K}$ on V_K is given by

$$\overline{\mu}_{q,K}(x) = \prod_{g\in\mathcal{G}_K}\overline{\nu}_{q,g}(x_g).$$

*A hypergraph $\mathcal{G} \subseteq \mathcal{G}_K$ which is **closed** in the sense that $\partial g \subseteq \mathcal{G}$ for every $g \in \mathcal{G}$, together with the spaces V_g and the weight functions $\overline{\nu}_{q,g}$ for $g \in \mathcal{G}$ is called a **weighted hypergraph bundle** over $\mathcal{H}$. The quantity $d' = \sup_{g\in\mathcal{G}} |g|$ is called the **order** of $\mathcal{G}$.*

Remark 4.5.5. *The underlying linear forms defining the weight system* $\{\overline{\nu}_{q,g}\}_{q\in Z, g\in\mathcal{G}_K}$,

$$\bar{L}_g(q,x_g) = L_{\pi(g)}(q,x_g), \quad supp_x\,(L_{\pi(g)}) = \pi(g)$$

are pairwise linearly independent. Indeed, if $g \neq g'$ *they depend on different sets of variables, and for a fixed sets of variables they are the same as the forms* $L(q,x_g)$. *What happens is that we sample a number variables from each space* V_j *and evaluate the forms* $L(q,x)$ *in the new variables. For example if we have* $x_1, x_{1'} \in V_1$ *and* $x_2, x_{2'} \in V_2$ *then to the edge* $(1,2) \in \mathcal{H}$ *there correspond the edges* $(1,2), (1,2'), (1',2)$ *and* $(1',2')$ *in* $\mathcal{G}$, *and to every linear form* $L(q,x_1,x_2)$ *there also correspond the forms* $L(q,x_1,x_{2'})$, $L(q,x_{1'},x_2)$ *and* $L(q,x_{1'},x_{2'})$ *defining the weights on the appropriate edges.*

Proposition 4.5.6. *[Generalized Counting Lemma] Let* $\mathcal{G} \subseteq \mathcal{G}_K$ *be a closed hypergraph bundle over* $\mathcal{H}$ *with the projection map* $\pi : K \to J$, *and* $d' := \sup_{g\in\mathcal{G}} |g|$ *be the order of* $\mathcal{G}$. *Then, for* F *growing sufficiently rapidly with respect to* d *and* K, *there exists a set* $\mathcal{E} \subseteq \Omega$ *of measure* $\psi(\mathcal{E}) = o_{N\to\infty;M_d,K,F}(1)$ *such that for* $q \in \Omega\backslash\mathcal{E}$ *we have*

$$\begin{aligned}&\int_{V_K} \prod_{g\in\mathcal{G}} \mathbf{1}_{A_{q,\pi(g)}}(x_g)\, d\overline{\mu}_{q,K}(x) \qquad (4.5.10)\\ &= (1+o_{M_d\to\infty,K}(1)) \prod_{g\in\mathcal{G}} \delta_{q,\pi(g)}(A_{q,\pi(g)}| \bigcap_{f\in\partial\pi(g)} A_{q,f}) + O_{K,M_1}(\frac{1}{F(M_1)}) + o_{N\to\infty,K,M_d}(1).\end{aligned}$$

Note that Proposition 4.5.2 is the special case when $\mathcal{G} = \mathcal{H}$ and π is the identity map.

Proof. We use a double induction. First we induct on d', the order of $\mathcal{G}$ (note that $d' \leq d$). Then, fixing K and π, we induct on the number of edges $r := |\{g \in \mathcal{G} : |g| = d'\}|$.

To start, assume that $d' = r = 1$, so that $\mathcal{G} = \{k\}$ and $j = \pi(k) \in J$. The left hand side of (4.5.10) becomes

$$\int_{V_k} \mathbf{1}_{A_{q,j}}(x_k)\, d\overline{\mu}_{q,k}(x_k) = \int_{V_j} \mathbf{1}_{A_{q,j}}(x_j)\, d\mu_{q,j}(x_j) = \mu_{q,j}(A_{q,j}) = \delta_{q,j}(A_{q,j}| \cap_{f\in\partial j} A_{q,f}).$$

Let $\{A_{q,e}\}_{e\in\mathcal{H}}$ be a regular collection of atoms for $q \in \Omega$, and define the functions $b_{q,e},\ c_{q,e} : V_e \to \mathbb{R}$ for $e \in \mathcal{H}$ by

$$b_{q,e} := \mathbb{E}_{\mu_{q,e}}(\mathbf{1}_{A_{q,e}}| \bigvee_{f\in\partial e} \mathcal{B}'_{q,f}) - \mathbb{E}_{\mu_{q,e}}(\mathbf{1}_{A_{q,e}}| \bigvee_{f\in\partial e} \mathcal{B}_{q,f}) \qquad (4.5.11)$$

$$c_{q,e} := \mathbf{1}_{A_{q,e}} - \mathbb{E}_{\mu_{q,e}}(\mathbf{1}_{A_{q,e}}| \bigvee_{f\in\partial e} \mathcal{B}'_{q,f}) \tag{4.5.12}$$

and introduce the shorthand notation

$$\delta_{q,e} = \delta_{q,e}(A_{q,e}| \bigcap_{f\in\partial e} A_{q,f}).$$

Note that if $x \in A_{q,e} \bigcap_{f\in\partial e} \mathcal{A}_{q,f}$ then

$$\delta_{q,e} = \mathbb{E}_{\mu_{q,e}}(\mathbf{1}_{A_e}| \bigvee_{f\in\partial e} \mathcal{B}_{q,f})(x_e), \tag{4.5.13}$$

and thus one has the decomposition

$$\mathbf{1}_{A_{q,e}}(x_e) = \delta_{q,e} + b_{q,e}(x_e) + c_{q,e}(x_e) \tag{4.5.14}$$

on the set $A_{q,e} \cap \bigcap_{f\in\partial e} \mathcal{A}_{q,f}$. To apply induction on r, let $g_0 \in \mathcal{G}$ such that $|g_0| = d'$ and use (4.5.14) to write

$$\prod_{g\in\mathcal{G}} \mathbf{1}_{A_{q,\pi(g)}}(x_g) = (\delta_{q,\pi(g_0)} + b_{q,\pi(g_0)}(x_{g_0}) + c_{q,\pi(g_0)}(x_{g_0})) \prod_{g\in\mathcal{G}\setminus\{g_0\}} \mathbf{1}_{A_{q,\pi(g)}}(x_g).$$

Consider the contribution of the terms separately:

Step1: Main term

$$\begin{aligned}
&\int_{V_K} \prod_{g\in\mathcal{G}} \mathbf{1}_{A_{q,\pi(g)}}(x_g) d\overline{\mu}_{q,K}(x) \\
&= \int_{V_K} (\delta_{q,\pi(g_0)} + b_{q,\pi(g_0)}(x_{g_0}) + c_{q,\pi(g_0)}(x_{g_0})) \prod_{g\in\mathcal{G}\setminus\{g_0\}} \mathbf{1}_{A_{q,\pi(g)}}(x_g) d\overline{\mu}_{q,K}(x) \\
&= M_q + E_q^1 + E_q^2
\end{aligned} \tag{4.5.15}$$

For main term M_q, by the second induction hypothesis we have

$$M_q = \delta_{q,\pi(g_0)} \int_{V_K} \prod_{g\in\mathcal{G}\setminus\{g_0\}} \mathbf{1}_{A_{q,\pi(g)}}(x_g) d\overline{\mu}_{q,K}(x)$$

$$= \delta_{q,\pi(g_0)} \left(1 + o_{M_d\to\infty}(1)\right) \prod_{g\in\mathcal{G}\setminus g_0} \delta_{q,\pi(g)} + \; O_{K,M_1}(\frac{1}{F(M_1)}) + \; o_{N,W\to\infty;K,M_d}(1),$$

and hence M_q agrees with the right side of (4.5.10).

