NOTES ON BJÖRKLUND-GORODNIK CLT

D. DOLGOPYAT, C. DONG, A. KANIGOWSKI, AND P. NÁNDORI

1. Introduction

In [1], Björklund and Gorodnik proved Central Limit Theorem for exponentially mixing actions of groups with subexponential growth. In this note we review the proof of this remarkable result in the case where the group is an abelian Lie group (which is thus isomorphic to $\mathbb{R}^{d_1} \times \mathbb{Z}^{d_2}$) and point out two useful extensions of the main result of [1]: the Central Limit Theorem for arrays (Theorem 2.3) and Functional Central Limit Theorem (Theorem 5.2). We follow the main ideas of the proof from [1], however, to make the text more self contained we use the method of moments rather than the method of cumulants.

2. The result.

Let $H = \mathbb{R}^{d_1} \times \mathbb{Z}^{d_2}$, M be a compact manifold and $G : H \times M \to M$ be an C^1 action of H preserving a measure μ. We denote $d = d_1 + d_2$.

Definition 2.1. We say that G is exponentially mixing of all orders if for each $p \in \mathbb{N}$ there exist constants $\bar{C}_p, \bar{c}_p > 0$ such that

$$\left| \mu \left(\prod_{j=1}^{p} A_j(G_{t_j} y) \right) - \prod_{j=1}^{p} \mu(A_j) \right| \leq \bar{C}_p e^{-\bar{c}_p \Delta(t_1, \ldots, t_p)} \prod_{j=1}^{p} \|A_j\|_{C^1} \tag{2.1}$$

where $\Delta(t_1, \ldots, t_p)$ is the gap

$$\Delta = \min_{i \neq j} \|t_i - t_j\|.$$

We shall assume throughout this note that $G = G_t$ is exponentially mixing of all orders and that the action G satisfies

$$\|G_t\|_{C^1} \leq C K^{\|t\|}, \tag{2.2}$$

for some $C, K > 0$.

Remark 2.2. We note that (2.2) is always satisfied if M is compact and the map $(t, x) \mapsto D_x G_t(x)$ is continuous. Indeed, in this case, we choose $K = \max_{t \in H, \|t\| \leq 1} \|G_t\|_{C^1}$ and use the fact that that for $t = (t_1, ..., t_d)$, $G_t = \prod_i G_{e_i}^{\lfloor t_i \rfloor} G_{(t_i - \lfloor t_i \rfloor) e_i}$, where e_i is the unit vector in the ith coordinate direction.

In the statement below we regard H as a subset of $\mathbb{R}^d$.

Theorem 2.3. Let $\{\mathfrak{m}_T\}$ be a family of finite measures on $\mathbb{R}^d$ and let $\{A_{t,T}\}_{t\in H, T\in\mathbb{R}}$ be a family of real valued functions on M such that $\|A_{t,T}\|_{C^1(M)}$ is uniformly bounded and $\mu(A_{t,T})\equiv 0$. Set $\mathcal{S}_T(x) := \int_{\mathbb{R}^d} A_{t,T}(G_t x)d\mathfrak{m}_T(t)$. Suppose that

(a) $\lim_{T\to\infty} \|\mathfrak{m}_T\| = \infty$ where $\|\mathfrak{m}_T\| = \mathfrak{m}(\mathbb{R}^d)$

(b) For each $r\in\mathbb{N}$, $r\geq 3$ and for each $K>0$,

$$\lim_{T\to\infty} \int \mathfrak{m}_T^{r-1}(B(t, K\ln\|\mathfrak{m}_T\|))d\mathfrak{m}_T(t) = 0,$$

where $B(t,u)\subset\mathbb{R}^d$ denotes a ball centered at $t\in\mathbb{R}^d$ with radius $u>0$;

(c) $\lim_{T\to\infty} V_T = \sigma^2$ where

$$V_T := \int \mathcal{S}_T^2(x)d\mu(x) = \iiint A_{t_1,T}(G_{t_1}x)A_{t_2,T}(G_{t_2}x)d\mathfrak{m}_T(t_1)\mathfrak{m}_T(t_2)d\mu(x).$$

Then $\mathcal{S}_T$ converges as $T\to\infty$ to normal distribution with zero mean and variance σ^2.

This theorem is proven in [1] in case $A_{t,T}$ does not depend on t,T however the proof does not use this assumption, see Section 4.

3. Preliminary estimates

3.1. Correlations with lonely terms.

We start with the following useful consequence of (2.1).

Lemma 3.1. For each $p\in\mathbb{N}$ there exist constants $C_p, c_p>0$ such that if $\mu(A_j)=0$ for all $1\leq j\leq p$ then

$$\left|\mu\left(\prod_{j=1}^p A_j(G_{t_j}y)\right)\right| \leq C_p e^{-c_p\mathfrak{l}(t_1,\dots,t_p)}\prod_{j=1}^p \|A_j\|_{C^1}, \tag{3.1}$$

where $\mathfrak{l}(t_1,\dots,t_p)$ is the *loneliness index*

$$\mathfrak{l} = \max_i \min_{j\neq i} |t_i - t_j|.$$

Proof. We use induction on p. If $p=1$ then the result is clear since A_1 has zero mean. If $p=2$ then (2.1) and (3.1) are equivalent.

Next suppose that (3.1) is known for $p<p_0$ where $p_0>2$, and let us show that it holds for $p=p_0$. Fix a small $\kappa>0$ and consider two cases:

(a) $\Delta(t_1,\dots,t_{p_0}) > \kappa\mathfrak{l}(t_1,\dots t_{p_0})$. In this case (3.1) follows from (2.1) provided that $c_{p_0}\leq\kappa\bar{c}_{p_0}$.

(b) $\Delta(t_1,\dots,t_{p_0}) = \|t_{i_0}-t_{j_0}\| \leq \kappa\mathfrak{l}(t_1,\dots t_{p_0})$. Write

$$A_{i_0}(G_{t_{i_0}}y)A_{j_0}(G_{t_{j_0}}y) = \tilde{A}(G_{t_{i_0}}y)$$

where $\tilde{A}(y) = A_{i_0}(y)A_{j_0}(G_{t_{j_0}-t_{i_0}}y)$. We now decompose $\tilde{A}(y) = \mu(\tilde{A}) + \left[\tilde{A}-\mu(\tilde{A})\right]$ and apply (3.1) with $p=p_0-2$ and $p=p_0-1$ respectively. In the case of $\mu(\tilde{A})$, we apply the inductive hypothesis to $\{t_1,\dots t_{p_0}\}\setminus\{t_{i_0},t_{j_0}\}$. Clearly, $\|\mu(\tilde{A})\|_{C^1}\leq\|A_{i_0}\|_{C^0}\|A_{j_0}\|_{C^0}$ and our assumption $\Delta(t_1,\dots,t_{p_0}) = \|t_{i_0}-t_{j_0}\|\leq\kappa\mathfrak{l}(t_1,\dots t_{p_0})$ implies that the maximum in the definition of $\mathfrak{l}$ is achieved for $i \notin \{i_0, j_0\}$ and so removing those points does not decrease the loneliness index.

