Equidistribution of random walks on compact groups II. The Wasserstein metric

BENCE BORDA

\textsuperscript{1}Graz University of Technology, Steyrergasse 30, 8010 Graz, Austria
E-mail: borda@math.tugraz.at

We consider a random walk $S_k$ with i.i.d. steps on a compact group equipped with a bi-invariant metric. We prove quantitative ergodic theorems for the sum $\sum_{k=1}^{N} f(S_k)$ with Hölder continuous test functions $f$, including the central limit theorem, the law of the iterated logarithm and an almost sure approximation by a Wiener process, provided the distribution of $S_k$ converges to the Haar measure in the $p$-Wasserstein metric fast enough. As an example we construct discrete random walks on an irrational lattice on the torus $\mathbb{R}^d/\mathbb{Z}^d$, and find their precise rate of convergence to uniformity in the $p$-Wasserstein metric. The proof uses a new Berry–Esseen type inequality for the $p$-Wasserstein metric on the torus, and the simultaneous Diophantine approximation properties of the lattice. These results complement the first part of this paper on random walks with an absolutely continuous component and quantitative ergodic theorems for Borel measurable test functions.

Keywords: ergodic theorem, empirical distribution, central limit theorem, law of the iterated logarithm, Berry–Esseen inequality, simultaneous Diophantine approximation.

1. Introduction

Consider a compact Hausdorff group $G$ with normalized Haar measure $\mu$, and a random walk $S_k = \prod_{j=1}^{k} X_j = X_1 X_2 \cdots X_k$ on $G$, where $X_1, X_2, \ldots$ are independent, identically distributed (i.i.d.) $G$-valued random variables. Assuming that the distribution of $X_1$ is a regular Borel probability measure $\nu$, the distribution of $S_k$ is $\nu^k$, the $k$-fold convolution of $\nu$. We say that $\nu$ is adapted if its support is not contained in any proper closed subgroup of $G$; further, $\nu$ is strictly aperiodic if its support is not contained in a coset of any proper closed normal subgroup of $G$.

The study of such random walks is a classical topic; in the first part of this paper \cite{10} we gave a brief overview of the early history. The most fundamental result is

**Theorem A** (Stromberg \cite{30}). Let $\nu$ be a regular Borel probability measure on a compact Hausdorff group $G$. Then $\nu^k$ converges weakly to the Haar measure $\mu$ as $k \to \infty$ if and only if $\nu$ is adapted and strictly aperiodic.

Throughout the paper we shall work under the technical assumption that $G$ is second countable. This will ensure that $G$ is metrizable; in fact, every second countable compact Hausdorff group is metrizable by a bi-invariant metric $d$, that is, by a metric $d$ satisfying
Let $G$ be a second countable compact Hausdorff group. Let $X_1, X_2, \ldots$ be i.i.d. $G$-valued random variables, and set $S_k = \prod_{j=1}^k X_j$. Then
\begin{equation}
\frac{1}{N} \sum_{k=1}^N f(S_k) \to \int_G f \, d\mu \quad \text{a.s.}
\end{equation}
for all continuous functions $f : G \to \mathbb{R}$ if and only if the distribution of $X_1$ is adapted.

Here a.s. (almost surely) means that the given relation holds with probability 1. Theorem B is an ergodic theorem; more precisely, it gives a necessary and sufficient condition for the random sequence $S_k$, $k = 1, 2, \ldots$ to be equidistributed in $G$ with probability 1. Our main goal is to prove quantitative ergodic theorems for the random walk $S_k$. In the terminology of probability theory, this means that we aim to improve the strong law of large numbers (CLT) or the law of the iterated logarithm (LIL). To prove a quantitative version of Theorem B, we will need stronger, more quantitative assumptions on the distribution of $X_1$ and on the test function $f$. Recall that the $p$-Wasserstein metric $W_p$ metrizes weak convergence of Borel probability measures on $G$ for any $p > 0$; in particular by Theorem A, $W_p(\nu^{*k}, \mu) \to 0$ as $k \to \infty$ if and only if $\nu$ is adapted and strictly aperiodic. In our quantitative form of Theorem B instead of the adaptedness of $\nu$, we shall assume that $W_p(\nu^{*k}, \mu) \to 0$ fast enough. Regarding the test function $f$, instead of continuity, we shall assume that it is $p$-Hölder for some $0 < p \leq 1$; recall that a function $f : G \to \mathbb{R}$ is called $p$-Hölder if there exists a constant $K \geq 0$ such that $|f(x) - f(y)| \leq Kd(x,y)^p$ for all $x, y \in G$, where $d$ is the metric on $G$. Our quantitative ergodic theorem is

**Theorem 1.** Let $G$ be a second countable compact Hausdorff group with normalized Haar measure $\mu$, and fix a bi-invariant metric on $G$ which metrizes its topology. Consider a sequence $X_1, X_2, \ldots$ of i.i.d. $G$-valued random variables with distribution $\nu$, and set $S_k = \prod_{j=1}^k X_j$. Let $0 < p \leq 1$, and assume that $\sum_{k=1}^\infty W_p(\nu^{*k}, \mu) < \infty$. Then for any $p$-Hölder function $f : G \to \mathbb{R}$ such that $\int_G f \, d\mu = 0$, the sum $\sum_{k=1}^N f(S_k)$ satisfies the central limit theorem
\begin{equation}
\sum_{k=1}^N f(S_k) \sqrt{\frac{d}{N}} \to \mathcal{N}(0, \sigma^2)
\end{equation}
Equidistribution of random walks on compact groups II.

and the law of the iterated logarithm

$$\limsup_{N \to \infty} \frac{\sum_{k=1}^{N} f(S_k)}{\sqrt{2N \log \log N}} = \sigma$$

with some constant $0 \leq \sigma < \infty$ depending only on $f$ and $\nu$.

Relation (2) means convergence in distribution to the normal distribution with mean zero and variance $\sigma^2$ (or to the constant 0 in case $\sigma = 0$). We emphasize that the $p$-Wasserstein metric and $p$-Hölder functions are defined with respect to the same bi-invariant metric. We conjecture that the condition $\sum_{k=1}^{\infty} W_p(\nu^{*k}, \mu) < \infty$ is best possible.

Theorem 1 is in fact a direct corollary of our main result, an almost sure approximation of the sum $\sum_{k=1}^{N} f(S_k)$ by a Wiener process, see Theorem 5.

In the first part of this paper [10] we considered the class of all Borel measurable test functions. As an analogue of Theorem B we proved that the strong law of large numbers (1) holds for all bounded Borel measurable functions $f$ if and only if the distribution of $X_1$ is adapted, and the distribution of $S_k$ has an absolutely continuous component for some $k \geq 1$. We also proved quantitative forms of equidistribution such as the CLT and the LIL, and showed that the assumption that $f$ is bounded can be weakened to delicate moment conditions. The main message of our first paper [10] is that random walks with an absolutely continuous component satisfy quantitative ergodic theorems for the much wider class of Borel measurable test functions. In the present paper we will thus focus on random walks with singular distributions; see Theorems 4 and 6 for examples of discrete distributions to which Theorem 1 applies.

Quantitative forms of Kakutani’s random ergodic theorem for ergodic group actions along random walks were given by Furman and Shalom [16]. Their results apply in particular to the action of certain subgroups of $\text{SL}_d(\mathbb{Z})$ on the torus $\mathbb{R}^d/\mathbb{Z}^d$. Such actions along random walks have more recently been studied by Bourgain, Furman, Lindenstrauss and Mozes [11] and by Boyer [12]. For random walks on $\text{GL}_d(\mathbb{R})$ see Cuny, Dedecker and Jan [14] and references therein. The case of Abelian groups is of course simpler; for instance, Theorem B for the circle group $G = \mathbb{R}/\mathbb{Z}$ dates back to Robbins in 1953 [24]. Berkes and Borda [6] proved quantitative ergodic theorems on the circle group for test functions of bounded variation. The same authors also showed that the LIL (3) with $f(x) = \exp(2\pi i x)$ on the circle group holds if and only if the distribution of $X_1$ is nondegenerate [5]. It is natural to ask whether this fact generalizes to arbitrary compact groups; for example, one might conjecture that the CLT (2) and the LIL (3) with some $0 \leq \sigma < \infty$ remain true for characters $f(x) = \text{tr} \pi(x)$, $\pi$ a nontrivial irreducible unitary representation of $G$, under the sole assumption that $\nu$ is adapted (without assuming fast enough convergence of $\nu^{*k}$ in some probability metric).

The proof of Theorem 1 (and our main result Theorem 5) is based on a perturbation method. Simply put, in the long product $S_k = X_1 X_2 \cdots X_k$ we will replace the product $S_J := \prod_{j \in J} X_j$ over a short block $J \subset [1, k]$ by a uniformly distributed variable $U_J$. Surprisingly, this perturbation will introduce independence, and limit theorems such as the CLT and the LIL follow from classical methods of probability theory. In order for the
B. Borda

To estimate the rate of convergence in the $p$-Wasserstein metric, we will use a Berry–Esseen type inequality for $W_p$. The corresponding result for the uniform metric on $\mathbb{R}^d/\mathbb{Z}^d$
is due to Niederreiter and Philipp [22]. The original Berry–Esseen inequality concerns the uniform metric on $\mathbb{R}$ [23, p. 142]; a generalization for the uniform metric on $\mathbb{R}^d$ is due to von Bahr [1]. Bobkov [9] gives a survey of similar smoothing inequalities for several other probability metrics on $\mathbb{R}$.

**Proposition 3** (Berry–Esseen inequality for $W_p$ on the torus). Let $\nu_1$ and $\nu_2$ be Borel probability measures on $\mathbb{R}^d/\mathbb{Z}^d$ with the Euclidean metric.

(i) If $0 < p < 1$, then for any integer $H > 1$,
\[
W_p(\nu_1, \nu_2) \leq \frac{5d}{1-p} \cdot \frac{1}{H^p} + \sum_{h \in \mathbb{Z}^d \setminus \{0\}} \frac{|\hat{\nu}_1(h) - \hat{\nu}_2(h)|}{\|h\|_\infty^p}.
\]

(ii) For any integer $H > 1$,
\[
W_1(\nu_1, \nu_2) \leq \frac{d}{2} \cdot \frac{2 + \log H}{H} + \sum_{h \in \mathbb{Z}^d \setminus \{0\}} \frac{|\hat{\nu}_1(h) - \hat{\nu}_2(h)|}{\|h\|_\infty}.
\]

The proof is based on smoothing with the $d$-dimensional Fejér kernel as an approximate identity. We will derive these estimates in somewhat stronger forms, see equations (18) and (19) in the proof. In the case $d = 1$, $p = 1$ the estimate is not sharp: the factor $\log H$ can be removed. Indeed, in one dimension every Lipschitz function is of bounded variation, and so Koksma’s inequality [19, p. 143] combined with the Berry–Esseen type inequality of Niederreiter and Philipp for the uniform metric yields a better estimate for $W_1$. This method no longer works in higher dimensions; the reason is that for every $d \geq 2$ there exist Lipschitz functions on $\mathbb{R}^d/\mathbb{Z}^d$ which are not of bounded variation in the sense of Hardy and Krause (see [2, Section 3.1] for a construction). Consequently, we do not know whether the factor $\log H$ is necessary in the case $d \geq 2$, $p = 1$.