Step2: We eliminate the error term $c_{q,e}$ by *Generalized von Neumann's Theorem argument.*

$$E_q^2 = \int_{V_K} c_{q,\pi(g_0)}(x_{g_0}) \prod_{g\in\mathcal{G}\setminus\{g_0\}} \mathbf{1}_{A_{q,\pi(g)}}(x_g) d\overline{\mu}_q(x) = \mathbb{E}_{x\in V_K}(c_{q,\pi(g_0)}\overline{\nu}_{q,g_0})(x_{g_0}) \prod_{g\in\mathcal{G}\setminus\{g_0\}} \mathbf{1}_{A_{q,\pi(g)}}\overline{\nu}_{q,g}(x_g)$$

$$= \mathbb{E}_{x\in V_K} \prod_{|g|=d',g\in\mathcal{G}} f_{q,g}(x_g) \prod_{g'\in\mathcal{G},|g'|<d'} \overline{\nu}_{q,g'}(x_{g'}), \tag{4.5.16}$$

where $f_{q,g_0} := c_{q,\pi(g_0)}\overline{\nu}_{q,g_0}$ and $f_{q,g} := h_{q,g}\overline{\nu}_{q,g}$, for $g\in\mathcal{G}, g\neq g_0$ and $|g| = d'$ for a function $h_{q,g}$ of magnitude at most 1. Thus we have $|f_{q,g}| \leq \overline{\nu}_{q,g}$ for all $g\in\mathcal{G}, |g| = d'$. Applying the Cauchy-Schwartz inequality d' times successively in the variables $x_j,\ j\in g_0$ as in the proof of generalized von Neumann's theorem (Theorem 3.2.5), to clear all functions $f_{q,g}(x_g)$, $g\neq g_0$, which does not depend on at least one of these variables, we obtain

$$|E_q^2|^{2^{d'}} \lesssim \|c_{q,\pi(g_0)}\|_{\square_{\overline{\nu}_{q,g_0}}}^{2^{d'}} + \mathbb{E}_{x_{g_0},y_{g_0}}|\mathcal{W}_q(x_{g_0},y_{g_0}) - 1| \prod_{h\subseteq g_0} \prod_{\omega\in\{0,1\}^h} \overline{\nu}_{q,h}(\omega_h(x_h,y_h)), \tag{4.5.17}$$

where $K' := K\backslash g_0$ and

$$\mathcal{W}_q(x_{g_0},y_{g_0}) = \mathbb{E}_{x\in V_{K'}} \prod_{g\in\mathcal{G},g\notin g_0} \prod_{\omega_{g\cap g_0}\in\{0,1\}^{g\cap g_0}} \overline{\nu}_{q,g}(\omega_{g\cap g_0}(x_{g,\cap g_0},y_{g\cap g_0}),x_{g\backslash g_0}). \tag{4.5.18}$$

Note that the first term on the right hand side of (4.5.17) is $O(F(M_1)^{-2^{d'}})$ by (4.4.36) and (4.5.12). To estimate the second term of (4.5.17) we apply the Cauchy-Schwartz inequality one more time in x_{g_0}, y_{g_0} variables to see that it is $o_{N,W\to\infty;M_d,K,F}(1)$ for $q\notin\mathcal{E}_1$, where $\mathcal{E}_1$ is a set of measure $o_{N,W\to\infty;M_d,K,F}(1)$ using the fact that the underlying linear forms are pairwise linearly independent in the variables $(q,x_{g_0},y_{g_0},x_{K'})$.

Step3: We estimate the error term E_q^1 defined as

$$E_q^1 = \int_{V_K} b_{q,\pi(g_0)}(x_{g_0}) \prod_{g\in\mathcal{G}\setminus\{g_0\}} \mathbf{1}_{A_{q,\pi(g)}}(x_g) d\overline{\mu}_q(x).$$

To apply induction hypothesis, take absolute values and *discarding all factors* $\mathbf{1}_{A_{q,\pi(g)}}(x_g)$ for $|g| = d'$, $g \neq g_0$ (this will be fine due to smallness of L^2-norm of $b_{q,g}$), one estimates

$$\begin{aligned}|E_q^1| \leq \int_{V_{g_0}} &\left(|b_{q,\pi(g_0)}(x_{g_0})| \prod_{g\subsetneq g_0} \mathbf{1}_{A_{q,\pi(g)}}(x_g) \right) \\ &\times \left(\mathbb{E}_{x_{K'}} \prod_{g\in\mathcal{G}',|g|<d'} \mathbf{1}_{A_{q,\pi(g)}} \overline{\nu}_{q,g}(x_g) \prod_{h\in\mathcal{G}',|h|=d'} \overline{\nu}_{q,h}(x_h) \right) d\overline{\mu}_{q,g_0}(x_{g_0}),\end{aligned}$$

where $\mathcal{G}' = \{g \in \mathcal{G};\ g \nsubseteq g_0\}$ and recall that $K' = K\setminus g_0$. Writing $A(x_{g_0})$ for the expression in the first parenthesis, and $B(x_{g_0})$ for the expression in the second parenthesis. Thus we have

$$|E_q^1| \leq \int_{V_{g_0}} A(x_{g_0})B(x_{g_0})\, d\overline{\mu}_{q,g_0}(x_{g_0}),$$

thus by the Cauchy-Schwartz inequality we get

$$|E_q^1|^2 \lesssim \left(\int_{V_{g_0}} A(x_{g_0})^2\, d\overline{\mu}_{q,g_0}(x_{g_0}) \right) \left(\int_{V_{g_0}} B(x_{g_0})^2\, d\overline{\mu}_{q,g_0}(x_{g_0}) \right). \tag{4.5.19}$$

Since $\overline{\nu}_{q,g_0}(V_{g_0}) = 1 + o_{M_d,K,F}(1)$ outside a set $\mathcal{E}_2 \subseteq \Omega$ of measure $\psi(\mathcal{E}_2) = o_{M_d,K,F}(1)$, the first factor on the left side of (4.5.19) is estimated by

$$\mathbb{E}_{x_{g_0}\in V_{g_0}} b_{q,\pi(g_0)}(x_{g_0})^2 \prod_{g\subsetneq g_0} \mathbf{1}_{A_{q,\pi(g)}}(x_g) \prod_{g\subseteq g_0} \nu_{q,\pi(g)}(x_g). \tag{4.5.20}$$

Let $f_0 = \pi(g_0)$, since $\pi : g_0 \to f_0$ is injective and $V_{g_0} = V_{f_0}$, we may write the expression in (4.5.20), by re-indexing the variables x_g to x_f, $f = \pi(g)$ for $g \subseteq g_0$, as

$$\int_{V_{f_0}} b_{q,f_0}(x_{f_0})^2 \prod_{f\subsetneq f_0} \mathbf{1}_{A_{q,f}}(x_f)\, d\mu_{q,f_0}(x_{f_0}) \lesssim \frac{1}{F(M_{d'})} \int_{V_{f_0}} \prod_{f\subsetneq f_0} \mathbf{1}_{A_{q,f}}(x_f) d\mu_{q,f_0}(x_{f_0}), \tag{4.5.21}$$

where the inequality follows from by assumption (4.5.3) on regular atoms. By the induction hypothesis we further estimate the right side (4.5.21) as

$$\frac{1}{F(M_{d'})}\,(1+o_{M_d\to\infty}(1))\prod_{f\subsetneq f_0}\delta_{q,f}\;+O_{M_d}(\frac{1}{F(M_1)})+o_{N,W\to\infty;M_d,K,F}(1). \tag{4.5.22}$$

The second factor in (4.5.19) may be expressed in terms of a hypergraph bundle $\tilde{K}$ over K, by using the construction given in [106]. Let $\tilde{K}=K_0\oplus_{g_0}K$, the set $K\times\{0,1\}$ with the elements $(k,0)$ and $(k,1)$ are identified for $k\subseteq g_0$. Let $\phi:\tilde{K}\to K$ be the natural projection, and $\pi\circ\phi:\tilde{K}\to J$ be the associated map down to J. Recall $\mathcal{G}\subseteq\mathcal{G}_K$ is a closed subhypergraph.