In the case of $\tilde{A}-\mu(\tilde{A})$, we apply the inductive hypothesis to $\{t_1, \dots t_{p_0}\} \setminus \{t_{i_0}\}$. Note that

$$\begin{aligned} \|\tilde{A}-\mu(\tilde{A})\|_{C^1} &\le \|\tilde{A}\|_{C^0} + \|\nabla \tilde{A}\|_{C^0} \\ &\le \|A_{i_0}\|_{C^0}\|A_{j_0}\|_{C^0} + (1+CK^{\|t_{j_0}-t_{i_0}\|})\|A_{i_0}\|_{C^1}\|A_{j_0}\|_{C^1} \end{aligned} \tag{3.2}$$

by (2.2). Thus

$$\left(\prod_{i\neq i_0}\|A_i\|_{C_1}\right)\|\tilde{A}-\mu(\tilde{A})\|_{C^1} \le (2+CK^{\|t_{j_0}-t_{i_0}\|})\prod_{i=1}^{p_0}\|A_i\|_{C^1} \le 2Ce^{(\ln K)\kappa\mathfrak{l}(t_1,\dots t_{p_0})}\prod_{i=1}^{p_0}\|A_i\|_{C^1}.$$

This proves (3.1) with $\bar{c}_{p_0} = \min\{\bar{c}_{p_0-2}, \bar{c}_{p_0-1} - \kappa \ln K\}$ provided that $\bar{c}_{p_0}$ is positive which can be ensured by choosing $\kappa = \bar{c}_{p_0-1}/(2\ln K)$. □

3.2. **Well separated clusters.** Fix a positive integer m and a partition $\mathcal{P} = \mathcal{P}_m$ of the set $\{1, 2, ..., m\}$. For some $R, T > 1$ and $(t_1, ..., t_m) \in \mathbb{R}^{md}$, let $\mathcal{G} = \mathcal{G}(R, T, t_1, ..., t_m)$ be the graph on $\{1, 2, ..., m\}$ where i and j are connected by an edge if and only if $\|t_i - t_j\| \le R \ln T$.

We say that $(t_1, ..., t_m)$ is in $\mathfrak{C}(\mathcal{P}, R, T)$ if the set of connected components of $\mathcal{G}$ is $\mathcal{P}$. We say that $(t_1, ..., t_m) \in \mathfrak{W}(\mathcal{P}, R, T)$ if $(t_1, ..., t_m) \in \mathfrak{C}(\mathcal{P}, R, T)$ and for any i, j not in the same atom of $\mathcal{P}$, $\|t_i - t_j\| \ge R^2 \ln T$.

Here $\mathfrak{C}$ stands for clusters and $\mathfrak{W}$ is stands for well separated clusters.

Lemma 3.2. For every positive integer m there is $R_0(m)$ so that for all $R \ge R_0(m)$ and for any partition $\mathcal{P} = \{P_1, ..., P_s\}$ of the set $\{1, ..., m\}$,

$$\sup_{(t_1,\dots,t_m)\in\mathfrak{W}(\mathcal{P},R,T)} \left| \mu\left(\prod_{j=1}^m A_j(G_{t_j}x)\right) - \prod_{k=1}^s \mu\left(\prod_{j\in P_k} A_j(G_{t_j}x)\right)\right| \le CT^{-10m}. \tag{3.3}$$

Proof. It is clearly enough to prove the lemma for a fixed partition $\mathcal{P}$. Denote $B_k = \prod_{j\in P_k} A_j(G_{t_j}x)$ and $\bar{B}_k = B_k - \mu(B_k)$. Then we have

$$\mu\left(\prod_{j=1}^m A_{t_j}(G_{t_j}x)\right) = \mu\left(\prod_{k=1}^s B_k(x)\right) = \mu\left(\prod_{k=1}^s \left[\bar{B}_k(x) + \mu(B_k)\right]\right)$$

$$= \sum_J \left(\prod_{k\in\{1,\dots,s\}\setminus J} \mu(B_k)\right)\mu\left(\prod_{k\in J}\bar{B}_k(x)\right) =: \sum_J Q_J$$

where the sum is over all subsets J of $\{1, \dots, s\}$.

If $J = \emptyset$, then Q_J is the second term on the left hand side of (3.3). If $J = \{k\}$, then $\mu(\bar{B}_k) = 0$ and so $Q_J = 0$. Let us now fix J with $|J| \ge 2$. Then, as in (3.2), we have $\|B_k\|_{C^1} \le CK^{Rm\ln T}$. Since $|J| \ge 2$ and the clusters are well separated, the exponential mixing of all orders implies

$$\mu\left(\prod_{k\in J}\bar{B}_k(x)\right) = O\left(e^{-\bar{c}_s R^2 \ln T} K^{Rm\ln T}\right).$$

We conclude

$$\sum_{J:|J|\geq 2} O\left(e^{-\bar{c}_s R^2 \ln T} K^{Rm\ln T}\right) = O\left(2^m e^{-\bar{c}_s R^2 \ln T} K^{Rm\ln T}\right) = O(T^{-10m}),$$

where the last equation holds for $R \geq R_0(m)$ sufficiently large. □

3.3. **Enforcing the separation.** We need some notations. Let $\mathfrak{P}_m$ be the set of all partitions of $\{1,...,m\}$ and let us write $\mathcal{Q} \leq \mathcal{P}$ for $\mathcal{P}, \mathcal{Q} \in \mathfrak{P}_m$ if for all $P \in \mathcal{P}$ there is $Q \in \mathcal{Q}$ with $P \subseteq Q$. As usual, $\mathcal{Q} < \mathcal{P}$ means that $\mathcal{Q} \leq \mathcal{P}$ and $\mathcal{Q} \neq \mathcal{P}$. Thus if $\mathcal{Q} < \mathcal{P}$, then $\mathcal{Q}$ is coarser than $\mathcal{P}$ in the sense that all atoms in $\mathcal{Q}$ are unions of some atoms of $\mathcal{P}$ and at least one union is non-trivial, i.e. $\mathcal{Q} \neq \mathcal{P}$. In particular if $\mathcal{Q} < \mathcal{P}$ then the number of atoms in $\mathcal{Q}$ is smaller than the number of atoms in $\mathcal{P}$.