We emphasize that neither the statement nor the proof of Proposition 3 uses concepts such as axis parallel boxes or functions of bounded variation; the only ingredients are the metric on $\mathbb{R}^d/\mathbb{Z}^d$ and the unitary dual. Generalizing this approach, smoothing inequalities can be proved in other classical compact groups as well, yielding examples of random walks on noncommutative compact groups to which Theorem 1 applies. The details will appear elsewhere.

With Proposition 3 at our disposal, it is not difficult to find examples of random walks with fast convergence in the $p$-Wasserstein metric. We work out the details for certain random walks on rank $r$ lattices. We mention a simple explicit construction here; for more general results see Section 2.3. Let $r, d \geq 1$ be integers, and consider a monic irreducible polynomial with integer coefficients of degree $r + d$, with distinct real roots $a_1, a_2, \ldots, a_{r+d}$. In other words, $a_1, a_2, \ldots, a_{r+d}$ are conjugate real algebraic integers. Let $V$ be the $r \times r$ matrix with entries $V_{ij} = a_i^{-j-1}$, and let $W$ be the $r \times d$ matrix with entries $W_{ij} = a_i^{j-r-1}$. Note that $V$ is an invertible Vandermonde matrix. Consider the $r \times d$ matrix $M = V^{-1}W$. 

---

*Equidistribution of random walks on compact groups II.*
Theorem 4. Let $\alpha_1, \alpha_2, \ldots, \alpha_r \in \mathbb{R}^d$ denote the $r$ rows of the matrix $M$, and let $\nu$ be the Borel probability measure on $\mathbb{R}^d/\mathbb{Z}^d$ with atoms $\pm \alpha_i + \mathbb{Z}^d$, $1 \leq i \leq r$ of weight $1/(2r)$ each. For any $0 < p \leq 1$,

$$k^{-pr/(2d)} \ll W_p(\nu^k, \mu) \ll k^{-pr/(2d)} (\log k)^{|p|}$$

with implied constants depending only on the matrix $M$ and $p$. In particular, if $pr > 2d$, then $\sum_{k=1}^\infty W_p(\nu^k, \mu) < \infty$, and Theorem 1 applies.

The entries of $M$ are explicitly computable rational functions of the algebraic integers $a_1, a_2, \ldots, a_r$. The conjugates $a_{r+1}, a_{r+2}, \ldots, a_{r+d}$ only appear in the proof of the fact that the system of vectors $\alpha_1, \alpha_2, \ldots, \alpha_r$ is badly approximable; see Section 2.3. Since the exponent of $\log k$ is the integer part of $p$, a logarithmic factor only appears in the upper bound when $p = 1$. We mention that in the special case $d = 1, p = 1$ this logarithmic factor can be removed; see the comments made after Proposition 3. We do not know if the logarithmic factor in the upper bound is necessary in the case $d \geq 2, p = 1$.

Theorem 4 should be compared with the rate of convergence in the $p$-Wasserstein metric in the classical CLT setting. Given $p \geq 1$ and i.i.d. real-valued random variables $\zeta_1, \zeta_2, \ldots$ with mean zero, unit variance and $E|\zeta_1|^{p+2} < \infty$, the distance of $k^{-1/2} \sum_{j=1}^k \zeta_j$ from the standard normal distribution in $W_p$ is $O(k^{-1/2})$, and this is best possible. In the case $0 < p \leq 1$ an application of the Hölder inequality yields the general inequality $W_p(\nu_1, \nu_2) \leq W_1(\nu_1, \nu_2)^p$; in particular, for $0 < p \leq 1$ the rate of convergence in $W_p$ is $O(k^{-p/2})$. For the proof and more general results see Bobkov [7] and references therein.

The notation used in the paper and definitions are given in Section 2.1. The main result about random walks on general compact groups is stated in Section 2.2, and proved in Section 3. The main results about the rate of convergence of random walks on the torus are stated in Section 2.3; their proof, together with the proof of Propositions 2 and 3 are given in Section 4.

2. Main results

2.1. Definitions and notation

For the rest of the paper $G$ denotes a compact group whose topology is generated by a bi-invariant metric $d$, with normalized Haar measure $\mu$. We write $\nu_1 \ast \nu_2$ for the convolution of the Borel probability measures $\nu_1$ and $\nu_2$ on $G$, and $\nu^k$ for the $k$-fold convolution of $\nu$. Recall that $\nu \ast \mu = \mu \ast \nu = \mu$ for any $\nu$.

Let $\mathcal{F}_p$ ($0 < p \leq 1$) denote the set of all $p$-Hölder functions $f : G \to \mathbb{R}$ such that $\int_G f \, d\mu = 0$. The optimal $p$-Hölder constant of $f$ will be denoted by

$$\|f\|_{p-\text{Höld}} := \sup_{x \neq y \in G} \frac{|f(x) - f(y)|}{d(x, y)^p}.$$
Observe that \( \| \cdot \|_{p-\text{Hold}} \) is a norm on \( \mathcal{F}_p \). We simply write \( \| f \|_q \) for the \( L^q(G, \mu) \)-norm of \( f \) (\( 1 \leq q \leq \infty \)).

The \( p \)-Wasserstein distance \( 0 < p \leq 1 \) of two Borel probability measures \( \nu_1 \) and \( \nu_2 \) on \( G \) is defined as\(^1\)

\[
W_p(\nu_1, \nu_2) := \inf_{\vartheta \in \text{Coup}(\nu_1, \nu_2)} \int_{G \times G} d(x, y)^p \, d\vartheta(x, y),
\]

where \( \text{Coup}(\nu_1, \nu_2) \) denotes the set of couplings of \( \nu_1 \) and \( \nu_2 \), i.e. the set of Borel probability measures \( \vartheta \) on \( G \times G \) with marginals \( \vartheta(B \times G) = \nu_1(B) \) and \( \vartheta(G \times B) = \nu_2(B) \) \((B \subseteq G \) Borel\)). By the Kantorovich duality theorem,

\[
W_p(\nu_1, \nu_2) = \sup_{f \in \mathcal{F}_p} \left| \int_G f \, d\nu_1 - \int_G f \, d\nu_2 \right|. 
\]

This duality is usually stated for \( p = 1 \), i.e. for Lipschitz functions; to see the general case simply note that \( d(x, y)^p \) is also a metric on \( G \) generating the same topology, and apply Kantorovich duality for Lipschitz functions in the \( d(x, y)^p \) metric. Observe that for any Borel probability measures \( \nu_1, \nu_2 \) and \( \nu \) on \( G \) we have \( W_p(\nu_1 \ast \nu, \nu_2 \ast \nu) \leq W_p(\nu_1, \nu_2) \).

A \( G \)-valued random variable is a Borel measurable map \( X \) from a probability space to \( G \); the distribution of \( X \) is the Borel probability measure on \( G \) which assigns \( \Pr(X \in B) \) to a Borel set \( B \subseteq G \). A \( G \)-valued random variable is called uniformly distributed if its distribution is the Haar measure \( \mu \). If the \( G \)-valued random variables \( X \) and \( Y \) are independent, then the distribution of \( XY \) is the convolution of the distributions of \( X \) and \( Y \).

Throughout \( | \cdot | \) denotes the Euclidean norm of a vector in \( \mathbb{R}^d \) or the cardinality of a set. We use \( \langle x, y \rangle = \sum_{i=1}^d x_i y_i \) for the usual scalar product on \( \mathbb{R}^d \). The \( d \)-dimensional torus \( \mathbb{R}^d / \mathbb{Z}^d \) is identified by \([0, 1)^d \) or \([-1/2, 1/2)^d \) the usual way, as are functions on \( \mathbb{R}^d / \mathbb{Z}^d \) by \( \mathbb{Z}^d \)-periodic functions on \( \mathbb{R}^d \). The normalized Haar measure \( \mu \) becomes the Lebesgue measure on a unit cube via this identification; we simply write \( dx \) in integrals with respect to the Lebesgue measure. Given \( x, y \in \mathbb{R}^d \), we define the Euclidean distance from the point \( x + \mathbb{Z}^d \) to the point \( y + \mathbb{Z}^d \) on the torus as \( \| x - y \|_{\mathbb{R}^d / \mathbb{Z}^d} := \min_{m \in \mathbb{Z}^d} | x - y - m | \).

In particular, \( \| \cdot \|_{\mathbb{R}^d / \mathbb{Z}^d} \) is the distance from the nearest integer function. We use the maximum norm \( \| h \|_{\infty} = \max_{1 \leq j \leq d} | h_j | \) for lattice points \( h \in \mathbb{Z}^d \). The Fourier coefficients of a function \( f : \mathbb{R}^d / \mathbb{Z}^d \to \mathbb{R} \) are denoted by \( \hat{f}(h) = \int_{[0, 1)^d} e^{-2\pi i \langle h, x \rangle} f(x) \, dx, \) \( h \in \mathbb{Z}^d \); those of a Borel probability measure \( \nu \) by \( \hat{\nu}(h) = \int_{[0, 1)^d} e^{-2\pi i \langle h, x \rangle} \, d\nu(x), \) \( h \in \mathbb{Z}^d \).

Given a (deterministic) function \( E(t), t \in [0, \infty) \), we define two stochastic processes \( Y(t) \) and \( Z(t) \) in the Skorokhod space \( D[0, \infty) \) to be \( o(E(t)) \)-equivalent the same way as in \([10]\). Somewhat informally, this notion simply means that \( Y(t) = Z(t) + o(E(t)) \) as \( t \to \infty \) a.s., except we allow a process to be replaced by another process defined on a new probability space with the same distribution in \( D[0, \infty) \).

\(^1\)The exponent \( 1/p \) is missing to ensure that \( W_p \) is a metric, i.e. it satisfies the triangle inequality.
For general facts about topological groups we refer to Hewitt and Ross [18]. The $p$-Wasserstein metric and Kantorovich duality are related to the theory of optimal transportation, see Villani [31]. Simultaneous Diophantine approximation is covered in detail in Cassels [13], Schmidt [27] and Sprindzuk [28].