Let $\mathcal{G}_0=\{g\in\mathcal{G},g\subseteq g_0\}$ and $\mathcal{G}'=\{g\in\mathcal{G},g\nsubseteq g_0,|g|<d'\}$ and define the hypergraph bundle $\tilde{\mathcal{G}}$ on $\tilde{K}$ to consist of the edges $g\times\{0\}$ and $g\times\{1\}$ for $g\in\mathcal{G}_0\cup\mathcal{G}'$, two edges bing identified for $g\in\mathcal{G}_0$. Define the following weights on $V_{\tilde{K}}$

$$\tilde{\nu}_{q,g\times\{i\}}(x_{g\times\{i\}}):=\overline{\nu}_{q,g}(x_{g\times\{i\}}), \tag{4.5.23}$$

for $q\in Z,\ g\in\mathcal{G}_K, i=0,1,$ (i.e. for all edges $\tilde{g}\in\mathcal{G}_{\tilde{K}}$), and let $\tilde{\mu}_{q,g\times\{i\}}$ be the associated family of measures. Then we have for the second factor appearing in (4.5.19)

$$\begin{aligned}
&\int_{V_{g_0}} B(x_{g_0})^2 d\overline{\mu}_{q,g_0}(x_{g_0})\\
&=\int_{V_{g_0}}\Big[\prod_{g\in g_0}\mathbf{1}_{A_{q,\pi(g)}}(x_g)\Big]\Big[\mathbb{E}_{x\in V_{K\backslash g_0}}\prod_{g\in\mathcal{G}\backslash\{g_0\}}\mathbf{1}_{A_{q,\pi(g)}}\overline{\nu}_{q,g}(x_g)\prod_{h\nsubseteq g_0,|h|=d'}\overline{\nu}_{q,h}(x_h)\Big]^2 d\overline{\mu}_{q,g_0}(x_{g_0})\\
&=\int_{V_{\tilde{K}}}\prod_{\tilde{g}\in\tilde{\mathcal{G}}}\mathbf{1}_{A_{q,\pi\circ\phi(\tilde{g})}}(x_{\tilde{g}})\,d\tilde{\mu}_{q,\tilde{K}}(x_{\tilde{K}}).
\end{aligned} \tag{4.5.24}$$

Indeed, when expanding the square of inner sum in (4.5.24) we double all points in $K\backslash g_0$ thus we eventually sum over $x_{\tilde{K}}\in V_{\tilde{K}}$, also double all edges $g\in\tilde{\mathcal{G}}$ to obtain the edges $g\times\{0\},g\times\{1\}$. As for the weights, the procedure doubles all weights $\overline{\nu}_{q,g}(x_g)$ for $g\not\subset g_0$, $g\in\mathcal{G}_K$ to obtain the weights $\overline{\nu}_{q,g}(x_{g\times\{i\}})$ for $i=0,1$ while leaves the weights $\overline{\nu}_{q,g}(x_g)$ for $g\subseteq g_0$ unchanged. The order of $\tilde{g}$ is less than d' thus by the first induction hypothesis, we have

$$\begin{aligned}
&\int_{V_{\tilde{K}}}\prod_{\tilde{g}\in\tilde{\mathcal{G}}}\mathbf{1}_{A_{q,\pi\circ\phi(\tilde{g})}}(x_{\tilde{g}})\,d\tilde{\mu}_{q,\tilde{K}}(x_{\tilde{K}})=\\
&=(1+o_{M_d\to\infty}(1))\prod_{\tilde{g}\in\tilde{\mathcal{G}}}\delta_{q,\pi\circ\phi(\tilde{g})}+O_{K,M_1}(\frac{1}{F(M_1)})+o_{N,W\to\infty;,M_d,K,F}(1)\\
&=(1+o_{M_d\to\infty}(1))\prod_{g\in\mathcal{G}_0}\delta_{q,\pi(g)}\prod_{g\in\mathcal{G}'}\delta^2_{q,\pi(g)}+O_{K,M_1}(\frac{1}{F(M_1)})+o_{N,W\to\infty;M_d,K,F}(1),
\end{aligned} \tag{4.5.25}$$

for $q \notin \mathcal{E}_{\tilde{K},\phi}$ where $\mathcal{E}_{\tilde{K},\phi} \subseteq \Omega$ is a set of measure $\psi(\mathcal{E}_{\tilde{K},\phi}) = o_{N,W\to\infty;M_d,K,F}(1)$. Note that there are only $O_K(1)$ choices for choosing the set $\tilde{K}$ and the projection map $\phi : \tilde{K} \to K$ thus taking the union of all possible exceptional sets $\mathcal{E}_{\tilde{K},\phi}$ we have that (4.5.25) holds for $q \notin \mathcal{E}'_K$ if measure $\psi(\mathcal{E}'_K) = o_{N,W\to\infty;M_d,K,F}(1)$. Combining the bounds (4.5.22) and (4.5.25) we obtain the error estimate

$$|E_q^1|^2 = (o_{F,M_d\to\infty}(1)) \prod_{g\in\mathcal{G}} \delta^2_{q,\pi(g)} + O_{K,M_1}(\frac{1}{F(M_1)}) + o_{N,W\to\infty;M_d,K,F}(1),$$

outside a set $\mathcal{E}'_K$ of measure $o_{N,W\to\infty;M_d,K,F}(1)$. This closes the induction and the Proposition follows. $\square$

4.6 Proof of Weighted Simplices Removal Lemma

Proof of Theorem 4.1.7. Let $\delta > 0$, $E_e \in \mathcal{A}_e$ and $g_e : V_e \to [0,1]$ for $e \in \mathcal{H}_d$ be given. Let $\mathcal{E}_1 \subseteq \Omega$ be a set of measure $\psi(\mathcal{E}_1) = o_{M_d,F}(1)$ so that (4.5.1), (4.5.7) and (4.5.9) hold for $q \in \Omega/\mathcal{E}_1$. Also by (4.3.8) conditions (4.1.4)-(4.1.5) hold for

$$\tilde{\mu}_J := \mu_{q,J} \;\; \textit{and} \;\; \tilde{\mu}_e := \mu_{q,e} \;\; (e \in \mathcal{H}_d), \tag{4.6.1}$$

for $q \notin \mathcal{E}_2$, for a set $\mathcal{E}_2 \subseteq \Omega$ be a set of measure $\psi(\mathcal{E}_2) = o_{M_d,F}(1)$.

Now fix $q \notin \mathcal{E}_1 \cup \mathcal{E}_2$ and define $\tilde{\mu}_J$ and $\tilde{\mu}_e$ for $e \in \mathcal{H}_d$ as is (4.6.1). We claim that this system of measures satisfy the conclusions of the theorem. By construction the system is symmetric so it remains to construct the sets E'_e and show (4.1.6)-(4.1.8) hold. For given $e \in \mathcal{H}_d$ define the sets

$$E'_{q,e} = V_J \backslash \, (B_{q,e,E_e} \cup \bigcup_{f\subsetneq e,\, A_{q,f}} (A_{q,f} \cap B_{q,f,A_{q,f}})), \tag{4.6.2}$$

where $A_{q,f}$ ranges over the atoms of $\mathcal{B}_{q,f}$. Hence $E'_{q,e}$ shpuld not contain a bad atom inside E_e (excluding E_e itself). As we have $\mathcal{B}_{q,e} = \mathcal{B}_e$, which is generated by a single set E_e, if $\bigcap_{e\in\mathcal{H}_d} E_e$ contains an atom $A_q = \bigcap_{f\in\mathcal{H}} A_{q,f}$ then $A_{q,e} = E_e$ for $e \in \mathcal{H}_d$. If such an atom A_q would be regular then by (4.1.3), (4.5.9), its measure would satisfy

$$\frac{1}{F^*(M_{d-1})} \leq \tilde{\mu}_J\,(\bigcap_{e\in\mathcal{H}_d} E_e\,) = \mu_J\,(\bigcap_{e\in\mathcal{H}_d} E_e\,) + o_{M_d,F}(1) \; < \; 2\delta.$$