Let us also write $\bar{t} = (t_1, ..., t_m)$.

Lemma 3.3. Given $\bar{t} \in \mathbb{R}^{md}$ we can find a unique non-negative integer $L = L(\bar{t}) \leq m$ and a sequence of partitions $\mathcal{P} = \mathcal{P}_0 > \mathcal{P}_2 > ... > \mathcal{P}_L$ so that $\bar{t} \in \mathfrak{C}(\mathcal{P}_\ell, R^{2^\ell}, T) \setminus \mathfrak{W}(\mathcal{P}_\ell, R^{2^\ell}, T)$ for all $\ell = 0, ..., L-1$and $\bar{t} \in \mathfrak{W}(\mathcal{P}_L, R^{2^L}, T)$.

Proof. We have $\bar{t} \in \mathfrak{C}(\mathcal{P}_0, R, T)$. If $\bar{t} \in \mathfrak{W}(\mathcal{P}_0, R, T)$, then $L = 0$. If not, then there are at least two connected components whose distance is less than $R^2 \ln T$. Thus $\bar{t} \in \mathfrak{C}(\mathcal{P}_1, R^2, T)$ with some $\mathcal{P}_1 < \mathcal{P}_0$. If $\bar{t} \in \mathfrak{W}(\mathcal{P}_1, R^2, T)$, then $L = 1$. We continue this process until $\bar{t} \in \mathfrak{W}(\mathcal{P}_L, R^{2^L}, T)$ holds. Since at each step we merge at least two components, the process will finish in less than m steps. □

Given a sequence of partitions $\mathcal{S} = \langle \mathcal{P}_0, ..., \mathcal{P}_L \rangle$ with $\mathcal{P}_0 > \mathcal{P}_1 ... > \mathcal{P}_L$, we denote by $\mathfrak{S}(\mathcal{S})$ for the set of points $\bar{t} \in \mathbb{R}^{md}$ for which Lemma 3.3 holds with the sequence $\mathcal{S}$. Thus

$$\mathbb{R}^{md} = \bigsqcup_{\mathcal{S}} \mathfrak{S}(\mathcal{S}),$$

where $\bigsqcup$ denotes disjoint union.

4. Multiple exponential mixing and CLT. Proof of Theorem 2.3

We will use Lemmas 3.1–3.3 to compute the asymptotics of moments of $\mathcal{S}_T$. It suffices to show that for each $p \in \mathbb{N}$

$$\lim_{T\to\infty} \mathbb{E}\left(\mathcal{S}_T^{2p}\right) = (2p-1)!!\sigma^{2p}, \tag{4.1}$$

and

$$\lim_{T\to\infty} \mathbb{E}\left(\mathcal{S}_T^{2p-1}\right) = 0. \tag{4.2}$$

To simplify the notation we suppose that $\|\mathfrak{m}_T\| = T$. We also drop the subscript T in $A_{t,T}$. We have

$$\mathbb{E}\left(\mathcal{S}_T^m\right) = \int \cdots \int \left(\int A_{t_1}(G_{t_1}x) \dots A_{t_m}(G_{t_m}x) d\mu(x) \right) d\mathfrak{m}_T^m(\bar{t}). \tag{4.3}$$

We will discuss the contribution to (4.3) of terms with different cluster combinatorics.

To further simplify notations, we write

$$\alpha(\bar{t}) = \mu\left(\prod_{j=1}^m A_{t_j}(G_{t_j}x) \right).$$

Thus we have to prove that

$$\sum_{\mathcal{S}} \int_{\bar{t}\in\mathfrak{S}(\mathcal{S})} \alpha(\bar{t}) d\mathfrak{m}_T^m(\bar{t}) = \mathbb{1}_{m \text{ is even}}(m-1)!!\sigma^m + o_{T\to\infty}(t). \tag{4.4}$$

We say that a partition is *pairing* if all atoms have size 2. In fact, we prove the following statement. For any sequence $\mathcal{S} = \langle \mathcal{P}_0, ..., \mathcal{P}_L \rangle$ of partitions in $\mathfrak{P}_m$

$$\int_{\bar{t}\in\mathfrak{S}(\mathcal{S})} \alpha(\bar{t}) d\mathfrak{m}_T^m(\bar{t}) = \mathbb{1}_{\ L=0, \mathcal{P}_0 \text{ is pairing}} \sigma^m + o(1). \tag{4.5}$$

(4.5) implies (4.4) since in case m is even, there are $(m-1)!!$ pairings and if m is odd, there are no pairings.

The proof of (4.5) relies on an auxiliary statement. We claim that for any sequence $\mathcal{S} = \langle \mathcal{P}_0, ..., \mathcal{P}_L \rangle$ of partitions in $\mathfrak{P}_m$ and for any $\mathcal{Q} \geq \mathcal{P}_L$ so that $\mathcal{Q}$ does not contain singletons,

$$\int_{\bar{t}\in\mathfrak{S}(\mathcal{S})} \beta_{\mathcal{Q}}(\bar{t}) d\mathfrak{m}_T^m(\bar{t}) = \mathbb{1}_{L=0, \mathcal{P}_0 \text{ is pairing}} \sigma^m + o(1) \tag{4.6}$$

where

$$\beta_{\mathcal{Q}}(\bar{t}) = \prod_{Q\in\mathcal{Q}} \mu\left(\prod_{j\in Q} A_{t_j}(G_{t_j}x) \right).$$

We will prove (4.5), (4.6) by induction on m.

It is easy to check (4.5) for $m = 1, 2$. If $m = 1$, then the only possible sequence is $\mathcal{S} = \langle\{\{1\}\}\rangle$ and since $\mu(A_{t,j}x) = 0$ for all t_j, (4.5) follows.

If $m = 2$, then there are three possible sequences:

$$\mathcal{S} = \langle\{\{1,2\}\}\rangle,\ \mathcal{S}' = \langle\{\{1\},\{2\}\},\{\{1,2\}\}\rangle,\ \mathcal{S}'' = \langle\{\{1\},\{2\}\}\rangle.$$

Thus

$$\mathfrak{S}(\mathcal{S}) = \{(t_1,t_2) : \|t_1 - t_2\| \leq R\ln T\},$$

$$\mathfrak{S}(\mathcal{S}') = \{(t_1,t_2) : R\ln T < \|t_1 - t_2\| \leq R^2\ln T\}, \quad \mathfrak{S}(\mathcal{S}'') = \{(t_1,t_2) : R^2\ln T < \|t_1 - t_2\|\}.$$

We see that

$$\int_{\mathfrak{S}(\mathcal{S})} \alpha d\mathfrak{m}^2(\bar{t}) = \int_{\mathbb{R}^{2d}} \alpha d\mathfrak{m}^2(\bar{t}) = \sigma^2 + o(1)$$

where the first equality uses the exponential mixing while the second one holds by assumption (c). Likewise $\int_{\mathfrak{S}(\mathcal{S}')} |\alpha| + \int_{\mathfrak{S}(\mathcal{S}'')} |\alpha| = o(1)$ by the exponential mixing.