2.2. $p$-Hölder test functions on a compact group

Our main result is an almost sure approximation of the sum $\sum_{k=1}^{N} f(S_k)$ by a Wiener process. For any $f \in F_p$ let

$$C(f, \nu) = \mathbb{E}f(U)^2 + 2 \sum_{k=1}^{\infty} \mathbb{E}f(U)f(U S_k),$$

(5)

where $U$ is a uniformly distributed $G$-valued random variable, independent of $X_1, X_2, \ldots$.

Theorem 5. Assume that the conditions of Theorem 1 hold, and let $f \in F_p$. Then the series in (5) is absolutely convergent, and $C(f, \nu) \geq 0$.

(i) If $C(f, \nu) > 0$, then there exists a (deterministic) nondecreasing positive function $\sigma(t), t \in [0, \infty)$ with $\sigma(t)^2 = C(f, \nu)t + o(t)$ such that the processes $\sum_{k \leq t} f(S_k)$ and $W(\sigma(t)^2)$ are $o(t^{5/12+\varepsilon})$-equivalent in the Skorokhod space $D[0, \infty)$ for any $\varepsilon > 0$, where $W(t), t \in [0, \infty)$ is a standard Wiener process.

(ii) If $C(f, \nu) = 0$, then

$$\frac{\sum_{k \leq t} f(S_k)}{\sqrt{t}} \overset{d}{\to} 0 \quad \text{and} \quad \frac{\sum_{k \leq t} f(S_k)}{\sqrt{t \log \log t}} \overset{a.s.}{\to} 0$$

Remark. As the proof will clearly show, the error term in $\sigma(t)^2$ is related to the tails of the series $\sum_{k=1}^{\infty} W_p(\nu^{*k}, \mu)$. For instance, if $W_p(\nu^{*k}, \mu) \ll k^{-(1+\varepsilon)}$ for some $0 < \varepsilon \leq 1/2$, then $\sigma(t)^2 = C(f, \nu)t + O(t^{1-c/3})$. In this case $\sum_{k \leq t} f(S_k)$ is $o(t^{1/2-c/6+\varepsilon})$-equivalent to $\sqrt{C(f, \nu)}W(t)$ for any $\varepsilon > 0$.

It was not our intention to find the optimal error term in Theorem 5; indeed, it is likely that the exponent $5/12$ can be improved. The point is that $5/12 < 1/2$, and so Theorem 1 with $\sigma = \sqrt{C(f, \nu)}$ follows immediately from Theorem 5 and classical limit theorems for Wiener processes. Furthermore, the almost sure asymptotics and the limit distribution of continuous functionals of the process $\sum_{k \leq t} f(S_k)$ also follow. Of course we could also use the piecewise linear functions $\sum_{k \leq |t|} f(S_k) + (|t| - t)f(S_{|t|+1})$ instead of the step functions $\sum_{k \leq t} f(S_k)$; in that case the same results hold in the space of continuous functions $C_{\infty}[0, \infty]$.

The normalizing factor $C(f, \nu)$ defined in (5) is the same as in the first part of this paper [10]. For a brief comparison with general Markov chain theory we refer to the same paper; in particular, note that the same factor $C(f, \nu)$ appears in the variance in the classical CLT and LIL for Markov chains.
2.3. Rate of convergence in the $p$-Wasserstein metric

Let $d \geq 1$ and consider the additive group of the $d$-dimensional torus $\mathbb{R}^d/\mathbb{Z}^d$ with the Euclidean metric and normalized Haar measure $\mu$. Let $r \geq 1$ and $\alpha_1, \alpha_2, \ldots, \alpha_r \in \mathbb{R}^d$. Further, let $\xi_1, \xi_2, \ldots, \xi_r$ be $\mathbb{Z}$-valued random variables, and let $I$ be a $\{1, 2, \ldots, r\}$-valued random variable with $\xi_1, \xi_2, \ldots, \xi_r, I$ independent. Consider the i.i.d. sequence $X_1, X_2, \ldots$ of $\mathbb{R}^d/\mathbb{Z}^d$-valued random variables where $X_1 = \xi_I \alpha_I + \mathbb{Z}^d$; that is, the distribution of $X_1$ is $\nu = \sum_{i=1}^{r} \sum_{n \in \mathbb{Z}} \Pr(I = i) \Pr(\xi_i = n) \delta_{n \alpha_i + \mathbb{Z}^d}$, where $\delta_{a + \mathbb{Z}^d}$ is the Dirac measure concentrated on the point $a + \mathbb{Z}^d$ on the torus. We thus have a random walk on the rank $r$ lattice $\left\{ \sum_{i=1}^{r} n_i \alpha_i + \mathbb{Z}^d : n_1, n_2, \ldots, n_r \in \mathbb{Z} \right\} \subset \mathbb{R}^d/\mathbb{Z}^d$, and to generate an elementary step of the walk we first randomly choose one of the basis vectors $\alpha_i$ of the lattice, and then move forward by a random integral multiple of $\alpha_i$.

The random walk in Theorem 4 is a special case, when $I$ is uniformly distributed on $\{1, 2, \ldots, r\}$, and $\xi_1, \xi_2, \ldots, \xi_r$ are Bernoulli random variables taking the values $\pm 1$ with probability $1/2$ each. It is not surprising that the rate of convergence of this random walk to uniformity is sensitive to the Diophantine properties of the given lattice. We will work under the assumption

$$\inf_{h \in \mathbb{Z}^d \setminus \{0\}} \left( \max_{1 \leq i \leq r} \|\langle h, \alpha_i \rangle\|_{\mathbb{R}/\mathbb{Z}} \cdot \psi(\|h\|_{\infty}) \right) > 0,$$

where $\psi(x)$ is a nondecreasing positive function on $[1, \infty)$. Dirichlet’s theorem [27, p. 28] states that (6) cannot hold with a function $\psi(x) = o(x^{d/r})$. A system of vectors $\alpha_1, \alpha_2, \ldots, \alpha_r$ is called badly approximable if (6) is satisfied with $\psi(x) = x^{d/r}$. It is known that badly approximable systems of vectors exist for any $d, r \geq 1$; in fact, the system used in Theorem 4 is badly approximable [27, p. 43–45]. A system of vectors is called a Roth system if every coordinate of every vector is algebraic, and (6) holds with $\psi(x) = x^{d/r + \varepsilon}$ for any $\varepsilon > 0$. Schmidt [27, p. 157] gave a simple necessary and sufficient condition for being a Roth system in terms of the rank of certain linear forms on rational subspaces. For instance, in the special case $d = 1$ a system $\alpha_1, \alpha_2, \ldots, \alpha_r$ is a Roth system if and only if $\alpha_1, \alpha_2, \ldots, \alpha_r$ are algebraic and linearly independent over the rationals. The corresponding metric result is the Khintchine–Groshev theorem [28]. Assume $\psi(x) = x^{d/r} L(x)^{1/r}$ with a nondecreasing positive function $L(x)$. If the series $\sum_{k=1}^{\infty} 1/(kL(k))$ is convergent (resp. divergent), then (6) is satisfied by almost every (resp. almost no) system $(\alpha_1, \alpha_2, \ldots, \alpha_r) \in \mathbb{R}^d$ in the sense of the Lebesgue measure.
Theorem 6. Assume that $\xi_1, \xi_2, \ldots, \xi_r$ are nondegenerate, $\Pr(I = i) > 0$ for all $1 \leq i \leq r$, and (6) is satisfied with a positive function $\psi(x)$ such that $\psi(x)x^{-d/r}$ is nondecreasing. For any $0 < p \leq 1$,

$$W_p(\nu^*, \mu) \leq \left( \frac{\log \psi^{-1}(\sqrt{k})}{\psi^{-1}(\sqrt{k})} \right)^{|p|},$$

where $\psi^{-1}(y) = \sup\{x \geq 1 : \psi(x) \leq y\}$ is the generalized inverse function of $\psi$. The implied constant depends only on the distributions of $\xi_1, \xi_2, \ldots, \xi_r, I$, the value of the infimum in (6), $d$, $r$ and $p$.

Theorem 6 with $\psi(x) = x^{d/r}$ yields the estimate

$$W_p(\nu^*, \mu) \ll k^{-(p r/(2d))},$$

valid whenever the system of vectors $\alpha_1, \alpha_2, \ldots, \alpha_r$ is badly approximable; in particular, the upper bound in Theorem 4 follows. If (6) is satisfied with $\psi(x) = x^{d/r} + \epsilon$ for any $\epsilon > 0$, then Theorem 6 gives

$$W_p(\nu^*, \mu) \ll k^{-(p r/(2d)) + \epsilon},$$

for any $\epsilon > 0$. This estimate holds for any Roth system $\alpha_1, \alpha_2, \ldots, \alpha_r$, and also for almost every $(\alpha_1, \alpha_2, \ldots, \alpha_r) \in \mathbb{R}^d$.

The logarithmic factor in Theorem 6 can be removed in the case $d = 1$, $p = 1$, but we do not know if it is necessary in the case $d \geq 2$, $p = 1$; see the comments made after Proposition 3. We conjecture that the estimate is tight whenever $\xi_1, \xi_2, \ldots, \xi_r$ have finite variance, possibly under certain regularity assumptions on the distributions of $\xi_1, \xi_2, \ldots, \xi_r, I$ and up to this logarithmic factor. We were only able to prove this conjecture for badly approximable systems.

Theorem 7. Assume that $\Pr(I = i)\mathbb{E}\xi_i^2 \leq B$ for all $1 \leq i \leq r$ with some constant $B > 0$. If the system of vectors $\alpha_1, \alpha_2, \ldots, \alpha_r \in \mathbb{R}^d$ is badly approximable, then for any $0 < p \leq 1$,

$$W_p(\nu^*, \mu) \gg k^{-pr/(2d)}$$

with an implied constant depending only on $\alpha_1, \alpha_2, \ldots, \alpha_r, B$ and $p$.

Diophantine approximation has been used by several authors to estimate the rate of convergence of random walks. Hensley and Su [17] showed that if $d = 1$, the system $\alpha_1, \alpha_2, \ldots, \alpha_r$ is badly approximable and $\nu$ has atoms $\pm \alpha_i + \mathbb{Z}$, $1 \leq i \leq r$ of weight $1/(2r)$ each, then the rate of convergence in the uniform metric is

$$k^{-r/2} \ll \delta_{\text{unif}}(\nu^*, \mu) \ll k^{-r/2}.$$ 

Berkes and Borda [4] studied the case $d = 1$, $r = 1$, and proved that if $\xi_1$ has finite variance and satisfies certain regularity assumptions (e.g. if $\xi_1$ is a Bernoulli variable), then

$$k^{-1/(2\gamma)} \ll \delta_{\text{unif}}(\nu^*, \mu) \ll k^{-1/(2\gamma)},$$

where $\gamma$ represents the sum of the degrees of the denominators of the approximants of $\alpha_1$. This establishes the tightness when $d = 1$.
provided \( \alpha \) satisfies \((6)\) with \( \psi(x) = x^\gamma \) with some \( \gamma \geq 1 \) in a tight way; however, the lower estimate only holds for infinitely many \( k \) (not necessarily for all \( k \)). See the same paper for results about heavy-tailed \( \xi_1 \). Bobkov \[8\] considered random walks on the real line defined in terms of irrational numbers in a similar way, and proved tight estimates for their distance from the normal distribution in the uniform metric, as well as Edgeworth expansions.