Choosing M_d to be the largest positive integer so that

$$F^*(M_{d-1}) \leq (2\delta)^{-1} \tag{4.6.3}$$

then we see that $\bigcap_{e\in\mathcal{H}_d} E_e$ could contain only irregular atoms.
Also, from (4.6.2) and (4.5.7) we have

$$\tilde{\mu}_J\,(E_e \backslash E'_{q,e}) \;=\; \tilde{\mu}_J\,(\bigcup_{f\subseteq e,\, A_{q,f}} (A_{q,f} \cap B_{q,f,A_{q,f}})) \;\leq\; 2^{-2^{M_d}}. \tag{4.6.4}$$

Also, all irregular atoms $A_q = \bigcap_{f\in\mathcal{H}} A_{q,f} \subseteq \bigcap_{e\in\mathcal{H}_d} E_e$ are contained in one of the sets $E_e \backslash E'_{q,e}$, thus

$$\bigcap_{e\in\mathcal{H}_d} E_e \subseteq \bigcup_{e\in\mathcal{H}_d} (E_e \backslash E'_{q,e})$$

so

$$\bigcap_{e\in\mathcal{H}_d} (E_e \cap E'_{q,e}) \;=\; \emptyset.$$

Finally, choosing $\varepsilon := 2^{-2^{M_d}}$, (4.1.7) holds by (4.6.4). Moreover $\delta \to 0$ implies $M_d \to \infty$ and hence $\varepsilon \to 0$ showing the validity of (4.1.8). This proves Theorem 4.1.7. $\square$

4.7 Proof of the Main Theorem

Proof (Theorem 4.1.7 implies Theorem 4.1.3). By assumption (4.1.2) in Theorem 4.1.3 and by (4.2.14),

$$\mathbb{E}_{x\in V_J} \prod_{e\in\mathcal{H}_d} \mathbf{1}_{E_e}(x)\,\mu_J(x) \leq \delta.$$

For a given $e' \in \mathcal{H}_d$ define the function $g_{e'} : V_{e'} \to [0,1]$ as follows. Let $\phi_{e'} : V_{e'} \to M$ be the inverse of the projection map $\pi_{e'} : V_J \to V_{e'}$ restricted to M, and $y \in V_{e'}$ let

$$g_{e'}(y) := \prod_{e\in\mathcal{H}_d} \mathbf{1}_{E_e}(\phi_{e'}(y)).$$

Applying Theorem 1.4 to the system of weights $\{\nu_e\}$ and functions $\{g_e\}$ gives a system of measures $\tilde{\mu}_e$ and sets $E'_e \in \mathcal{A}_e$ satisfying (4.1.4)-(4.1.8). By (4.2.11) we have that $x \in M \cap \bigcap_{e\in\mathcal{H}_d} E_e$ if and only if $\Phi(x) = (y, 0)$ with $y \in A$. Moreover in that case $w(y) = \mu_e(x)$ for all $e \in \mathcal{H}_d$ by (4.2.13), thus for any given $e' \in \mathcal{H}_d$,

$$\mathbb{E}_{y\in\mathbb{Z}_N^d} \mathbf{1}_A(y) w(y) = \mathbb{E}_{x\in M} \prod_{e\in\mathcal{H}_d} \mathbf{1}_{E_e}(x)\mu_{e'}(x) = \mathbb{E}_{z\in V_{e'}} g_{e'}(z)\mu_{e'}(z)$$

$$= \mathbb{E}_{z \in V_{e'}} g_{e'}(z) \tilde{\mu}_{e'}(z) + o_{N,W\to\infty}(1)$$
$$= \mathbb{E}_{x \in M} \prod_{e \in \mathcal{H}_d} \mathbf{1}_{E_e}(x) \tilde{\mu}_{e'}(x) + o_{N,W\to\infty}(1).$$

By (4.1.6), $\prod_{e\in\mathcal{H}_d} \mathbf{1}_{E_e} \leq \sum_{e\in\mathcal{H}_d} \mathbf{1}_{E_e \setminus E'_e}$. Then the symmetry of the measures $\tilde{\mu}_e$ (i.e. the fact that $\tilde{\mu}_e(x) = \tilde{\mu}_{e'}(x)$ for $x \in M$), (4.1.7) and the fact that $\mathbf{1}_{E_e \setminus E'_e}$ is constant on the fibers $\pi_e^{-1}(x)$ implies

$$\mathbb{E}_{x\in M} \prod_{e\in\mathcal{H}_d} \mathbf{1}_{E_e}(x)\tilde{\mu}_{e'}(x) \leq \sum_{e\in\mathcal{H}_d} \mathbb{E}_{x\in M} \mathbf{1}_{E_e \setminus E'_e}(x)\tilde{\mu}_{e'}(x).$$

Changing the sum over M to sum over V_e, we obtain

$$\sum_{e\in\mathcal{H}_d} \mathbb{E}_{x\in M} \mathbf{1}_{E_e \setminus E'_e}(x)\tilde{\mu}_{e'}(x) = \sum_{e\in\mathcal{H}_d} \mathbb{E}_{x\in V_e} \mathbf{1}_{E_e \setminus E'_e}(x)\tilde{\mu}_{e}(x) \leq (d+1)\,\epsilon(\delta) + o_{N,W\to\infty}(1).$$

Choosing N, W sufficiently large with respect to δ gives

$$\mathbb{E}_{y\in\mathbb{Z}_N^d} \mathbf{1}_A(y) w(y) \leq \epsilon'(\delta),$$

with, say $\epsilon'(\delta) := (d+2)\epsilon(\delta)$. $\square$

First, let us identify $[1,N]^d$ with $\mathbb{Z}_N^d$ and recall that constellations in $\mathbb{Z}_N^d$ defined by the simplex Δ which are contained in a box $B \subseteq [1,N]^d$ of size εN, are in fact genuine constellations contained in B. Note that we can assume that the simplex Δ is *primitive* in the sense that $t\Delta \nsubseteq \mathbb{Z}^d$ for any $0 < t < 1$, as any simplex is a dilate of a primitive one. To any simplex $\Delta \subseteq \mathbb{Z}^d$ there exists a constant $\tau(\Delta) > 0$ depending only on Δ such that the following holds.

Lemma 4.7.1 ($\mathbb{Z}_N$ to $\mathbb{Z}$)**.** *Let $\Delta \subseteq \mathbb{Z}^d$ be a primitive simplex. Then there is constant $0 < \varepsilon < \tau(\Delta)$ so that the following holds.*
Let N be sufficiently large, and let $B = I^d$ be a box of size εN contained in $[1,N]^d \simeq \mathbb{Z}_N^d$. If there exist $x \in \mathbb{Z}_N^d$ and $1 \leq t < N$ such that $x \in B$ and $x + t\Delta \subseteq B$ as a subset on $\mathbb{Z}_N^d$, then either $x + t\Delta \subseteq B$ or $x + (t-N)\Delta \subseteq B$, also as a subset of $\mathbb{Z}^d$.