Next, (4.6) is vacuous if $m = 1$ since a partition of a one point set should contain singletons. If $m = 2$, the only non trivial case if when $\mathcal{Q} = \mathcal{P}_L = \{1, 2\}$ in which case (4.6) reduces to (4.5).

Now assume that (4.5), (4.6) hold for $1, ..., m - 1$ with $m \geq 3$.

To prove (4.5), (4.6) for m, we first note that in the case when $\mathcal{P}_0$ contains singletons, the required estimate follows from Lemma 3.1. Namely (4.5) is a direct consequence of Lemma 3.1, while (4.6) is obtained by applying Lemma 3.1, to the atom of $\mathcal{Q}$ which contains one of the singletons of $\mathcal{P}_0$.

Thus we assume henceforth that all atoms of $\mathcal{P}_0$ (and so all atoms of $\mathcal{P}_l$ for $l \geq 0$) have size at least 2. In particular, the only way $\mathcal{P}_L$ can be pairing is when $L = 0$.

Now for the given $m \geq 3$, we proceed by induction on $|\mathcal{P}_L|$: the number of atoms in $\mathcal{P}_L$. If $|\mathcal{P}_L| = 1$, then we need to prove that for all $\mathcal{Q} \in \mathfrak{P}_m$, (4.6) holds (the case $\mathcal{Q} = \mathcal{P}_L$ gives (4.5). This is immediate from

$$\left| \int_{\bar{t} \in \mathfrak{S}(\mathcal{S})} \beta_{\mathcal{Q}}(\bar{t}) d\mathfrak{m}_T^m(\bar{t}) \right| \leq \int_{\bar{t} \in \mathfrak{S}(\mathcal{S})} \|A\|_\infty^m d\mathfrak{m}_T^m(\bar{t})$$

$$\leq \|A\|_\infty^m \int_{t_1 \in \mathbb{R}^d} \mathfrak{m}_T^{m-1}(B(t_1, mR \ln T)) d\mathfrak{m}_T(t_1) = o(1),$$

where we used assumption (b).

Now assume that (4.5), (4.6) hold for all sequences $\mathcal{S}$ so that $|\mathcal{P}_L| < k$ and fix a sequence $\mathcal{S}$ with $|\mathcal{P}_L| = k$. First we prove (4.5). Note that by Lemma 3.2,

$$\int_{\bar{t} \in \mathfrak{S}(\mathcal{S})} \alpha(\bar{t}) d\mathfrak{m}_T^m(\bar{t}) = \int_{\bar{t} \in \mathfrak{S}(\mathcal{S})} \beta_{\mathcal{P}_L}(\bar{t}) d\mathfrak{m}_T^m(\bar{t}) + o(1) =: I_1 + o(1). \tag{4.7}$$

Since now the integrand is a product over variables in $P \in \mathcal{P}_L$, we wish to replace the integration domain with such a product domain as well. Thus we approximate I_1 by

$$I_2 = \int_{\bar{t} \in \mathfrak{S}'(\mathcal{S})} \beta_{\mathcal{P}_L}(\bar{t}) d\mathfrak{m}_T^m(\bar{t})$$

where for $\mathcal{S} = \langle \mathcal{P}_0, ..., \mathcal{P}_L \rangle$, $\mathfrak{S}'(\mathcal{S})$ is defined as

$$\mathfrak{S}'(\mathcal{S}) = \{\bar{t} : \forall P \in \mathcal{P}_L, \forall \ell = 0, ..., L : (t_j, j \in P) \in \mathfrak{C}(\{Q \in \mathcal{P}_\ell : Q \subseteq P\}, R^{2^\ell}, T)\}.$$

In other words, $\mathfrak{S}'(\mathcal{S})$ is obtained from $\mathfrak{S}(\mathcal{S})$ by waiving all restrictions in $\mathcal{S}$ which come from different atoms of $\mathcal{P}_L$. Hence $\mathfrak{S}'(\mathcal{S})$ is a product $\mathfrak{S}'(\mathcal{S}) = \prod_{P \in \mathcal{P}_L} \mathfrak{S}(\mathcal{S}_P)$ where $\mathcal{S}_P = \langle \mathcal{R}_0, \ldots \mathcal{R}_{L^*} \rangle$ with $\mathcal{R}_\ell = \mathcal{R}_\ell(P) = \{Q \in \mathcal{P}_\ell : Q \subseteq P\}$ and

$$L^* = L^*(P) = L \wedge \inf\{\ell = 0, ..., L - 1 : \mathcal{R}_\ell = \mathcal{R}_{\ell+1}\}.$$

Thus

$$I_2 = \prod_{P \in \mathcal{P}_L} \int_{\mathfrak{S}(\mathcal{S}_P)} \mu \left(\prod_{j \in P} A_{t_j}(G_{t_j} x) \right) \prod_{j \in P} d\mathfrak{m}_T(t_j).$$

Note that $\mathcal{P}_L$ is pairing iff ($\mathcal{P}_0$ is pairing and $L = 0$) iff for all P, $\mathcal{S}_P$ has a single partition which is the pairing of a set of two elements. Therefore by the inductive hypothesis (namely, (4.5) for smaller m)

$$I_2 = \mathbb{1}_{\mathcal{P}_L \text{ is pairing}} \sigma^m + o(1).$$

It remains to verify that $|I_1 - I_2| = o(1)$. To this end, we note that $\mathfrak{S}(\mathcal{S}) \subset \mathfrak{S}'(\mathcal{S})$ and

$$\mathfrak{S}'(\mathcal{S}) \setminus \mathfrak{S}(\mathcal{S}) = \bigsqcup_{L'=L}^{|\mathcal{P}_0|-1} \bigsqcup_{\substack{\mathcal{P}'_0 > \mathcal{P}'_2 > \ldots > \mathcal{P}'_{L'}: \\ \forall \ell = 0, \ldots, L: \mathcal{P}'_\ell \leq \mathcal{P}_\ell,\, \mathcal{P}'_{L'} < \mathcal{P}_L}} \mathfrak{S}(\langle \mathcal{P}'_0, \ldots, \mathcal{P}'_{L'} \rangle).$$

Now the inductive hypothesis, namely (4.6) applied with $\mathcal{S}' = \langle \mathcal{P}'_0, \ldots, \mathcal{P}'_{L'} \rangle$ as in the previous display and with $\mathcal{Q} = \mathcal{P}_L$, implies that $|I_1 - I_2| = o(1)$ (indeed, since $\mathcal{P}_0$ does not contain singletons, $\mathcal{P}_L$ does not contain singletons either and since $\mathcal{P}'_{L'}$ has strictly less atoms than $\mathcal{P}_L$, it follows that $\mathcal{P}'_{L'}$ cannot be pairing). We have verified (4.5).