3. Random walks on general compact groups

Throughout this section we assume that the conditions of Theorem 1 hold. Let us fix \( 0 < p \leq 1 \), and let \( \Delta_k = W_p(\nu^{*k}, \mu) \) denote the distance of \( S_k \) from the uniform distribution in the \( p \)-Wasserstein metric. Further, let \( U \) be a uniformly distributed \( G \)-valued random variable independent of \( X_1, X_2, \ldots \).

First of all note that \( \|f\|_\infty \ll \|f\|_{p - \text{Hölder}} \) for any \( f \in F_p \). Indeed, for any \( x \in G \) we have
\[
|f(x)| = \left| \int_G (f(x) - f(y)) \, d\mu(y) \right| \leq \|f\|_{p - \text{Hölder}} \int_G d(x, y)^p \, d\mu(y),
\]
where the last integral does not depend on \( x \) by the translation invariance of \( d \) and \( \mu \), and is finite by the compactness of \( G \). Observe also that
\[
W_p(\nu^{*(k+1)}, \mu) = W_p(\nu^{*k} * \nu, \mu) \leq W_p(\nu^{*k}, \mu),
\]
therefore the sequence \( \Delta_k \) is nonincreasing. The assumption \( \sum_{k=1}^{\infty} \Delta_k < \infty \) thus implies that \( k\Delta_k \to 0 \) as \( k \to \infty \). Our main tool is the fact that by Kantorovich duality, any \( p \)-Hölder function \( g : G \to \mathbb{R} \) satisfies
\[
|\mathbb{E}g(S_k) - \mathbb{E}g(U)| = \left| \int_G g \, d\nu^{*k} - \int_G g \, d\mu \right| \leq \|g\|_{p - \text{Hölder}} \Delta_k. \tag{7}
\]

3.1. The variance

We start by finding the precise asymptotics of the variance of \( \sum_{k=1}^{N} f(U S_k) \), and estimate the variance of \( \sum_{k=1}^{N} f(S_k) \). It would not be difficult to find the precise asymptotics of the latter sum as well; however, we will not need it later.

Lemma 1. For any \( f \in F_p \) the series in \((5)\) is absolutely convergent, and
\[
\mathbb{E} \left( \sum_{k=1}^{N} f(U S_k) \right)^2 = C(f, \nu) N + o(N) \quad \text{as } N \to \infty.
\]
In particular, \( C(f, \nu) \geq 0 \).
Proof. Let $A_k = \mathbb{E}f(U)f(US_k)$. Since $U$ is independent of $S_k$, we have

$$A_k = \int_G \int_G f(u)f(ux) \, d\mu(u) \, d\nu^k(x).$$

Letting $g(x) = \int_G f(u)f(ux) \, d\mu(u)$ denote the inner integral, we can thus write $A_k = \mathbb{E}g(S_k)$. Observe that $\int_G g \, d\mu = 0$, and $\|g\|_{p-\text{H{"o}l}} \leq \|f\|_1 \cdot \|f\|_{p-\text{H{"o}l}}$. Applying (7) we thus obtain $|A_k| = |\mathbb{E}g(S_k)| \leq \|f\|_1 \cdot \|f\|_{p-\text{H{"o}l}} \Delta_k$, and the absolute convergence of the series in (5) follows.

Next, let us expand the square in the claim:

$$\mathbb{E} \left( \sum_{k=1}^N f(US_k) \right)^2 = \sum_{k=1}^N \mathbb{E}f(US_k)^2 + 2 \sum_{1 \leq k < \ell \leq N} \mathbb{E}f(US_k)f(US_\ell). \tag{8}$$

Here $US_k$ is uniformly distributed, because $\mu \nu^k = \mu$. Let us write $US_\ell = US_k \prod_{j=k+1}^\ell X_j$, and observe that $\prod_{j=k+1}^\ell X_j$ has the same distribution as $S_{\ell-k}$, and is independent of $US_k$. Therefore $\mathbb{E}f(US_k)^2 = \mathbb{E}f(U)^2$ and $\mathbb{E}f(US_k)f(US_\ell) = \mathbb{E}f(U)f(US_{\ell-k})$; in particular, (8) simplifies as

$$\mathbb{E} \left( \sum_{k=1}^N f(US_k) \right)^2 = N \mathbb{E}f(U)^2 + 2 \sum_{1 \leq k < \ell \leq N} \mathbb{E}f(U)f(US_{\ell-k})$$

$$= N \mathbb{E}f(U)^2 + 2 \sum_{d=1}^{N-1} (N-d) \mathbb{E}f(U)f(US_d)$$

$$= C(f,\nu)N + \mathcal{O}\left( \sum_{d=1}^{N-1} d |A_d| + N \sum_{d=N}^\infty |A_d| \right).$$

As observed before, here $|A_d| \leq \|f\|_1 \cdot \|f\|_{p-\text{H{"o}l}} \Delta_d$. The error term in the previous estimate is thus $o(N)$. \hfill \Box

Lemma 2. For any $f \in \mathcal{F}_p$ and any integer $N \geq 1$,

$$\mathbb{E} \left( \sum_{k=1}^N f(S_k) \right)^2 \ll \|f\|_1 \cdot \|f\|_{p-\text{H{"o}l}} N + \|f\|_{p-\text{H{"o}l}}^2$$

with an implied constant depending only on $\nu$, $p$ and the metric $d$.

Proof. We may assume that $\|f\|_{p-\text{H{"o}l}} = 1$. Expanding the square we get

$$\mathbb{E} \left( \sum_{k=1}^N f(S_k) \right)^2 = \sum_{k=1}^N \mathbb{E}f(S_k)^2 + 2 \sum_{1 \leq k < \ell \leq N} \mathbb{E}f(S_k)f(S_\ell). \tag{9}$$
To estimate the first sum, note that \( g(x) = f(x)^2 \) satisfies
\[
\mathbb{E} g(U) = \| f \|_2^2 \leq \| f \|_1 \cdot \| f \|_\infty \ll \| f \|_1 \cdot \| f \|_{p-Höld} = \| f \|_1
\]
and
\[
\| g \|_{p-Höld} \leq 2\| f \|_\infty \cdot \| f \|_{p-Höld} \ll \| f \|_{p-Höld}^2 = 1.
\]
Hence by (7) we have
\[
\sum_{k=1}^N \mathbb{E} f(S_k)^2 \ll \sum_{k=1}^N (\| f \|_1 + \Delta_k) \ll \| f \|_1 N + 1.
\] (10)

Next, consider the second sum in (9). Since \( S_\ell = S_k \prod_{j=k+1}^\ell X_j \), we have
\[
\mathbb{E} f(S_k)f(S_\ell) = \int_G \int_G f(x) f(xy) \, d\nu^{*k}(x) d\nu^{*(\ell-k)}(y).
\]
Let \( g(y) = \int_G f(x) f(xy) \, d\nu^{*k}(x) \) denote the inner integral, and observe \( \mathbb{E} g(U) = \int_G g \, d\mu = 0 \). From (7) we thus deduce
\[
\mathbb{E} |f(S_k)f(S_\ell)| \leq \| g \|_{p-Höld} \Delta_{\ell-k}.
\]
For any \( y, y' \in G \) we have
\[
|g(y) - g(y')| = \left| \int_G f(x) (f(xy) - f(xy')) \, d\nu^{*k}(x) \right|
\]
\[
\leq \int_G |f(x)| \cdot \| f \|_{p-Höld} d(xy, xy')^p \, d\nu^{*k}(x)
\]
\[
= \mathbb{E} |f(S_k)| \cdot d(y, y')^p
\]
where we used the translation invariance of the metric \( d \). Applying (7) to \( h(x) = |f(x)| \) we deduce \( \mathbb{E} |f(S_k)| \leq \| f \|_1 + \| f \|_{p-Höld} \Delta_k = \| f \|_1 + \Delta_k \). Therefore \( \| g \|_{p-Höld} \leq \| f \|_1 + \Delta_k \), and altogether we get
\[
|\mathbb{E} f(S_k)f(S_\ell)| \leq (\| f \|_1 + \Delta_k) \Delta_{\ell-k}.
\]
Summing over \( 1 \leq k < \ell \leq N \) we finally obtain that the second sum in (9) satisfies
\[
\sum_{1 \leq k < \ell \leq N} \mathbb{E} f(S_k)f(S_\ell) \ll \| f \|_1 N + 1.
\]
This estimate together with (10) finishes the proof.

3.2. Approximation by independent random variables

The main goal of this section is to approximate \( \sum_{k=1}^N f(S_k) \) by a sum of independent random variables. We use the same exact construction as in [10]; for the sake of completeness, we repeat the key details. This construction in its simplest form was first used
by Schatte [25], [26] on the circle group, and is admittedly somewhat ad hoc; we give an
intuitive explanation first. We start by partitioning the set of positive integers into con-
secutive, nonempty, finite intervals of integers (called blocks for short) \(H_1, J_1, H_2, J_2, \ldots\).
Consider the block sums \(Y_i = \sum_{k \in J_i} f(S_k)\) and \(Z_i = \sum_{k \in H_i} f(S_k)\). In each block sum
we perform a small perturbation of \(S_k\); we replace the product corresponding to the
preceding block by a uniformly distributed random variable. Thus e.g. in each term of
\(Y_i = \sum_{k \in J_i} f(S_k)\) we replace
\[
S_k = \left( \prod_{\ell \in H_i \cup J_i \cup \cdots \cup H_{i-1} \cup J_{i-1}} X_\ell \right) \prod_{\ell \in H_i} X_\ell \prod_{\ell \leq k} X_\ell
\]
by
\[
S_k^* = \left( \prod_{\ell \in H_i \cup J_i \cup \cdots \cup H_{i-1} \cup J_{i-1}} X_\ell \right) U_i \prod_{\ell \leq k} X_\ell,
\]
where \(U_i\) is uniformly distributed. The reason why we use a uniformly distributed variable
is that this way the perturbed block sums \(Y_i^* = \sum_{k \in J_i} f(S_k^*)\) become independent.
Indeed, the uniform distribution has the peculiar ability to imply independence; this is
related to translation invariance, a defining property of the Haar measure. For the techni-
cal details we refer to [10, Lemmas 1 and 2]. To ensure that the error of this perturbation
is small, we will use a coupling to construct \(U_i\). If the size of \(H_i\) is large, then the
distribution of \(\prod_{\ell \in H_i} X_\ell\) is close to the Haar measure, and thus replacing it by a carefully
chosen uniformly distributed variable has a small effect; this makes the perturbation error
\(Y_i - Y_i^*\) small. Note that in the formal construction the sums \(Y_i, Y_i^*, Z_i, Z_i^*\) will be
defined in terms of variables \(W_k, W_k^*\) which only have the same distribution as \(S_k, S_k^*\)
but are not actually equal to them; however, this is just a technicality.