Proof. Consider $\Delta = \{e_1, \ldots, e_d\}$ as an element of $\mathbb{Z}^{d^2}$ where $e_i \in \mathbb{Z}^d$. Define

$$\tau(\Delta) = \inf_{m\notin\{0,e\}, x\in[0,e]} |m - x|_\infty$$

Let $0 < \varepsilon < \tau(\Delta)$. We may assume that the simplex is primitive. By our assumption, there is $x \in [1,N]^d, t \in [1, N-1]$ such that $x + te_j \in B + N\mathbb{Z}^d$ for all $1 \leq j \leq d$. Hence for each j, there is $m_j \in \mathbb{Z}^d$ such that $|te_j - Nm_j|_\infty \leq \varepsilon N$ i.e. $|(t/N)\Delta - m|_\infty \leq \varepsilon$ where $m = (m_1, \ldots, m_d) \in \mathbb{Z}^{d^2}$. Since $0 < t/N < 1$ and $\varepsilon < \tau(\Delta)$ we have $m = 0$ or Δ. If $m = 0$ then $|te|_\infty \leq \varepsilon N$. Since $x \in B$ we have $x + te \subseteq B \subseteq \mathbb{Z}^d$. Similarly, if $m = e$ then $x + (t-N)e \subseteq B \subseteq \mathbb{Z}^d$. $\square$

Lemma 4.7.2 (Pigeonhole Principle for W-trick). [14] *Let* $A_1 := \{n \in [1, N/W]^d;\ Wn + b \in A\}$ *and* $A' = A_1 \cap [\varepsilon_1 N', \varepsilon_2 N']^d$. *By the Prime Number Theorem there is a prime* N' *so that* $\varepsilon_2 N' = N_1(1 + o_{N_1\to\infty}(1))$. *We will work on* $\mathbb{Z}/N'$. *We can choose* $b \in \mathbb{Z}^d, \varepsilon_1, \varepsilon_2$ *in the definition of* ν *so that*

$$|A_1 \cap [\varepsilon_1 N', \varepsilon_2 N']^d| \geq \frac{\alpha\, \varepsilon_2^d}{2} \; \frac{(N')^d\, W^d}{(\log N')^d\, \phi(W)^d}.$$

Proof. Let N, W be sufficiently large positive integers and assume that $|A| \geq \alpha\, |\mathcal{P}_N|^d$ for a set $A \subseteq \mathcal{P}_N^d$. By the pigeonhole principle choose $b = (b_j)_{1\leq j\leq d}$ so that b_j is relative prime to W for each j, and

$$|A \cap ((W\mathbb{Z})^d + b)| \geq \alpha\, \frac{N^d}{(\log N)^d\, \phi(W)^d}, \tag{4.7.1}$$

where ϕ is the Euler totient function. Set $N_1 := N/W$ and $A_1 := \{n \in [1, N_1]^d;\ Wn + b \in A\}$. Choose $\varepsilon_2 > 0$ so that $2\varepsilon_2 < \tau(\Delta)$. Hence

$$\alpha \frac{W^d N_1^d}{\phi(W)^d \log^d(WN_1)} = \alpha \frac{\epsilon_2^d W^d (N')^d}{\phi(W)^d \log^d(\epsilon_2 W N')} + o_{N,W\to\infty}(1) = \alpha \frac{\varepsilon_2^d (N')^d\, W^d}{(\log N')^d\, \phi(W)^d} + o_{N,W\to\infty}(1)$$

where we used that we choose $W\epsilon_2 = O(1)$. We have from (4.7.1) that

$$|A_1 \cap [1, \varepsilon_2 N']^d| \geq \frac{\alpha\, \varepsilon_2^d}{2} \; \frac{(N')^d\, W^d}{(\log N')^d\, \phi(W)^d}. \tag{4.7.2}$$

By Dirichlet's theorem on primes in arithmetic progressions the number of $n \in [1, N']^d \backslash [\varepsilon_1 N', N']^d$ for which $Wn + b \in \mathcal{P}^d$ is of $O(\varepsilon_1 \frac{WN'}{\phi(W) \log \varepsilon_1 WN'} \times \frac{(WN')^{d-1}}{\phi(W)^{d-1} (\log WN')^{d-1}}) = O(\varepsilon_1 \frac{N'^d W^d}{(\log N')^d\, \phi(W)^d})$, thus (4.7.2) holds for the set $A' := A_1 \cap [\varepsilon_1 N', \varepsilon_2 N']^d$ as well, if $\varepsilon_1 \leq c_d\, \varepsilon_2^d \alpha$ for a small enough constant $c_d > 0$. □

Theorem 4.1.3 implies Theorem 4.1.2. If $x \in A'$ then $\varepsilon_1 N' \leq x_i \leq \varepsilon_2 N'$ and $Wx_i + b_i \in \mathcal{P}$ for $1 \leq i \leq d$, thus by the definition of the Green-Tao measure $\nu_b : [1, N'] \to \mathbb{R}_+$, we have

$$w(x) = \prod_{i=1}^d \nu_{b_i}(x_i) = c_d \left(\frac{\phi(W) \log N}{W} \right)^d. \tag{4.7.3}$$

as $\log N' - \log N \approx \log(\frac{1}{\epsilon_2 W}) = O(1)$, assuming N sufficiently large with respect to W. Thus

$$\mathbb{E}_{x\in\mathbb{Z}_{N'}^d} \mathbf{1}_{A'}(x)\, w(x) \;=\; \frac{c_d |A'|}{(N')^d} \left(\frac{\phi(W) \log N}{W} \right)^d \geq c_d' \varepsilon_2^d \alpha \tag{4.7.4}$$

for some constant $c_d, c_d' > 0$. Applying the contrapositive of Theorem (4.1.3) for the set A' with

[14] If we allow W to grow with N then the choice of b will depend on N; $b = b(N)$.

$\varepsilon := c_d \varepsilon_2^d \alpha$ gives

$$\mathbb{E}_{x \in \mathbb{Z}_{N'}^d,\, t \in \mathbb{Z}_{N'}} \left(\prod_{j=0}^{d} \mathbf{1}_{A'}(x + t v_j) \right) w(x + t\Delta) \geq \delta \tag{4.7.5}$$

with a constant $\delta = \delta(\alpha, \Delta) > 0$ depending only on α and the simplex $\Delta = \{v_0, \dots, v_d\}$ and $\alpha \to 0$ if $\delta \to 0$. Hence in our case δ is bounded above by a positive constant. Now we transfer (4.7.5) to statement about numbers of prime simplices in A'. As in (4.7.3)

$$w(x + t\Delta) \leq C_d \left(\frac{\phi(W) \log N}{W} \right)^{l(\Delta)}, \tag{4.7.6}$$

since all coordinates of $x + t\Delta$ are primes, bigger then R. Thus the number of copies $\Delta' = x + t\Delta$ which are contained in A' as a subset of $\mathbb{Z}_{N'}^d$ is at least $c\, N^{d+1} (\log N)^{-l(\Delta)}$, for some constant $c = c(\alpha, \Delta, W) > 0$ depending only on the initial data α, Δ and the number W. Since $A' \subseteq [\varepsilon_1 N', \varepsilon_2 N']^d$, by Lemma 4.7.1 at least half of the simplices Δ' are contained in A' as a subset of $\mathbb{Z}^d$, and then the simplices $\Delta'' := W\Delta' + b$ are contained in A.

Now choose $W = W(\alpha, \Delta)$ large enough so that Theorem 4.1.3 holds for all sufficiently large N, and then A contain at least $c'(\alpha, \Delta)\, N^{d+1} (\log N)^{-l(\Delta)}$ similar copies of Δ for some constant $c'(\alpha, \Delta) > 0$ depending only on α and the simplex Δ. This proves Theorem 4.1.2 $\square$

4.8 Concluding Remarks

In this chapter, we obtain a more general version of the weighted hypergraph removal lemma. Our analysis is a kind of averaging arguments. A more details analysis in these measure system may be an interesting problem, given many recent developments e.g. citeTZ3 in the theory of uniformity norms ,say, in $\mathbb{Z}^d$.