The proof of (4.6) is similar. Consider first the case where $\mathcal{P}_L$ is pairing. In this case $\mathcal{Q} = \mathcal{P}_L$ since $\mathcal{Q}$ does not contain singletons and so the left hand side of (4.6) is I_1. Thus (4.6) follows as above.

Next, consider the case where $\mathcal{P}_L$ is not pairing. We need to show that

$$\bar{I}_2 = o(1), \quad \bar{I}_1 - \bar{I}_2 = o(1)$$

where

$$\bar{I}_1 = \int_{\bar{t} \in \mathfrak{S}(\mathcal{S})} \beta_{\mathcal{Q}}(\bar{t}) d\mathfrak{m}_T^m(\bar{t}), \quad \bar{I}_2 = \int_{\bar{t} \in \mathfrak{S}'(\mathcal{S})} \beta_{\mathcal{Q}}(\bar{t}) d\mathfrak{m}_T^m(\bar{t})$$

(note that the difference between $\bar{I}_j$ and I_j is that $\mathcal{P}_L$ is replaced by $\mathcal{Q}$). To bound $\bar{I}_2$ we write it as

$$\bar{I}_2 = \prod_{P \in \mathcal{P}_L} \int_{\mathfrak{S}(\mathcal{S}_P)} \beta_{\mathcal{Q}(P)}(t_j, j \in P) \prod_{j \in P} d\mathfrak{m}_T(t_j)$$

where $\mathcal{Q}(P)$ denotes the restriction of $\mathcal{Q}$ to P and apply the inductive hypothesis (noting that at least one atom of $\mathcal{P}_L$ has size greater than 2 and so the corresponding integral is in $o(1)$, while all other integrals are (at least) bounded).

The estimate $\bar{I}_1 - \bar{I}_2 = o(1)$ is proven similarly using the fact that $\mathcal{Q} \geq \mathcal{P}_L > \mathcal{P}'_{L'}$. This completes the proof.

5. Functional CLT

5.1. **Multidimensional observables.**

Corollary 5.1. Theorem 2.3 remains valid if $A_{t,Y}$ are $\mathbb{R}^p$ valued functions for some $p > 1$, provided that condition (c) is updated to require that the matrix V_T with components

$$(V_T)_{\alpha,\beta} = \int (\mathcal{S}_T)_{(\alpha)}(x) (\mathcal{S}_T)_{(\beta)}(x) d\mu(x)$$

satisfies $\lim\limits_{T \to \infty} V_T = \sigma^2$ and the conclusion is convergence to the p-dimensional Gaussian law with covariance matrix σ^2.

Proof. By the Cramér-Wold Theorem, it suffices to show that for each $\xi \in (\mathbb{R}^p)^*$ the product $\langle \xi, S_T \rangle$ converges to a centered normal random variable with variance $\sigma^2_\xi = \langle \sigma^2 \xi, \xi \rangle$. This follows, since for each ξ the assumptions of Theorem 2.3 are satisfied for $A_{t,T,\xi} = \langle \xi, A_{t,T} \rangle$ with the asymptotic variance σ^2_ξ. □

5.2. Følner sets.

Theorem 5.2. Let $S_T = \int_{D_T} A(G_t x) dt$ where $D_T \subset \mathbb{R}^d$ is an increasing ($D_{T_1} \subset D_{T_2}$ for $T_1 \leq T_2$) Følner sequence, such that $\mathrm{Vol}(D_T) = T$ and $\mu(A) = 0$. Then $(W_T(uT))_{u \geq 0} = \left(\frac{1}{\sqrt{T}} S_{uT} \right)_{u \geq 0}$ converges weakly to a Brownian Motion with asymptotic variance

$$\sigma^2 = \int_{\mathbb{R}^d} \mu(A(x) A(G_t x)) dt. \tag{5.1}$$

Proof. STEP I. The fact that $\frac{S_T}{\sqrt{T}}$ converges to normal distribution is proven in [1]. We repeat the argument since it will be used at steps II and III. We verify the assumptions of Theorem 2.3. Let $\mathfrak{m}_T = \frac{\mathrm{Leb}_{D_T}}{\sqrt{T}}$. Then (a) is evident. To check (b), note that for each K, t, $\mathfrak{m}_T(t, K \ln T) \leq C(K) \frac{\ln^d T}{\sqrt{T}}$, so

$$\frac{1}{\sqrt{T}} \int_{D_T} \mathfrak{m}_T^{r-1}(t, K \ln T) dt \leq \frac{\mathrm{Vol}(D_T)}{\sqrt{T}} \times \left(\frac{C(K) \ln^d T}{\sqrt{T}} \right)^{r-1} \to 0$$

as $r > 2$. To check (c), we note that

$$V_T = \frac{1}{T} \iiint_{M \times D_T \times D_T} A(G_{t_1} x) A(G_{t_2} x) d\mu(x) dt_1 dt_2.$$

Denote $\rho(t) = \int A(x) A(G_t x) d\mu(x)$. Fix a large R and split this integral into two parts.

(i) The integral over $\|t_1 - t_2\| > R$ is bounded by

$$\frac{1}{T} \int_{D_T} \left(\int_{\mathbb{R}^d \setminus B(t_1, R)} |\rho(t_2 - t_1)| dt_2 \right) dt_1 \leq C e^{-cR}$$

by the exponential mixing.

(ii) The integral over $\|t_1 - t_2\| \leq R$ is equal to

$$\frac{1}{T} \int_{D_T} \left(\int_{B(t_1, R)} \rho(t_2 - t_1) dt_2 \right) dt_1 + \varepsilon_{R,T}.$$

The first term equals to

$$\frac{\mathrm{Vol}(D_T)}{T} \left(\int_{B(0,R)} \rho(t) dt \right) = \int_{B(0,R)} \rho(t) dt,$$

while the error term is bounded by

$$|\varepsilon_{R,T}| \leq \|A\|^2_{C^0} \times \frac{1}{T} \int_{B(0,R)} \mathrm{Vol}(D_T \setminus (D_T + u)) du \to 0$$

since D_T is a Følner sequence.