The perturbation method thus approximates \(\sum_{k=1}^N f(S_k)\) by two sums of independent
variables \(Y_1, Y_2, \ldots, Z_2, Z_3, \ldots\); the two sums, however, are not independent of one
another. To overcome this problem, we need to make sure that, say, the variables \(Z_i^*\)
are negligible compared to the variables \(Y_i^*\); this way \(\sum_{k=1}^N f(S_k)\) is approximated by a
single sum of independent variables, and limit theorems such as the CLT and the LIL
follow. Therefore our main strategy is to choose the sizes of the blocks \(H_1, J_1, H_2, J_2, \ldots\)
optimally: on the one hand, we need \(H_i\) to be large enough to make the perturbation error small; on the other hand, we need \(H_i\) to be small enough so that \(Z_i^*\) is negligible compared to \(Y_i^*\).

We now give the formal construction. By the definition of the \(p\)-Wasserstein metric,
for any \(k \geq 1\) there exists a coupling \(\vartheta_k\) of \(\nu^{*k}\) and the Haar measure \(\mu\) such that, say\(^2\),
\[
\int_{G \times G} d(x, y)^p \, d\vartheta_k(x, y) \leq 2 W_p(\nu^{*k}, \mu) = 2 \Delta_k.
\]
\(^2\)Since \(G\) is compact, the infimum in the definition of \(W_p\) is actually attained by some coupling; however, we do not need this fact.
For any nonempty, finite interval of positive integers \( J \subseteq \mathbb{N} \) let \( S_J = \prod_{j \in J} X_j \). After a suitable extension of the probability space, for any such \( J \subseteq \mathbb{N} \) we can introduce a pair of \( G \)-valued random variables \((T_J, U_J)\) with joint distribution \( \vartheta_{|J|} \), where \(|J|\) denotes the cardinality of \( J \). That is, \( T_J \overset{d}{=} S_J \), the variable \( U_J \) is uniformly distributed, and \( \mathbb{E} d(T_J, U_J)^p \leq 2\Delta_{|J|} \). Further, we may assume that \( X_1, X_2, \ldots \) and \((T_J, U_J), J \subseteq \mathbb{N} \) are independent.

Fix a block decomposition \( H_1, J_1, H_2, J_2, \ldots \) of the set of positive integers. For any \( i \geq 1 \) and \( k \in J_i \) let

\[
W_k = \prod_{j=1}^{i-1} (T_{H_j} S_{J_j}) T_{H_k} \prod_{\ell \leq k} X_\ell, \quad W_k^* = \prod_{j=1}^{i-1} (T_{H_j} S_{J_j}) U_{H_k} \prod_{\ell \leq k} X_\ell.
\]

Similarly, for any \( i \geq 2 \) and \( k \in H_i \) let

\[
W_k = \prod_{j=1}^{i-2} (S_{H_j} T_{J_{j-1}}) S_{H_{i-1}} T_{J_{i-1}} \prod_{\ell \leq k} X_\ell, \quad W_k^* = \prod_{j=1}^{i-2} (S_{H_j} T_{J_{j-1}}) S_{H_{i-1}} U_{J_{i-1}} \prod_{\ell \leq k} X_\ell.
\]

We use the convention that an empty product equals the identity element \( 1 \in G \). Note that \( W_k^* \) is obtained from \( W_k \) by replacing a block product by a uniformly distributed variable; consequently, \( W_k^* \) is uniformly distributed for all \( k \). Observe that the (joint) distribution of the sequence \( W_k \), \( k \in \bigcup_{i=1}^\infty J_i \) is the same as the (joint) distribution of the sequence \( S_k, k \in \bigcup_{i=2}^\infty J_i \); the same holds for the sequences \( W_k, k \in \bigcup_{i=2}^\infty H_i \) and \( S_k, k \in \bigcup_{i=2}^\infty H_i \). Finally, for a given \( f \in \mathcal{F}_p \) let us introduce the random variables

\[
Y_i = \sum_{k \in J_i} f(W_k), \quad Y_i^* = \sum_{k \in J_i} f(W_k^*) \quad (i \geq 1),
\]

\[
Z_i = \sum_{k \in H_i} f(W_k), \quad Z_i^* = \sum_{k \in H_i} f(W_k^*) \quad (i \geq 2).
\]

Observe that along the sequence \( N_R = \max J_R (R = 1, 2, \ldots) \) we have

\[
\sum_{k=1}^{N_R} f(S_k) = \sum_{k \in H_1} f(S_k) + \sum_{i=1}^{R} \sum_{k \in J_i} f(S_k) + \sum_{i=2}^{R} \sum_{k \in H_i} f(S_k),
\]

where the (joint) distribution of the sequence \( \sum_{i=1}^{R} \sum_{k \in J_i} f(S_k) \) (\( R = 1, 2, \ldots \)) is the same as the (joint) distribution of the sequence \( \sum_{i=1}^{R} Y_i \) (\( R = 1, 2, \ldots \)), and the same holds for the sequences \( \sum_{i=2}^{R} \sum_{k \in H_i} f(S_k) \) and \( \sum_{i=2}^{R} Z_i \).

The key observation is that replacing \( Y_i \) by \( Y_i^* \), and \( Z_i \) by \( Z_i^* \) introduces independence; for a detailed proof see [10, Lemmas 1 and 2]. Also note that \( Y_i^* \overset{d}{=} \sum_{k=1}^{[J_i]} f(U S_k) \) and \( Z_i^* \overset{d}{=} \sum_{k=[H_i]} f(U S_k) \), thus \( Y_i^* \) and \( Z_i^* \) are mean zero random variables with variance given by Lemma 1. The main properties of the approximating variables are summarized as follows.
Lemma 3. Let $f \in \mathcal{F}_p$, and let $H_1, J_1, H_2, J_2, \ldots$ be an arbitrary block decomposition of $\mathbb{N}$. Then $Y_1^*, Y_2^*, \ldots$ are independent, $\mathbb{E}Y_i^* = 0$, and $\mathbb{E}(Y_i^*)^2 = C(f, \nu)|J_i| + o(|J_i|)$. The same holds for $Z_2^*, Z_3^*, \ldots$ with $|J_i|$ replaced by $|H_i|$.

Next, we estimate the perturbation error. The main idea is that in the error $Y_i - Y_i^* = \sum_{k \in J_i} f(W_k - f(W_k^*))$ the variable $W_k^*$ is obtained from $W_k$ by replacing $T_{H_i}$ by $U_{H_i}$. Since $\mathbb{E} d(T_{H_i}, U_{H_i})^p \leq 2\Delta_{|H_i|}$ by construction, the error cannot be too large.

Lemma 4. Let $f \in \mathcal{F}_p$, and let $H_1, J_1, H_2, J_2, \ldots$ be an arbitrary block decomposition of $\mathbb{N}$. For any integers $1 \leq R < S$,

$$
\mathbb{E} \left( \sum_{i=R+1}^{S} (Y_i - Y_i^*) \right)^2 \ll \|f\|_{p-\text{H"{o}ld}}^2 \sum_{i=R+1}^{S} (\Delta_{|H_i|}|J_i| + 1),
$$

$$
\mathbb{E} \left( \sum_{i=R+1}^{S} (Z_i - Z_i^*) \right)^2 \ll \|f\|_{p-\text{H"{o}ld}}^2 \sum_{i=R+1}^{S} (\Delta_{|J_i|}|H_i| + 1)
$$

with implied constants depending only on $\nu$, $p$ and the metric $d$.

Proof. We only give a proof of the first estimate; the proof of the second one is exactly the same. We may assume that $\|f\|_{p-\text{H"{o}ld}} = 1$. We need to estimate

$$
\mathbb{E} \left( \sum_{i=R+1}^{S} (Y_i - Y_i^*) \right)^2 = \sum_{i=R+1}^{S} \mathbb{E}(Y_i - Y_i^*)^2 + 2 \sum_{R+1 \leq i < j \leq S} \mathbb{E}(Y_i - Y_i^*)(Y_j - Y_j^*). \quad (11)
$$

First, fix $i \geq 2$ and consider $Y_i - Y_i^* = \sum_{k \in J_i} f(W_k - f(W_k^*))$. By construction, here $W_k$ and $W_k^*$ are of the form $W_k = AT_{H_i} \prod_{k \in J_i, \ell \leq k} X_{\ell}$ and $W_k^* = AU_{H_i} \prod_{k \in J_i, \ell \leq k} X_{\ell}$, where $A$ is independent of the other factors. Let $\alpha$ denote the distribution of $A$ (in fact, it is a convolution power of $\nu$), and recall that $(T_{H_i}, U_{H_i})$ has joint distribution $\vartheta_{|H_i|}$. By the independence of the factors we have

$$
\mathbb{E}(Y_i - Y_i^*)^2 = \int_G \int_{G \times G} \mathbb{E} \left( \sum_{k=1}^{|J_i|} (f(xyS_k) - f(xy'S_k)) \right)^2 d\vartheta_{|H_i|}(y,y') d\alpha(x).
$$

Observe that for any given $x, y, y' \in G$, the function $g(z) = f(xy) - f(xy')$ satisfies $\int_G g d\mu = 0$, and $\|g\|_{p-\text{H"{o}ld}} \leq 2\|f\|_{p-\text{H"{o}ld}} = 2$. Further,

$$
\|g\|_1 = \int_G |f(xy) - f(xy')| d\mu(z)
\leq \|f\|_{p-\text{H"{o}ld}} \int_G d(xy, xy')^p d\mu(z)
= d(y, y')^p
$$
by the bi-invariance of the metric \( d \). Applying Lemma 2 to \( g \), we thus get

\[
E \left( \sum_{k=1}^{\lfloor J_i \rfloor} (f(xyS_k) - f(xy'S_k)) \right)^2 \ll d(y, y')^p |J_i| + 1
\]

uniformly in \( x, y, y' \in G \). By the construction of \( \vartheta_{|J_i|} \) we finally deduce

\[
E(Y_i - Y_i^*)^2 \ll \int_G \int_{G \times G} (d(y, y')^p |J_i| + 1) \, d\vartheta_{|J_i|}(y, y') \, d\alpha(x) \tag{12}
\]