As we have seen, Szemerédi's theorem type problems in higher dimension are quite interesting. Our method indeed give an explicit bound on the number of prime configurations but it is terrible (of tower type) due to the application of the regularity lemma (This is necessary as demonstrated in [40] but it may not be necessary in removal lemma or multidimensional Szemerédi's theorem, see e.g. [24]). Also we use the weight ν which we may obtain narrow progressions result similar to [113] but we don't know how to model such problems on graph. Another interesting question would also to prove a polynomial progression in this setting or finding asymptotic of linear equation in primes [52] in higher dimensions. There could be an interesting phenomena happen from the correlations of points. There are other interesting modern approaches to multidimensional Szemerédi's theorem in the primes by Tao-Ziegler [111] and Fox-Zhao [68], both relied on the following more advanced tools, the inverse Gowers norm theorem, which currently cannot give any quantitative bound.

4.8.1 Inverse Gowers Norm Theorem and Infinite Linear Forms Condition

The very first application of the full inverse norm conjecture is used to find the asymptotic of number of prime solutions to a system of linear equation of finite complexity [52]. Basically this means no two linear parts of the system is a multiple of each other. The complexity basically measures how many times you have to apply Cauchy-Schwartz's inequality to obtain the *generalized von Neumann theorem*. The notion of complexity is further discussed in [41].

Another number theoretical application is that we can now define a weight with more general linear form conditions. Define a new weight

$$\nu'_{b,W}(n) := \frac{\phi(W)}{W}(\log n)\mathbf{1}_{\mathcal{P}'}(n) \tag{4.8.1}$$

Then by a result of Green-Tao, we have the following infinite linear forms conditions. Due to the technical restriction of the sieve method, we cannot get the infinite linear form conditions for ν in [51] or [52].

Theorem 4.8.1 ([102], Thm. 5.1). *Let $(\psi_1,\dots,\psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t$ be a system of linear forms (hence $\psi_i(0) = 0$). Let $K \subseteq \Big[\lfloor -N/W \rfloor, \lfloor N/W \rfloor\Big]^d$ be a convex body and $b_1,\dots,b_t$ are coprime to W. Then*[15]

$$\sum_{n\in K\cap\mathbb{Z}^d}\prod_{j\in[t]}\nu'_{b_i,W}(\psi_j(n)) = \#\{n \in K\cap\mathbb{Z}^d : \psi_j(n) > 0\ \forall j\} + o((N/W)^d) \tag{4.8.2}$$

This condition is used in [112] and [68] to give different proofs of the main result in this chapter. In [110], they prove analogue of corresponding principle in ergodic theory to the weight setting that allows them deduce this theorem from the analogue theorem in integer case. They constructed a $\mathbb{Z}^d-$ system $(X,\mathcal{B},\mu,(T_h)_{h\in\mathbb{Z}^d})$ and they have to consider the measure of the form $\mu(T_{h_1}\cap\cdots\cap T_{h_k}(A))$ to shows the result. Here k can be arbitrarily large.

Proof in [68] using sampling argument and their method also gives a polynomial progression version of this theorem if we assume the Bateman-Horn conjecture [4] on the asymptotic number of prime points in a given set of polynomials.

Finally we state a nontrivial version of Inverse Gowers Norms conjecture [59].

Theorem 4.8.2 (U^3-inverse theorem). *Let $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ and $\mathcal{H}$ be the Heisenberg group. Let $N > 2$ be a prime and $0 < \eta < \frac{1}{2}$ and $f : \mathbb{Z}_N \to \mathbb{C}$, bounded by 1. Suppose $\|f\|_{U^3} \geq \eta$. Then for some positive integer $m \leq \eta^{-C}$, then is an $N-th$ root of $g \in \mathcal{H}^m$ (i.e. $g^N \in \Gamma$) and a continuous $1-$bounded*

[15] recall that $\nu \geq 0$.

function F on $\mathcal{N}^m$ with Lipschitz constant at most $\exp(\eta^{-C})$ such that

$$\mathbb{E}_{n\in\mathbb{Z}_N} f(n)\overline{F(g^n\Gamma^m)} \geq \exp(-\eta^{-C}) \tag{4.8.3}$$

The constant in (4.8.3) relies on Frieman's theorem on sumset where the current best bound is due to Sander [89]. It is also mentioned (e.g.in [48]) that the proof of the general U^k of this theorem is not very conceptually explain the role of nilsequence and the bound in terrible (due to the use of ultrafilter arguments). It is also mentioned in [48] also possible that we can use smaller class of nilsequences such as eigenfunctions of the Laplacian on free nilpotent Lie group. Link to *approximate subgroup of* $\mathbb{Z}$ may be a new interesting approach to inverse U^3-theorem.

Bibliography

[1] N. ALON, J.H. SPENCER , *The Probabilistic Method* Wiley Series in Discrete Mathematics and Optimization, Fourth edition, 2016.

[2] A. BALOG, *Linear equations in primes*, Mathematika 39.2 (1992): 367-378

[3] M. BATEMAN, N. KATZ, *New bounds on Cap sets*, J. Amer. Math. Soc. 25 (2012), 585-613.

[4] P.T. BATEMAN, R.A. HORN, *A heuristic asymptotic formula concerning the distribution of prime numbers*, Math. Comp., 16 (1962), 363-367.

[5] F. BEHREND, *On sequences of integers containing no arithmetic progression*, Proc. Nat. Acad. Sci. U. S. A., 32:331332, 1946.

[6] V. BERGELSON, A. LIEBMAN, *Polynomial extension of Van der Waerden's and Szemerédi's theorems*, J. Amer. Math. Soc.9 (1996), 3, 725-753.

[7] V. BERGELSON, B. HOST, B. KRA, *Multiple recurrence and nilsequences.*, Inventiones Math. 160, 2. (2005), 261-303.

[8] V. BERGELSON, *Combinatorial and Diophantine Applications of Ergodic Theory, with appendix by A. Liebman, appendix B by A. Quas and M. Weirdl.*, Handbook of Dynamical Systems Volume 1, Part B, 2006, 745869.

[9] T. BLOOM, *A quantitative improvement for Roth's theorem in arithmetic progression*, preprint.

[10] T. BLOOM, G JONES, *A sum product theorem in function fields*, Int. Math. Res. Note. IMRN, rnt125, June 2013.

[11] B. BIRCH, *Forms in many variables*, Invent. Math. 179,(1962):559-644.

[12] J. BOURGAIN *On triples in arithmetic progression* , Geom. Funct. Anal., 9(5):968 984, 1999.

[13] J. BOURGAIN, A. GAMBURD, P. SARNARK, *Affine linear sieve, expander, and sum-product* , Invent. Math. 179,(2010):559-644.

[14] J. BRUDERN, J. DIETMANN, J. LIU, T. WOOLEY, *A Birch-Goldbach theorem*, Archiv der Mathematik, 94(1), (2010): 53-58

[15] COJOCARU, ALINA CARMEN; MURTY, M. RAM, *An introduction to sieve methods and their applications*, London Mathematical Society Student Texts 66. Cambridge University Press (2005). pp. 113134.

[16] D. CONLON, J. FOX, Y. ZHAO, *A relative Szemeredi theorem*, Geom. Func. Anal. 25(3) (2015), 733-762.

[17] D. CONLON, J. FOX, Y. ZHAO, *The Green-Tao Theorem: An Exposition*, EMS Survey in Mathematical Sciences. 1(2) (2014), 249-282.

[18] B. Cook, A. Magyar and T. Titichetrakun, *A Multidimensional Szemerédi's theorem in the primes.* Preprint.

[19] B. Cook, A. Magyar, *Diophantine equations in the primes.* Invent. Math.: v 198/3, 701-737, 2014.

[20] B. Cook, A. Magyar, *On restricted arithmetic progressions over finite fields,* Online J. Anal. Comb., v.7, (2012): 1-10

[21] H. Davenport, *Cubic forms in thirty two variables* Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences Vol. 251, No. 993 (Mar. 12, 1959), 193-232.

[22] H. Davenport, *Analytic Methods for Diophantine equations and Diophantine inequalities* Cambridge University Press, 2005.

[23] W. Duke, Z. Rudnick, P. Sarnak, *Density of integer points on affine homogeneous varieties*, Duke Math. J. 71.1 (1993): 143-179.