Taking R to infinity we obtain that $\lim_{T\to\infty} \dfrac{V(S_T)}{T} = \int_{\mathbb{R}^d} \rho(t)dt$ proving (5.1).

Step II. Convergence of finite dimensional distributions. We need to show that for each $u_1 < u_2 < \dots u_k$

$$\frac{S_{u_1T}}{\sqrt{T}}, \frac{S_{u_2T}-S_{u_1T}}{\sqrt{T}}, \dots, \frac{S_{u_kT}-S_{u_{k-1}T}}{\sqrt{T}}$$

are asymptotically independent Gaussian random variables. By Step I and by the properties of the multivariate Gaussian law, it is sufficient to check that for each $0 \le u_1 < u_2 \le u_3 < u_4$, $\frac{S_{u_2T}-S_{u_1T}}{\sqrt{T}}$ and $\frac{S_{u_4T}-S_{u_3T}}{\sqrt{T}}$ are asymptotically uncorrelated. Notice that for u_1, u_2, u_3, u_4 as above, we have

$$\frac{1}{T}\mu((S_{u_2T}-S_{u_1T})(S_{u_4T}-S_{u_3T})) = \frac{1}{T}\int_{(D_{u_4T}\setminus D_{u_3T})\times(D_{u_2T}\setminus D_{u_1T})} \rho(t_2-t_1)dt_1dt_2$$

Splitting this integral as in Step I and using that the terms with $\|t_2-t_1\| < R$ contribute at most

$$\frac{\|A\|^2_{C^0}}{T}\int_{D_{u_2T}} \mathrm{Vol}(B(t_1,R)\setminus D_{u_2T})dt_1,$$

which tends to zero due to the Følner condition, we conclude the proof.

Step III. Tightness. By a standard argument, it is enough to verify tightness for $u \in [0,1]$. Denote

$$\mathcal{A}_{K,\gamma} = \{\mathcal{W} : [0,1] \to \mathbb{R} : \forall u_1, u_2 \in [0,1] \;\; |\mathcal{W}(u_2) - \mathcal{W}(u_1)| \le K|u_2-u_1|^\gamma\}.$$

We shall show that if $\gamma < \frac{1}{4}$, then for every $\varepsilon > 0$ there is K such that for large T,

$$\mathbb{P}(W_T \notin \mathcal{A}_{K,\gamma}) \le \varepsilon.$$

We say that $[t_1, t_2]$ is *a diadic interval*, if $t_1 = \frac{j}{2^m}$, $t_2 = \frac{j+1}{2^m}$ for some $m \in \mathbb{N}$ and $j \in 0, \dots, 2^m-1$. It suffices to prove that with probability $1-\varepsilon$ it holds that for each diadic interval

$$|W_T(u_2) - W_T(u_1)| \le \bar{K}|u_2-u_1|^\gamma \tag{5.2}$$

since in this case $W_T(\cdot) \in \mathcal{A}_{K,\gamma}$ for a sufficiently large $K > \bar{K}$ as can be seen by decomposing an arbitrary interval into diadic intervals. Let $L = |u_2-u_1|T$.

We claim that

$$\mathbb{E}\left((S_{u_2T}-S_{u_1T})^4\right) \le CL^2. \tag{5.3}$$

To prove this claim, denote $\Delta = S_{u_2T} - S_{u_1T}$. In the proof of (4.1) in Section 4 we saw that

$$\mathbb{E}\left(\Delta^4\right) = 3(\mathbb{E}(\Delta^2))^2(1+o(1)).$$

In the present situation we do not need precise asymptotics of $\mathbb{E}(\Delta^2)$ in case $L \ll T$, instead, we use the upper bound

$$\mathbb{E}(\Delta^2) \le \int_{D_{u_2T}\setminus D_{u_1T}} \int_{\mathbb{R}^d} \rho(t_2-t_1)d\mathfrak{m}(t_1)d\mathfrak{m}(t_2) \le C\mathrm{Vol}(D_{u_2T}\setminus D_{u_1T}) = CL$$

which proves (5.3)

Next, (5.3) and the Markov inequality implies

$$\mathbb{P}\left(\exists j < 2^m : \left|W_T\left(\frac{j+1}{2^m}\right) - W_T\left(\frac{j}{2^m}\right)\right| \geq \bar{K}2^{-m\gamma}\right)$$
$$\leq \sum_{j=0}^{2^m-1} \mathbb{P}\left(\left|S_{\frac{j+1}{2^m}T} - S_{\frac{j}{2^m}T}\right| \geq \bar{K}2^{-m\gamma}\sqrt{T}\right) \leq C\frac{2^{(4\gamma-1)m}}{\bar{K}^4}.$$

Summing over $m \geq 1$ and then choosing $\bar{K} = \bar{K}(\varepsilon)$, we obtain that (5.2) holds for all dyadic intervals with probability $1 - \varepsilon$. The tightness follows. □

Remark 5.3. The condition that $\gamma < \frac{1}{4}$ was imposed since we use the fourth moment in the argument above. Considering higher moments one can make γ arbitrarily close to 1/2 which is an optimal result. Namely, using the estimate $\mathbb{E}(\Delta^{2p}) \leq CL^p$ one obtains the bound

$$\mathbb{P}\left(\exists j < 2^m \quad \left|W_T\left(\frac{j+1}{2^m}\right) - W_T\left(\frac{j}{2^m}\right)\right| \geq \bar{K}2^{-m\gamma}\right) \leq \frac{2^{m(1+p(2\gamma-1))}}{\bar{K}^{2p}}$$

which decays exponentially in m if p is sufficiently large and $\gamma < 1/2$. However, taking $\gamma < 1/4$ is sufficient for proving tightness.

5.3. **Brownian sheet.** Recall that the Brownian sheet is a Gaussian process $\{W(u_1, u_2, \ldots, u_d)\}_{u_j \in \mathbb{R}^+}$ such that $\mathbb{E}(W(u_1, u_2, \ldots, u_d)) = 0$ and

$$\text{Cov}(W(u_1', u_2', \ldots, u_d'), W(u_1'', u_2'', \ldots, u_d'')) = \sigma^2 \prod_{j=1}^{d} \min(u_j', u_j'').$$

Let $W_T(s_1, s_2, \ldots s_d) = \dfrac{1}{T^{d/2}} \displaystyle\int_{u \in \prod_j [0, s_j T]} A(G_u y) du.$

Theorem 5.4. W_T converges weakly as $T \to \infty$ to a Brownian sheet.