Next, fix \( 2 \leq i < j \) and consider \( Y_j - Y_j^* = \sum_{k \in J_i}^{} (f(W_k) - f(W_k^*)) \). By construction, here \( W_k \) and \( W_k^* \) can be written in the form \( W_k = BC_k \) and \( W_k^* = BC_k^* \), where \( B = T_{H_1, S_{J_1}} \cdots T_{H_i, S_{J_i}} \). Let \( \mathcal{G}_i \) be the \( \sigma \)-algebra generated by the variables \( X_{\ell}, \ell \in J_i \cup \cdots \cup J_j \), the variables \( T_{H_1}, \ldots, T_{H_i} \), and \( U_{H_j} \). Then \( Y_i, Y_i^* \) and \( B \) are \( \mathcal{G}_i \)-measurable, and \( C_k \) and \( C_k^* \) are independent of \( \mathcal{G}_i \). Therefore

\[
|\mathbb{E} ( (Y_i - Y_i^*)(Y_j - Y_j^*) \mid \mathcal{G}_i ) | = |Y_i - Y_i^*| \left| \sum_{k \in J_j} \mathbb{E} ( f(BC_k) - f(BC_k^*) \mid \mathcal{G}_i ) \right|.
\]

Since \( C_k^* \) is uniformly distributed and \( \mathbb{E} f(bC_k^*) = 0 \) for all \( b \in G \), here \( \mathbb{E} ( f(BC_k^*) \mid \mathcal{G}_i ) = 0 \). A trivial estimate gives \( |\mathbb{E} ( f(BC_k) \mid \mathcal{G}_i )| \leq \sup_{b \in G} |\mathbb{E} f(bC_k)| \). Note that \( C_k \) has distribution \( \nu^m \) with \( m = k - \max J_i \). Applying (7) to \( g(x) = f(bx) \) we thus obtain \( |\mathbb{E} f(bC_k)| \leq \Delta_{k - \max J_i} \) for any \( b \in G \). Hence

\[
|\mathbb{E} ( (Y_i - Y_i^*)(Y_j - Y_j^*) \mid \mathcal{G}_i ) | \leq |Y_i - Y_i^*| \left| \sum_{k \in J_j} \Delta_{k - \max J_i} \right|.
\]

By taking the (total) expectation and summing over \( j = i + 1, i + 2, \ldots, \)

\[
\sum_{j=i+1}^{\infty} |\mathbb{E} ( (Y_i - Y_i^*)(Y_j - Y_j^*) )| \leq |Y_i - Y_i^*| \sum_{j=i+1}^{\infty} \sum_{k \in J_j} \Delta_{k - \max J_i} \ll |Y_i - Y_i^*|.
\]

This estimate together with (12) shows that in (11) we have

\[
\sum_{i=R+1}^{S} \mathbb{E}(Y_i - Y_i^*)^2 + 2 \sum_{R+1 \leq i < j \leq S} \mathbb{E} ( (Y_i - Y_i^*)(Y_j - Y_j^*) ) \ll \sum_{i=R+1}^{S} (\Delta_{|H_i|}|J_i| + 1)
\]

for any \( 1 \leq R < S \), as claimed. \( \square \)

### 3.3. Approximation by a Wiener process

In this section we finish the proof of Theorem 5, and prove the Remark thereafter.
Proof of Theorem 5. Fix \( f \in \mathcal{F}_p \), and consider the block decomposition \( H_1, J_1, H_2, J_2, \ldots \) of \( \mathbb{N} \) with \( |H_i| = [i^{1/4}] \) and \( |J_i| = [i^{1/2}] \). For any \( t \geq 1 \) let \( R(t) \) denote the positive integer for which \( [t] \in H_{R(t)} \cup J_{R(t)} \); in particular, \( R(t) = \Theta(t^{2/3}) \). Replacing \( t \) by \( \max J_{R(t)} \) in the sum \( \sum_{k \leq t} f(S_k) \), we thus have

\[
\sum_{k \leq t} f(S_k) = \sum_{i=1}^{R(t)} \sum_{k \in J_i} f(S_k) + \sum_{i=2}^{R(t)} \sum_{k \in H_i} f(S_k) + O\left(t^{1/3}\right). \tag{13}
\]

The first double sum has the same distribution as \( \sum_{i=1}^{R(t)} Y_i \), whereas the second double sum has the same distribution as \( \sum_{i=2}^{R(t)} Z_i \). Recall that \( k \Delta_k \to 0 \) as \( k \to \infty \); in particular, \( \Delta_{|H_i||J_i|} + 1 \ll i^{1/4} \) and \( \Delta_{|J_i||H_i|} + 1 \ll 1 \).

Lemma 4 and the Rademacher–Menshov inequality [21, Theorem F] give that for any integer \( m \geq 1 \),

\[
\mathbb{E} \left( \max_{1 \leq R \leq 2^m} \left| \sum_{i=1}^{R} (Y_i - Y_i^*) \right|^2 \right) \ll m^2 \sum_{i=1}^{2^m} (\Delta_{|H_i||J_i|} + 1) \ll m^2 2^{m/2}. \]

By the Chebyshev inequality and the Borel–Cantelli lemma, for any \( \varepsilon > 0 \)

\[
\max_{1 \leq R \leq 2^m} \left| \sum_{i=1}^{R} (Y_i - Y_i^*) \right| \ll 2^{(5/8+\varepsilon)m} \text{ a.s.}
\]

Repeating the same arguments for \( Z_i - Z_i^* \), we deduce

\[
\left| \sum_{i=1}^{R} (Y_i - Y_i^*) \right| \ll R^{5/8+\varepsilon} \text{ a.s. and } \left| \sum_{i=2}^{R} (Z_i - Z_i^*) \right| \ll R^{1/2+\varepsilon} \text{ a.s.} \tag{14}
\]

Recall from Lemma 3 that \( Z_3^*, Z_3^*, \ldots \) are independent, mean zero random variables with \( \mathbb{E}(Z_i^*)^2 \ll |H_i| \leq i^{1/4} \). In particular, \( \sum_{i=1}^{\infty} \mathbb{E}(Z_i^*)^2 / i^{(5/8+\varepsilon)2} < \infty \), and by the strong law of large numbers \( \sum_{i=1}^{R} Z_i^* \ll R^{5/8+\varepsilon} \) a.s. for all \( \varepsilon > 0 \). This estimate, together with (14), show that the processes \( \sum_{k \leq t} f(S_k) \) and \( \sum_{i=1}^{R(t)} Y_i^* \) are \( o(R(t)^{5/8+\varepsilon}) \)-equivalent, i.e. \( o(R(t)^{5/12+\varepsilon}) \)-equivalent for all \( \varepsilon > 0 \). It will thus be enough to prove the claims for \( \sum_{i=1}^{R(t)} Y_i^* \) instead of \( \sum_{k \leq t} f(S_k) \).

Recall from Lemma 3 that \( Y_1^*, Y_2^*, \ldots \) are independent, mean zero random variables. We start with the easier case \( C(f, \nu) = 0 \). In this case \( \sum_{i=1}^{R} \mathbb{E}(Y_i^*)^2 = \sum_{i=1}^{R} o(|J_i|) = o(R^{3/2}) \), and \( |Y_i^*| \ll |J_i| \leq i^{1/2} \) by construction. From a general form of the strong law of large numbers and Kolmogorov’s exponential bounds it follows that \( \sum_{i=1}^{R} Y_i^* = o(\sqrt{R^{1/2} \log \log R}) \) a.s.; for a detailed proof see the almost sure stability criterion in Loève [20, p. 270]. Therefore \( \sum_{i=1}^{R(t)} Y_i^* = o(\sqrt{t \log \log t}) \) a.s., and consequently \( \sum_{k \leq t} f(S_k) = o(\sqrt{t \log \log t}) \) a.s. Further, we have \( \sum_{i=1}^{R(t)} Y_i^* = o(R(t)^{3/4}) = o(t^{1/2}) \) in \( L^2 \), and consequently \( \sum_{k \leq t} f(S_k) = o(t^{1/2}) \) in \( L^2 \). In particular, \( t^{-1/2} \sum_{k \leq t} f(S_k) \to 0 \).
Next, assume $C(f, \nu) > 0$. It remains to find an almost sure approximation of $\sum_{i=1}^{R(t)} Y_i^*$ by a Wiener process. In fact, we will prove that $\sigma(t)^2 = \sum_{i=1}^{R(t)} \mathbb{E}(Y_i^*)^2$ satisfies the claims of Theorem 5. First, note that by Lemma 3 we have

$$\sigma(t)^2 = \sum_{i=1}^{R(t)} (C(f, \nu)|J_i| + o(|J_i|)) = C(f, \nu)t + o(t).$$

Now let $V_R = \sum_{i=1}^{R} \mathbb{E}(Y_i^*)^2 = \Theta(R^{3/2})$, $R = 1, 2, \ldots$, and for any $t \geq V_1$ let $R(t)$ be the positive integer for which $V_{R(t)} \leq t < V_{R(t)+1}$. An invariance principle proved by Strassen [29, Theorem 4.4] implies that if $\sum_{i=1}^{\infty} \mathbb{E}|Y_i|^p |V_i^{-\theta/2} < \infty$ for some $p > 2$ and $0 \leq \theta < 1$, then the stochastic processes $\sum_{i=1}^{R(t)} Y_i^*$ and $W(t)$ are $o(t^{(1+\theta)/4} \log t)$-equivalent, where $W(t)$ is a standard Wiener process. By construction $|Y_i^*| \ll |J_i| \ll t^{1/2}$, and so $\mathbb{E}|Y_i|^p \ll t^{p/2}$. Therefore for any $\theta > 2/3$ there exists a large enough $p > 2$ such that

$$\sum_{i=1}^{\infty} \mathbb{E}|Y_i|^p \ll \sum_{i=1}^{\infty} i^{p(1/2-3\theta/4)} < \infty.$$

Hence by the Strassen invariance principle $\sum_{i=1}^{R(t)} Y_i^*$ and $W(t)$ are $o(t^{5/12+\epsilon})$-equivalent for all $\epsilon > 0$. Using $\sigma(t)^2 \sim C(f, \nu)t$, it is not difficult to see that $\sum_{i=1}^{R(t)} Y_i^*$ and $W(\sigma(t)^2)$ are also $o(t^{5/12+\epsilon})$-equivalent for all $\epsilon > 0$. Since $R(\sigma(t)^2) = R(t)$ for all $t$, the last relation is exactly what we wanted to prove.