[24] J. Fox, *A new proof of graph removal lemma*, Annals. of Math. 174.1 (2011), 561-579.

[25] J. Friedlander, H. Iwaniec, *Opera de Cribro*, Colloquium Publications, American Mathematical Society (2010).

[26] H. Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerdi on arithmetic progressions*, J. Analyse Math. 31 (1977), 204-256.

[27] H. Furstenberg, Y. Katznelson, *An ergodic Szemerdi theorem for commuting transformations*, J. d'Analyse Math., 34 (1978), 275-291.

[28] H. Furstenberg, Y. Katznelson, *An ergodic Szemerdi theorem for IP-systems and combinatorial theory*, J. d'Analyse Math., 45 (1985), 117-168.

[29] H. Furstenberg, Y. Katznelson, *A density version of the Hales-Jewett theorem*, J. d'Analyse Math., 57 (1991), 64-119.

[30] H. Furstenberg, Y. Katznelson, B. Weiss, *Ergodic theory and configurations in set of positive upper density*, Mathematics of Ramsey Theory, Algorithms Combin, no 5 (1990) 184-198.

[31] H. Furstenberg, Y. Katznelson, D. Ornstein, *The ergodic theoretical proof of Szemeredis theorem*, Bull. Amer. Math. Soc. (N.S.) 7 (1982), no. 3, 527552

[32] H. Furstenberg, Y. Katznelson, B. Weiss, *Ergodic theory and configurations in set of positive upper density*, Mathematics of Ramsey Theory, Algorithms Combin, no 5 (1990) 184-198.

[33] H. Furstenberg, Y. Katznelson, *An ergodic Szemerédi theorem for commuting trnasformations*, J. Analyse Math. 31 (1978), 275-291

[34] W. Hurewicz, H. Wallman, *Dimension Theory*, Princeton, N.J. Princeton University Press (1951).

[35] W.T. Gowers, *Quasirandomness, Counting and Regularity for 3-Uniform Hypergraphs*, Combin. Probab. Comput. 15 (2006), no. 1-2, 143-184.

[36] W.T. Gowers, *Hypergraph regularity and the multidimensional Szemerédi theorem*, Annals of Math. 166/3 (2007), 897-946

[37] W.T. GOWERS, *Decompositions, approximate structure, transference, and the Hahn-Banach theorem*, Bull. London Math. Soc. 42 (4) (2010), 573-606

[38] W.T. GOWERS, *A New Proof of Szemerédi's Theorem for Arithmetic Progression of length four*, Geom. Func. Anal. 8(3)(1998), 529-551

[39] W.T. GOWERS, *A New Proof of Szemerédi's Theorem* , Geom. Func. Anal. 11(3)(2001), 465-558

[40] W.T. GOWERS, *Lower bounds of tower-type for Szemeredi's uniformity lemma* , Geom. Func. Anal. 7(2)(1997)

[41] W.T. GOWERS, J. WOLF, *The true complexity of a system of linear equations*, Proc. London. Math. Soc. 100 (3) (2010), no. 1, 155-176.

[42] D. GOLDSTON, J. PINTZ, YILDIRIM, *Primes in tuples I*, Ann. of Math. 170 (2009): 819-862

[43] D. GOLDSTON, C. YILDIRIM, *Higher correlations of divisor sums related to primes I: triple correlations*, Integers: Electronic Journal of Combinatorial Number theory, 3 (2003), 1-66

[44] R. GRAHAM, *Recent trend in Ergodic Ramsey Theory.* Discrete Math, no. 136, (1994), 119-227

[45] B. GREEN, *Roth's Theorem in the Primes.*Annals. of Math. 161.2 (2005), 1609-1636.

[46] B. GREEN, *Finite Field Model in Additive Combinatorics.* In Bridget S Webb, editor, Surveys in combinatorics 2005, pages 127. Cambridge Univ. Press, Cambridge, Cambridge, 2005.

[47] B. GREEN, *Montral notes on quadratic Fourier analysis.* In Additive combinatorics, pages 69102. Amer. Math. Soc., Providence, RI, 2007.

[48] B. GREEN, *Approximate Algebraic Structure.* Proceedings of ICM 2014.

[49] B. GREEN, *A Szemerdi-type regularity lemma in abelian groups, with applications.* Geom. Funct. Anal. 15 (2005), no. 2, 340376.

[50] B. GREEN AND T. TAO, *An arithmetic regularity lemma, an associatd counting lemma, and applications.* An Irregular Mind: Szemerédi is 70 Bolyai Society Mathematical Studies, 335-342.

[51] B. GREEN AND T. TAO, *The Primes contain arbitrarily long arithmetic progressions.* Annals. of Math.(2) 167.2 (2008), 481-547.

[52] B. GREEN AND T. TAO, *Linear equations in primes.* Annals. of Math.(2) 171.3 (2010), 1753-1850.

[53] B. GREEN AND T. TAO, *The quantitative behaviour of polynomial orbits on nilmanifolds.* Annals. of Math.(2) 175 (2012), 465-540.

[54] B. GREEN AND T. TAO, *The Möbius Function is Strongly Orthogonal to Nilsequences.* Annals. of Math.(2) 175 (2012), 541-566.

[55] B. GREEN AND T. TAO, *Restriction theory of Selberg's sieve with applications.* Journal de Théorie des Nombres.18 (2006), 147-182.

[56] B. GREEN AND T. TAO, *New bounds for Szemeredi's theorem, Ia: Progressions of length 4 in finite field geometries revisited.* Preprint.

[57] B. GREEN AND T. TAO, *New bounds for Szemeredi's Theorem, II: A new bound for $r_4(N)$.* Analytic number theory: essays in honour of Klaus Roth, W. W. L. Chen, W. T. Gowers, H. Halberstam, W. M. Schmidt, R. C. Vaughan, eds, Cambridge University Press, 2009. 180-204.

[58] B. GREEN AND T. TAO, *New bounds for Szemeredi's Theorem, III: A polylog bound for* $r_4(N)$. In preparation.

[59] B. GREEN AND T. TAO, *An inverse theorem for the Gowers* U^3-*norm, with applications* Proc. Edinburgh Math. Soc. 51, no. 1, 71-153.

[60] B. GREEN, T. TAO AND T. ZIEGLER, *An inverse theorem for Gowers* $U^{s+1}[N]$*-norm.* Annals. of Math.(2) 176 (2) (2012), 1231-1372.

[61] S. GHORPAGE, T. TAO AND G. LACHAUD, *Etale cohomology, Lefschetz theorems and number of points on singuar varieties over finite fields* Moscow. Math. J.(2)(2002), 589-631.

[62] R. HARTSHORNE, *Algebraic geometry*, Graduate Texts in Mathematics Vol. 52. Springer (1977)

[63] H. HELFGOTT, A. ROTON, *Improving Roth's Theorem in the Primes.* Int. Math. Res. Not. IMRN, (4) (2011), 767783.

[64] B. HOST, N. FRANTZIKINAKIS, *Higher order Fourier analysis of multiplicative functions and applications.*, preprint.

[65] B. HOST AND B. KRA, *Non conventional ergodic averages and nilmanifolds.* Annals. of Math., 161 (2005) 398-488.

[66] B. HOST AND B. KRA, *A point of view in Gowers uniformity norms.* New York J. Math., 18 (2012), 213-248.

[67] L.K. HUA, *Additive theory of prime numbers*, Translations of Mathematical Monographs Vol. 13." Am. Math. Soc., Providence (1965).

[68] J . FOX, Y. ZHAO, *A short proof of multidimensional szemerédi's theorem in the primes* American Journal of Mathematics 137 (2015), 11391145.

[69] J. KOMLS, M. SIMONOVITS, *Szemerdi's regularity lemma and its applications in graph theory*, Combinatorics, Paul Erds is eighty, Vol. 2 (Keszthely, 1993), 295-352, Bolyai Soc. Math. Stud., 2, Jnos Bolyai Math. Soc., Budapest, 1996.