Proof. The proof of this result is similar to the proof of Theorem 5.2 so we just sketch the argument.

(I) Similar to step I in the proof of Theorem 5.2 we obtain that $W_T(s_1, s_2, \ldots s_d)$ is asymptotically normal with zero mean and variance $T^{-d}\text{Vol}\left(\prod_{j=1}^{d}[0, s_j T]\right)\sigma^2 = \left(\prod_{j=1}^{d} s_j\right)\sigma^2$ where σ^2 is given by (5.1).

(II) Similarly to Step II in the proof of Theorem 5.2 we obtain that

$$\lim_{T\to\infty} \text{Cov}(W_T(u_1', \ldots, u_d')W_T(u_1'', \ldots u_d'')) = \prod_{j=1}^{d} \min(u_j', u_j'')\sigma^2.$$

(III) We say that $[(u_1', \ldots u_d'), (u_1'', \ldots u_d'')]$ is *a diadic segment* if there is $\ell \in \{1, \ldots d\}$ and $m \in \mathbb{N}$ such that $u_\ell' = \frac{j}{2^m}$, $u_\ell'' = \frac{j+1}{2^m}$, and for all $i \neq \ell$, $u_i' = u_i''$ and $2^m u_i'' \in \mathbb{N}$. Similarly to Remark 5.3, one can show that for each $\varepsilon > 0$ there exist $\bar{K}$ and $\gamma > 0$ such that with probability $1 - \varepsilon$, for each diadic segment

$$|W_T(u_1'', \ldots u_d'') - W_T(u_1', \ldots u_d')| \leq \bar{K}|u_\ell'' - u_\ell'|^\gamma.$$

Joining every two points by a union of diadic segments we see that with probability $1-\varepsilon$ it holds that for each $(u_1', \dots u_d'), (u_1'', \dots u_d'') \in [0,1]^d$ we have

$$|W_T(u_1'', \dots u_d'') - W_T(u_1', \dots u_d')| \le K \sum_{j=1}^{d} |u_j'' - u_j'|^\gamma.$$

This shows that the family $W_T(\cdot)$ is tight and completes the proof of the theorem. □

5.4. **Limiting Gaussian field.**

Lemma 5.5. Let $\mathfrak{C}_T = [0,T]^d$ and $\phi : \mathfrak{C}_1 \to \mathbb{R}$ be a continuous function. Denote $W_T^\phi = \displaystyle\int_{\mathfrak{C}_T} A(G_t(x))\phi(t/T)dt$ then $W_T^\phi / T^{d/2}$ converges as $T \to \infty$ to a normal random variable with zero mean and variance $\sigma^2 \|\phi\|^2_{L^2}$ where σ^2 is given by (5.1).

Proof. We verify the conditions of Theorem 2.3 with $\mathfrak{m}_T = T^{-d/2} \times$ the Lebesgue measure on $\mathfrak{C}_T$ and $A_t = \phi(t)A$. Condition (b) follows since $\mathfrak{m}(B(t, R \ln T)) \le C(R) \frac{\ln^d T}{T^{d/2}}$. Condition (c) follows since

$$V_T = \frac{1}{T^d} \int_{t_1, t_2 \in \mathfrak{C}_T} \mu(A(G_{t_1}x)A(G_{t_2}x))\phi(t_1/T)\phi(t_2/T)dt_1 dt_2$$

$$= \frac{1}{T^d} \int_{\substack{t_1, t_2 \in \mathfrak{C}_t \\ |t_1 - t_2| \le \ln^2 T}} \mu(A(G_{t_1}x)A(G_{t_2}x))\phi(t_1/T)\phi(t_2/T)dt_1 dt_2 + o(1)$$

$$= \frac{1}{T^d} \int_{\substack{t_1, t_2 \in \mathfrak{C}_t \\ |t_1 - t_2| \le \ln^2 T}} \mu(A(x)A(G_{t_2 - t_1}x))\phi^2(t_1/T)dt_1 dt_2 + o(1)$$

$$= \frac{1}{T^d} \int_{t \in \mathfrak{C}_T, s \in \mathbb{R}^d} \mu(A(x)A(G_s x))\phi^2(t/T)dt ds + o(1) = \sigma^2 \int_{\mathfrak{C}_1} \phi^2(t)dt + o(1)$$

where the second and the fourth line follow by exponential mixing and the third line follows by the uniform continuity of ϕ. □

Corollary 5.6. The family $\{W_T^\phi\}_{\phi \in C^{2d}}$ converges as $T \to \infty$ to a Gaussian random field $\mathcal{W}^\phi$ on $C^{2d}(\mathfrak{C}_1)$ with zero mean and covariance $\mathrm{Cov}(\mathcal{W}^\phi, \mathcal{W}^\psi) = \langle \phi, \psi \rangle_{L^2}$.

Proof. The convergence of finite dimensional distributions holds because by Lemma 5.5 for any $a_1, \dots, a_k,\ \phi_1, \dots \phi_k\ W_T^{\sum_j a_j \phi_j}$ is asymptotically normal with zero mean and variance $\left\| \displaystyle\sum_j a_j \phi_j \right\|^2_{L^2}$. To prove tightness we will show that for each ε there is a finite dimensional space V such that with probability $1-\varepsilon$ for each $\phi \in C^{2d}(\mathfrak{C}_1)$ we have $\left| W_T^\phi - W_T^{\pi_V(\phi)} \right| \le \varepsilon \|\phi\|_{C^{2d}}$. For $n \in \mathbb{Z}^d$ denote $e_n(t) = \exp\left(2\pi i \displaystyle\sum_{j=1}^d n_j t_j\right)$. Using the same argument as in Remark 5.3 one can show that for each $\kappa > 0$ there is N_0 such that with probability $1-\varepsilon$ if n satisfies $\max_j |n_j| \ge N_0$ then $|W_T^{e_n}| \le n^\kappa$. Take $\kappa < 1$ and $V = \mathrm{span}(e_n : \displaystyle\max_j |n_j| \le N_0)$. Let $\psi = \phi - \pi_V \phi$. Then $\psi = \displaystyle\sum_{n \in \mathcal{I}} a_n e_n(t)$ where

$\mathcal{I}$ includes the terms satisfying $\max_j |n_j| > N_0$. Note that $\|a_n\| \leq C\Pi^{-2}(n)\|\psi\|$ where $\Pi(n) := \prod_j (|n_j| + 1)$. It follows that with probability $1 - \varepsilon$

$$|W_T^\psi| = \left| \sum_{n \in \mathcal{I}} a_n W^{e_n} \right| \leq C \sum_{n \in \mathcal{I}} \Pi^{\kappa-2}(n).$$

The last sum is smaller than ε if N_0 is large enough. □

6. Applications.

We say that a function $w : \mathbb{Z}^d \to \mathbb{R}$ is *selfaveraging* if the limit $\bar{w} = \frac{1}{N^d} \sum_{\mathfrak{C}_N} w(n)$ exists where $\mathfrak{C}_N$ denotes the cube $[0, N]^d$.