We now prove the remark made after Theorem 5. If $\Delta_k \ll k^{-(1+c)}$ for some $0 < c \leq 1/2$, then in Lemma 1 the error term $o(N)$ can be replaced by $O(N^{1-c})$. In particular, in Lemma 3 we have $\mathbb{E}(Y_i^*)^2 = C(f, \nu)|J_i| + O(|J_i|^{1-c})$, and so

$$\sigma(t)^2 = \sum_{i=1}^{R(t)} (C(f, \nu)|J_i| + O(|J_i|^{1-c}))$$

$$= C(f, \nu) \sum_{i=1}^{R(t)} (|J_i| + |H_i|) + O\left( R(t)^{5/4} + R(t)^{3/2-\epsilon/2} \right)$$

$$= C(f, \nu)t + O\left( t^{1-c/3} \right).$$

Finally, we show that in this case $W(\sigma(t)^2)$ and $\sqrt{C(f, \nu)}W(t)$ are $o(t^{1/2-c/6+\epsilon})$-equivalent for all $\epsilon > 0$. Recalling the distribution of the running maximum of a Wiener process, for an arbitrarily large constant $K > 0$ and any $\epsilon > 0$,

$$\Pr\left( \sup_{s \in [n-Kn^{1-c/3}, n+Kn^{1-c/3}]} |W(s) - W(n)| \geq n^{1/2-c/6+\epsilon/2} \right) \ll \frac{1}{n^2}.$$

The Borel–Cantelli lemma thus shows that $W(\sigma(t)^2)$ and $W(C(f, \nu)t)$ are both $o(t^{1/2-c/6+\epsilon})$-equivalent, say, $W(|C(f, \nu)t|)$. By the scaling property of the Wiener process, $W(C(f, \nu)t)$ and $\sqrt{C(f, \nu)}W(t)$ have the same distribution.
4. Random walks on the torus

Recall that the Fourier coefficients of any $p$-Hölder function $f : \mathbb{R}^d / \mathbb{Z}^d \rightarrow \mathbb{R}$ decay at the rate $|\hat{f}(h)| \leq \|f\|_{p\text{-Höld}} \cdot 2^{-1-p}\|h\|^{-p}_\infty$. This follows easily from the identity

$$\hat{f}(h) = \frac{1}{2} \int_{[0,1)^d} \left( f(x) - f\left(x - \frac{1}{2h_j} e_j\right) \right) e^{-2\pi i \langle h, x \rangle} \, dx,$$

where $e_j = (0, \ldots, 1, \ldots, 0)$ is the $j$th standard basis vector, applied with the index $1 \leq j \leq d$ for which $\|h\|_\infty = |h_j|$.

For any positive integer $H$, let

$$F_H(x) = \sum_{|h| < H} \left( 1 - \frac{|h|}{H} \right) e^{2\pi i hx} = \frac{1}{H} \frac{\sin^2(\pi Hx)}{\sin^2(\pi x)}$$

denote the 1-dimensional Fejér kernel of order $H$, and let

$$F_H^{(d)}(x_1, x_2, \ldots, x_d) = F_H(x_1) F_H(x_2) \cdots F_H(x_d)$$

denote its $d$-dimensional counterpart. Recall that given a function $f \in L^1(\mathbb{R}^d / \mathbb{Z}^d)$, the convolution $f_H = f * F_H^{(d)}$ is a Cesàro average of the formal Fourier series of $f$; more precisely, $f_H$ is a trigonometric polynomial whose Fourier coefficients are $\hat{f}_H(h) = \prod_{j=1}^d (1 - |h_j|/H) \hat{f}(h)$ if $\|h\|_\infty < H$, and $\hat{f}_H(h) = 0$ otherwise.

**Proof of Proposition 2.** We will prove that under the hypotheses of Theorem 1 for any $f \in F_p$,

$$\sigma^2 = C(f, \nu) = \sum_{h \in \mathbb{Z}^d, \ h \neq 0} |\hat{f}(h)|^2 \frac{1 - |\hat{\nu}(h)|^2}{1 - |\hat{\nu}(h)|^2}, \quad (15)$$

where $C(f, \nu)$ is as in (5). Here $|\hat{\nu}(h)| < 1$ for every $h \neq 0$. Indeed, by the assumption $W_p(\nu^* k, \mu) \rightarrow 0$ we have $\nu^* k \rightarrow \mu$ weakly as $k \rightarrow \infty$, and hence

$$\hat{\nu}(h)^k = \nu^* k(h) = \int_{\mathbb{R}^d/\mathbb{Z}^d} e^{-2\pi i \langle h, x \rangle} \, d\nu^* k(x) \rightarrow \int_{\mathbb{R}^d/\mathbb{Z}^d} e^{-2\pi i \langle h, x \rangle} \, d\mu(x) = 0.$$

In particular, (15) implies that $\sigma = 0$ if and only if $\hat{f}(h) = 0$ for all $h \in \mathbb{Z}^d$. The latter condition is equivalent to $f = 0$ a.e., and by the continuity of $f$, to $f = 0$.

We establish (15) in two steps. First, assume that $f$ is a trigonometric polynomial; that is, assume $f(h) \neq 0$ for finitely many $h \in \mathbb{Z}^d$. Since the variable $U$ in the definition (5) is independent of $X_1, X_2, \ldots,$

$$\mathbb{E} f(U) f(U + S_k) = \int_{\mathbb{R}^d/\mathbb{Z}^d} \int_{\mathbb{R}^d/\mathbb{Z}^d} f(u) f(u + x) \, d\mu(u) d\nu^* k(x) = \mathbb{E} g(S_k),$$

for some function $g$. This follows from the identity

$$\mathbb{E} f(U) f(U + S_k) = \int_{\mathbb{R}^d/\mathbb{Z}^d} \mathbb{E} f(U) f(U + x) \, d\mu(x) = \int_{\mathbb{R}^d/\mathbb{Z}^d} f(x) \, d\mu(x) = \mathbb{E} g(S_k).$$

The second step involves a careful analysis of the properties of $f$ and $\nu^* k$.
Equidistribution of random walks on compact groups II.

where \( g(x) = \int_{\mathbb{R}^d/\mathbb{Z}^d} f(u) f(u+x) \, d\mu(u) \). Since \( \hat{g}(h) = |\hat{f}(h)|^2 \) and \( g \) is also a trigonometric polynomial,

\[
Eg(S_k) = \sum_{h \in \mathbb{Z}^d} \hat{g}(h)Ee^{2\pi i \langle h, S_k \rangle} = \sum_{h \in \mathbb{Z}^d} |\hat{f}(h)|^2 \hat{\nu}(h)^{k}.
\]

Using \( \hat{f}(0) = 0 \) and the Parseval formula \( Ef(U)^2 = \sum_{h \in \mathbb{Z}^d} |\hat{f}(h)|^2 \), in (5) we thus have

\[
C(f, \nu) = Ef(U)^2 + 2 \sum_{k=1}^{\infty} Eg(S_k) = \sum_{h \in \mathbb{Z}^d, h \neq 0} |\hat{f}(h)|^2 \left( 1 + 2 \frac{\hat{\nu}(h)}{1 - \hat{\nu}(h)} \right).
\]

Combining the \( h \) and \(-h\) terms, (15) follows.

Finally, let us show (15) for arbitrary \( f \in F_p \). Consider the convolutions \( f_H = f * F_H^{(d)} \), \( H = 1, 2, \ldots \). By the special case shown above,

\[
C(f_H, \nu) = \sum_{h \in \mathbb{Z}^d} \prod_{0 < |h| < H} \left( 1 - \frac{|h|}{H} \right)^2 |\hat{f}(h)|^2 \frac{1 - |\hat{\nu}(h)|^2}{|1 - \hat{\nu}(h)|^2} \sum_{h \in \mathbb{Z}^d, h \neq 0} |\hat{f}(h)|^2 \frac{1 - |\hat{\nu}(h)|^2}{|1 - \hat{\nu}(h)|^2}
\]

as \( H \to \infty \) (see e.g. from the monotone convergence theorem).

Now let \( g(x) = \int_{\mathbb{R}^d/\mathbb{Z}^d} f(u) f(u+x) \, d\mu(u) \) and \( g_H(x) = \int_{\mathbb{R}^d/\mathbb{Z}^d} f_H(u) f_H(u+x) \, d\mu(u) \). Applying (7),

\[
|C(f, \nu) - C(f_H, \nu)| = \left| Ef(U)^2 + 2 \sum_{k=1}^{\infty} Eg(S_k) - Ef_H(U)^2 - 2 \sum_{k=1}^{\infty} Eg_H(S_k) \right|
\]

\[
\leq |Ef(U)^2 - Ef_H(U)^2| + 2 \sum_{k=1}^{\infty} |E(g - g_H)(S_k)|
\]

\[
\leq \|f\|_2^2 - \|f_H\|_2^2 + 2 \sum_{k=1}^{\infty} \|g - g_H\|_{p-Höld} W_p(\nu^k, \mu)
\]

\[
\ll \|f\|_2^2 - \|f_H\|_2^2 + \|g - g_H\|_{p-Höld}.
\]

Since \( f_H \to f \) in \( L^2(\mathbb{R}^d/\mathbb{Z}^d) \), the first term goes to 0 as \( H \to \infty \). It will thus be enough to show \( \|g - g_H\|_{p-Höld} \to 0 \) as \( H \to \infty \); indeed, this will imply \( C(f_H, \nu) \to C(f, \nu) \), and (15) will follow.

Writing \( f = f_H + (f - f_H) \) in the definition of \( g \), and applying the integral transformation \( u \mapsto u - x \) in one of the terms, we get

\[
g(x) - g_H(x) = \int_{\mathbb{R}^d/\mathbb{Z}^d} (f - f_H)(u) (f(u + x) + f_H(u - x)) \, d\mu(u)
\]

for all \( x \in \mathbb{R}^d/\mathbb{Z}^d \). Therefore \( \|g - g_H\|_{p-Höld} \leq \|f_H\|_{1} (\|f\|_{p-Höld} + \|f_H\|_{p-Höld}) \). Using the fact that \( F_H^{(d)} \) is a nonnegative kernel, it is not difficult to see that \( \|f_H\|_{p-Höld} \leq \|f\|_{p-Höld} \). Hence \( \|g - g_H\|_{p-Höld} \leq \|f\|_{p-Höld} + \|f\|_{1} \cdot 2 \|f\|_{p-Höld} \to 0 \) as \( H \to \infty \). This finishes the proof of (15) for arbitrary \( f \in F_p \). \( \square \)
Proof of Proposition 3. Let \( f \in \mathcal{F}_p \) be such that \( \|f\|_{\text{Hold}} \leq 1 \), and let \( H > 1 \) be an integer. Consider the convolution \( f_H = f \ast F_H^{(d)} \). Clearly,

\[
\left| \int_{\mathbb{R}^d / \mathbb{Z}^d} f \, dv_1 - \int_{\mathbb{R}^d / \mathbb{Z}^d} f \, dv_2 \right|
\]

\[
\leq 2\|f - f_H\|_\infty + \left| \int_{\mathbb{R}^d / \mathbb{Z}^d} f_H \, dv_1 - \int_{\mathbb{R}^d / \mathbb{Z}^d} f_H \, dv_2 \right|
\]

\[
= 2\|f - f_H\|_\infty + \left| \sum_{h \in \mathbb{Z}^d} \hat{f}_H(h) (\hat{\nu}_1(-h) - \hat{\nu}_2(-h)) \right|
\]

\[
\leq 2\|f - f_H\|_\infty + \sum_{0 < \|h\|_\infty < H} \prod_{j=1}^d \left( 1 - \frac{|h_j|}{H} \right) |\hat{\nu}_1(h) - \hat{\nu}_2(h)|
\]

\[
\leq 2\|f - f_H\|_\infty + \frac{1}{2^{1+p}} \sum_{h \in \mathbb{Z}^d} \prod_{j=1}^d \left( 1 - \frac{|h_j|}{H} \right) \frac{|\hat{\nu}_1(h) - \hat{\nu}_2(h)|}{\|h\|_\infty^p}.
\]