[70] Y. KOHAYAKAWA, T. LUCZSAK, V. RÖDL, *Arithmetic progressions of length three in subsets of a random set*, Acta Arith. 75 (1996), no. 2, 133-163.

[71] B. KRA, *Ergodic methods in additive combinatorics*,In Additive combinatorics, pages 69102. Amer. Math. Soc., Providence, RI, 2007.

[72] J. LENZ, D. MUBAYI, *The poset of hypergraph quasirandomness*, Random Structures & Algorithms. Volume 46, Issue 4, July 2015, 762800.

[73] J. LIU, *Integral points on quadrics with prime coordinates*, Monats. Math. 164.4 (2011): 439-465.

[74] A. MAGYAR AND T. TITICHETRAKUN, *Corners in dense subset of* $\mathcal{P}^d$. Preprint.

[75] A. MAGYAR AND T. TITICHETRAKUN, *Almost prime solutions to diophantine equation of high rank.* Preprint.

[76] L. MATTHIESEN, *Generalized Fourier coefficients of multiplicative functions* , Preprint.

[77] H.L. MONTGOMERY AND R.C. VAUGHN, *Multiplicative number theory I. classical theory.* , Cambridge studies in advanced mathematics, 97, 2006.

[78] E. NASLUND, *On improving Roth's Theorem in the primes.* Mathematica (61) , (2015), 49-62.

[79] K. O. BRYANT, *Sets of integers that do not contain long arithmetic progressions.* The Electronic Journal of Combinatorics, Volume 18, Issue 1 (2011)

[80] S. PRENDIVILLE, *Four variants of the Fourier-analytic transference principle*, Preprint available at http://arxiv.org/abs/1509.09200.

[81] J. PINTZ, *Are there arbitrarily long arithmetic progressions In the sequence of twin primes?* An irregular mind, Volume 21 of the series Bolyai Society Mathematical Studies, 525-559.

[82] B. NAGLE, V. RÖDL, M. SCHACHT, *The counting lemma for regular k-uniform hypergraphs*, Random Structures and Algorithms, 28 (2), (2006), 113-179

[83] V. RÖDL, J. SKOKAN, *Regularity lemma for uniform hypergraphs*, Random Structures and Algorithms, 2004, vol. 25, no. 1, 1-42.

[84] V. RÖDL, M. SCHACHT, *Applications of the regularity lemma for uniform hypergraphs*, Random Structures and Algorithms, 2006, vol. 28, no. 2, 180-194.

[85] V. RÖDL, M. SCHACHT, *Regular partitions of hypergraphs*, Regularity Lemma, Combin. Prob. Comput. 16 (2007), 833-885.

[86] V. RÖDL, M. SCHACHT, *Regular partitions of hypergraphs*, Counting Lemma, Combin. Prob. Comput. 16 (2007), 887-901.

[87] O. REINGOLD, L. TREVISAN, M. TULSIANI, S. VADHAM, *Dense subsets of pseudorandom sets* Electronic Colloquium of Computational Complexity, Report TR08-045 (2008)

[88] O. REINGOLD, L. TREVISAN, M. TULSIANI, S. VADHAM, *New proofs of the Green-Tao-Ziegler dense model theorem: an exposition* arXiv:0806.0381.

[89] T. SANDER, *On the Bogolyubov-Ruzsa lemma.* Anal. PDE, 5 (2) (2012), 627-655.

[90] T. SANDER, *The structure theory of set additions revisited.* Bull. Amer. Math. Soc 50, (2013), 93-127.

[91] K.F. ROTH, *On certain sets of integers* J. London Math Society, 28 (1953): 104-109.

[92] I. RUZSA, E. SZEMERDI, *Triple systems with no six points carrying three triangles,* Colloq. Math. Soc. J. Bolyai, 18 (1978), 939-945.

[93] T. SCHEON, I. D. SHKREDOV, *Roth's theorem in many variables*, Israel Journal of Mathematics January 2014, Volume 199, Issue 1, pp 287-308

[94] W. SCHMIDT, *The density of integral points on homogeneous varieties*, Acta Math. 154.3 (1985):243-296.

[95] G. SHIMURA, *Reduction of algebraic varieties with respect to a discrete valuation of the basic field*, Amer. J. Math. (1955): 134-176.

[96] I.D. SHKREDOV, *On one problem of Gowers*, Izv. RAN. Ser. Mat. 70.2 (2006) 179-221

[97] J. SOLYMOSI, *Note on a generalization of Roth theorem*, Combinatorial Number Theory and Additive Group Theory, Part of the series Advanced Courses in Mathematics - CRM Barcelona, pp 299-314.

[98] J. SOLYMOSI, *Incidences and the spectra of graphs*, Discrete and Computational Geometry, Algorithms Combin. 25, (2003), 825-827

[99] E. SZEMERÉDI, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. 27 (1975), 299-345

[100] T. TAO, *The ergodic and combinatorial approaches to Szemerédi's theorem*, Centre de Recerches Mathématiques CRM Proceedings and Lecture Notes, 43 (2007), 145–193.

[101] T. TAO, *A quantitative ergodic theory proof of Szemerédi's theorem* , Electron. J. Combin. 13 (2006). 1 No. 99, 1–49.

[102] T. TAO, *The Gaussian primes contain arbitrarily shaped constellations*, J. Analyse Math., 99/1 (2006), 109-176

[103] T. TAO, *Szemerdi's regularity lemma revisited*, Contrib. Discrete Math. 1 (2006), pp.828.

[104] T. TAO, *A variant of the hypergraph removal lemma*, Journal of Combinatorial Theory, Series A 113.7 (2006): 1257-1280

[105] T. TAO, *Higher order Fourier analysis*, American Mathematical Society, Graduate Studies in Mathematics Volume: 142 (2012).

[106] T. TAO, *The prime tuples conjecture, sieve theory, and the work of Goldston-Pintz-Yildirim, Motohashi-Pintz, and Zhang* (2013). Retrieved on July 16, 2015. Available at https://terrytao.wordpress.com/2013/06/03/the-prime-tuples-conjecture-sieve-theory-and-the-work-of-goldston-pintz-yildirim-motohashi-pintz-and-zhang/

[107] T. TAO, *254A, Notes 4: Some sieve theory* (2015). Retrieved on July 16, 2015. Available at https://terrytao.wordpress.com/2015/01/21/254a-notes-4-some-sieve-theory/

[108] T. TAO, *A symmetric formulation of the Croot-Lev-Pach-Ellenberg-Gijswijt capset bound* (2016). Retrieved on May 24, 2016. Available at https://terrytao.wordpress.com/2016/05/18/a-symmetric-formulation-of-the-croot-lev-pach-ellenberg-gijswijt-capset-bound/

[109] T. TAO, V. VU, *Additive Combinatorics* Cambridge University Press, 2006.

[110] T. TAO AND T. ZIEGLER, *The primes contain arbitrarily long polynomial progressions*, Acta Math., 201 (2008), 213-305

[111] T. TAO AND T. ZIEGLER, *multi-dimensional Szemerdi theorem for the primes via a correspondence principle*, Israel Journal of mathematics, 207 (1) (2015), 203-228.

[112] T. TAO AND T. ZIEGLER, *Narrow gaps in the primes*, "Analytic Number Theory" in honour of Helmut Maier's 60th birthday, to appear.

[113] T. TAO AND T. ZIEGLER, *Concatenation theorems for the Gowers uniformity norm and application* , Preprint available at http://arxiv.org/abs/1603.07815

[114] P. VARNAVIDES, *On Certain Sets of Positive Density* , J. London Math Soc. 39 (1959), 358-360.

[115] T. ZIEGLER, *Universal characteristic factors and Furstenberg average*, J. Amer. Math. Soc. 20 (2007), 53-97

[116] T. ZIEGLER, *Linear equations in primes and dynamics of nilmanifolds*, Proceedings of ICM 2014.