Theorem 6.1. Let $w : \mathbb{Z}^d \to \mathbb{R}$ be a bounded function such that for each m the map $\mathfrak{w}_m(n) = w(n)w(n+m)$ is selfaveraging. Let $S_N = \sum_{n \in \mathfrak{C}_N} w(n)A(G_n y)$. Then $\frac{S_N}{N^{d/2}}$ converges as $N \to \infty$ to normal distribution with zero mean and varaince

$$\sigma^2 = \sum_{m \in \mathbb{Z}^d} \overline{\mathfrak{w}_m}\, \rho(m), \tag{6.1}$$

where $\rho(m) = \mu(A(y)A(G_m y))$.

Proof. We define $\mathfrak{m}_N = \frac{1}{N^{d/2}} \sum_{n_j \in [0,N]} \delta_n$. Then property (a) is evident, (b) is shown at the first step of the proof of Theorem 5.2.

It remains to prove property (c). We have

$$\frac{V(S_N)}{N^d} = \frac{1}{N^d} \sum_{n \in \mathfrak{C}_N} \sum_{m \in \mathbb{Z}^d} \mu(A(y)A(G_m y))\mathfrak{w}_m(n) 1_{n+m \in \mathfrak{C}_N} d\mu(x) = \sum_{m \in \mathbb{Z}^d} \rho(m)h(m, N)$$

where $h(m, N) = \frac{1}{N^d} \sum_{n \in \mathfrak{C}_N} \mathfrak{w}_m(N) 1_{n+m \in \mathfrak{C}_N}$. Note that for each m, $h(m, N) \leq \|w\|_\infty^2$ and moreover $\lim_{N \to \infty} h(m, N) = \overline{\mathfrak{w}_m}$ because $\mathfrak{w}_m$ are selfaveraging. Thus (6.1) follows by Dominated Convergence Theorem. □

Corollary 6.2. Let $f : Y \to Y$ be exponentially mixing and $\boldsymbol{\mu}^2$ be the square of the Mobius function, that is $\boldsymbol{\mu}^2$ is the indicator function of square free numbers. Then for each smooth zero mean A, $\frac{1}{\sqrt{N}} \sum_{n=1}^{N} \boldsymbol{\mu}^2(n) A(f^n y)$ converges to a normal distribution.

The corollary follows since $n \mapsto \boldsymbol{\mu}^2(n)\boldsymbol{\mu}^2(n+m)$ is self averaging by [4].

Remark 6.3. The above results would also hold for the Möbius function $\boldsymbol{\mu}$ provided it is self averaging. However this property is precisely Chowla's conjecture [3] for $\boldsymbol{\mu}$ for correlations of order 2 which is widely open. One can also consider *logarithmic* averages, i.e. expressions of the form $S_N = \sum_{n\in[0,N]} \frac{\boldsymbol{\mu}(n)A(G_n y)}{\sqrt{n}}$. The above reasoning can be used to show that $\frac{S_N}{\sqrt{\log N}}$ converges to the normal distribution with zero mean and variance $\sigma^2 = \tilde{\boldsymbol{\mu}} \cdot \mu(A^2)$, where $\tilde{\boldsymbol{\mu}} = \lim_{N\to\infty} \frac{1}{\log N} \sum_{n\in[0,N]} \frac{\boldsymbol{\mu}^2(n)}{n}$. Here we use the fact that the Möbius functions is logarithmically self averaging, i.e. for every $m \in \mathbb{Z} \setminus \{0\}$, $\frac{1}{\log N} \sum_{n\in[0,N]} \frac{\boldsymbol{\mu}(n)\boldsymbol{\mu}(n+m)}{n} \to 0$, as $N \to \infty$, which follows from [5], Theorem 1.3.

Lemma 6.4. Let $f : Y \to Y$ be exponentially mixing and A be a smooth zero mean function. Let $S_N = \sum_{p\leq N} A(f^p y)$ where the sum is over primes. Then $\frac{S_N}{\sqrt{N/\ln N}}$ converges as $N \to \infty$ to a normal random variable with zero mean and variance $\nu(A^2)$.

Proof. Let $\mathfrak{m}_N = \sqrt{\frac{\ln N}{N}} \sum_{p\leq N} \delta_p$. Then (a) is evident. (b) follows because

$$\mathfrak{m}_N(B(n, K \ln \|m_N\|)) \leq \mathfrak{m}_N(B(n, K \ln N)) \leq \frac{CK \ln^{3/2} N}{\sqrt{N}}.$$

To prove (c) note that $V_N = \sum_{m\in\mathbb{Z}} \frac{\ln N}{N} Q(m,N)\rho(m)$ where $Q(m,N)$ is the number of primes p less that $N-m$ such that $p+m$ is a prime. Observe that m-th term in this sum is $O\left(\theta^{|m|}\right)$. Therefore by Dominated Convergence Theorem it suffices to compute the limit for each term. Now for $m = 0$, $\rho(0) = \mu(A^2)$ while $\lim_{N\to\infty} \frac{\ln N\, Q(0,N)}{N} = 1$ by the Prime Number Theorem, while for $m \neq 0$ $\lim_{N\to\infty} \frac{\ln N\, Q(m,N)}{N} = 0$ by [2]. □

References

[1] Björklund M., Gorodnik A. *Central Limit Theorems for group actions which are exponentially mixing of all orders,* J. d'Analyse Mathematiques **141** (2020) 457–482.

[2] Chen J.R. *On Goldbach's problem and the sieve methods*, Sci. Sinica Ser. A **21** (1978), 701–738.

[3] Chowla S. *The Riemann hypothesis and Hilberts tenth problem*, Mathematics and Its Applications, Vol. 4, Gordon and Breach Science Publishers, New York, 1965.

[4] Mirsky L. *Arithmetical pattern problems relating to divisibility by r-th powers,* Proc. London Math. Soc. **50** (1949) 497–508.

[5] Tao T. *The logarithmically averaged Chowla and Elliot conjectures for two-point correlations*, Forum Math. Pi **4** (2016) paper e8, 36 pp.