On the other hand,

\[
\|f - f_H\|_\infty = \sup_{x \in \mathbb{R}^d / \mathbb{Z}^d} \left| \int_{\mathbb{R}^d / \mathbb{Z}^d} (f(x) - f(x - u)) F_H^{(d)}(u) \, d\mu(u) \right|
\]

\[
\leq \int_{[-1/2,1/2]^d} |u|^p F_H^{(d)}(u) \, du
\]

\[
\leq \int_{-1/2}^{1/2} \cdots \int_{-1/2}^{1/2} (|u_1|^p + \cdots + |u_d|^p) F_H(u_1) \cdots F_H(u_d) \, du_1 \cdots du_d
\]

\[
= d \int_{-1/2}^{1/2} |u|^p F_H(u) \, du.
\]

Note that the 1-dimensional Fejér kernel satisfies \( F_H(u) \leq H^{-1} \sin^2(\pi H u)/(4u^2) \) on \([-1/2,1/2] \). Applying the integral transformation \( y = Hu \),

\[
\|f - f_H\|_\infty \leq d \int_{-H/2}^{H/2} |y|^{p-2} \sin^2(\pi y) \, dy.
\]

Assume now, that \( 0 < p < 1 \). Extending the range of integration, we arrive at the Mellin transform of \( \sin^2(\pi y) \):

\[
\|f - f_H\|_\infty \leq \frac{d}{4H^p} \int_{-\infty}^{\infty} |y|^{p-2} \sin^2(\pi y) \, dy
\]

\[
= \frac{d}{4H^p} (2\pi)^{1-p} \sin(p\pi/2) \Gamma(p+1).
\]
where $\Gamma$ is the Gamma function. This estimate yields an upper bound in (16) independent of $f$; in particular, by Kantorovich duality

$$W_p(\nu_1, \nu_2) \leq \frac{dC_p}{H^p} + \frac{1}{2^{1+p}} \sum_{h \in \mathbb{Z}^d} \frac{1}{0 < \|h\|_\infty < H} \prod_{j=1}^d \left( 1 - \frac{|h_j|}{H} \right) \frac{|\hat{\nu}_1(h) - \hat{\nu}_2(h)|}{\|h\|_\infty^p}$$  \hspace{1cm} (18)

where

$$C_p = 2^{-p} \pi^{1-p} \frac{\sin(p\pi/2)}{p(1-p)} \Gamma(p+1).$$

To deduce the form stated in Proposition 3, note that $\sin(p\pi/2) / p(1-p)$, and that $\max_{0 \leq p \leq 1} \Gamma(p+1) = \Gamma(1) = 1$ by the convexity of $\Gamma$ on $(0, \infty)$; therefore $C_p \leq (\pi^2/2)/(1-p) \leq 5/(1-p)$.

Finally, assume that $p = 1$. Then (17) can be written as

$$\|f - fh\|_\infty \leq \frac{d}{2H} \left( \int_0^1 \frac{\sin^2(\pi y)}{y} dy + \int_{1}^{H/2} \frac{\sin^2(\pi y)}{y} dy \right).$$

Applying a trigonometric identity, integration by parts and the second mean value theorem for integrals:

$$\int_1^{H/2} \frac{\sin^2(\pi y)}{y} dy = \int_1^{H/2} \frac{1 - \cos(2\pi y)}{2y} dy$$

$$= \frac{\log(H/2)}{2} - \int_1^{H/2} \frac{\sin(2\pi y)}{4\pi y^2} dy$$

$$= \frac{\log(H/2)}{2} - \frac{1}{4\pi} \int_1^{\xi} \sin(2\pi y) dy$$

$$\leq \frac{\log(H/2)}{2}$$

with some $1 \leq \xi \leq H/2$. Once again we obtain an upper bound in (16) independent of $f$; in particular, by Kantorovich duality

$$W_1(\nu_1, \nu_2) \leq \frac{d}{2} \frac{C + \log H}{H} + \frac{1}{4} \sum_{h \in \mathbb{Z}^d} \prod_{0 < \|h\|_\infty < H}^d \left( 1 - \frac{|h_j|}{H} \right) \frac{|\hat{\nu}_1(h) - \hat{\nu}_2(h)|}{\|h\|_\infty}$$  \hspace{1cm} (19)

where

$$C = 2 \int_0^1 \frac{\sin^2(\pi y)}{y} dy - \log 2 \approx 1.7445.$$

$\square$
Proof of Theorem 6. In this proof constants and implied constants depend only on the distributions of $\xi_1, \xi_2, \ldots, \xi_r, I$, the value of the infimum in (6), $d$, $r$ and $\rho$. We start by estimating $|\tilde{\nu}(h)|$, $h \in \mathbb{Z}^d$. For every $1 \leq i \leq r$ let $D_i > 0$ be the greatest common divisor of the (finite or infinite) set of integers

$$\{a - b : \Pr(\xi = a), \Pr(\xi = b) > 0}\}.$$ 

That is, $\xi_i$ is a lattice variable with maximal span $D_i$. Note that these are well defined because $\xi_1, \xi_2, \ldots, \xi_r$ are assumed to be nondegenerate. It is not difficult to see that the characteristic function $\varphi_i$ of $\xi_i$ satisfies

$$|\varphi_i(2\pi x)| \leq 1 - c_i \|D_i x\|_{\mathbb{R}/\mathbb{Z}}^2 \ (20)$$

for all $x \in \mathbb{R}$ with some constant $c_i > 0$. Indeed, $|\varphi_i(2\pi x)| = 1$ if and only if $x$ is an integer multiple of $1/D_i$; in addition, an estimate of the form (20) holds for all $x$ in an open neighborhood of $0$, see Petrov [23, p. 14]. Since both sides of (20) have period $1/D_i$, (20) follows with a small enough constant $c_i > 0$.

Let $e(x) = e^{2\pi i x}$, and let $D > 0$ denote the least common multiple of the integers $D_i$, $1 \leq i \leq r$. By the construction of $\nu$ and (20),

$$|\tilde{\nu}(h)| = \left| \sum_{i=1}^{r} \sum_{n=1}^{D_i} \Pr(I = i) \Pr(\xi = n)e(-\langle h, n\alpha_i \rangle) \right| \leq \sum_{i=1}^{r} \Pr(I = i) \left| \sum_{n=1}^{D_i} \Pr(\xi = n)e(-\langle h, \alpha_i \rangle n) \right| = \sum_{i=1}^{r} \Pr(I = i) |\varphi_i(-2\pi \langle h, \alpha_i \rangle)| \leq 1 - \sum_{i=1}^{r} \Pr(I = i) c_i \|D_i \langle h, \alpha_i \rangle\|_{\mathbb{R}/\mathbb{Z}}^2 \leq 1 - \sum_{i=1}^{r} \Pr(I = i) \frac{D_i}{D} \|D \langle h, \alpha_i \rangle\|_{\mathbb{R}/\mathbb{Z}}^2 \leq 1 - c \max_{1 \leq i \leq r} \|D \langle h, \alpha_i \rangle\|_{\mathbb{R}/\mathbb{Z}}^2$$

(21)

with some constant $c > 0$. Note that we used the general inequality $\|D_i x\|_{\mathbb{R}/\mathbb{Z}} \geq (D_i/D)\|Dx\|_{\mathbb{R}/\mathbb{Z}}, x \in \mathbb{R}$, which follows from the subadditivity of the function $\|\cdot\|_{\mathbb{R}/\mathbb{Z}}$, and in the last step the assumption $\Pr(I = i) > 0$.

Let us apply Proposition 3 to $\nu_1 = \nu^k, \nu_2 = \mu$ with $H = \psi^{-1}(\sqrt{k})(2D)$; in particular, $\psi(2DH)^2 \leq k$. By (21),

$$|\nu^k(h) - \tilde{\mu}(h)| = |\tilde{\nu}(h)|^k \leq \exp \left( -ck \max_{1 \leq i \leq r} \|D \langle h, \alpha_i \rangle\|_{\mathbb{R}/\mathbb{Z}}^2 \right)$$
Equidistribution of random walks on compact groups II.

for any $h \in \mathbb{Z}^d$, $h \neq 0$. It will thus be enough to prove

$$\sum_{h \in \mathbb{Z}^d, 0 < \|h\|_\infty < DH} \exp \left( -ck \max_{1 \leq i \leq r} \|\langle h, \alpha_i \rangle\|_{\mathbb{R}/\mathbb{Z}}^2 \right) \ll \frac{1}{H^p}. \quad (22)$$

For any integer $1 \leq m \leq DH$, let

$$B_m = \sum_{h \in \mathbb{Z}^d, 0 < \|h\|_\infty < m} \exp \left( -ck \max_{1 \leq i \leq r} \|\langle h, \alpha_i \rangle\|_{\mathbb{R}/\mathbb{Z}}^2 \right). \quad (23)$$

Let $C_0 = [-K/(2\psi(2m)), K/(2\psi(2m))]^r$ denote the axis parallel cube in $\mathbb{R}^r$ centered at the origin with edge length $K/\psi(2m)$, where the constant $K > 0$ denotes the value of the infimum in (6). The translates $C_q := C_0 + qK/\psi(2m)$, $q \in \mathbb{Z}^r$ decompose $\mathbb{R}^r$ into congruent axis parallel cubes. Consider the point set

$$P = \{g(h) : h \in \mathbb{Z}^d, 0 < \|h\|_\infty < m \},$$

where $g(h)$ denotes the unique representative of $((h, \alpha_1), (h, \alpha_2), \ldots, (h, \alpha_r)) + \mathbb{Z}^r$ in $[-1/2, 1/2]^r$. Observe that $\|g(h)\|_\infty = \max_{1 \leq i \leq r} \|\langle h, \alpha_i \rangle\|_{\mathbb{R}/\mathbb{Z}}$. We claim that every cube $C_q$ contains at most one point of $P$. Indeed, let $h, h' \in \mathbb{Z}^d$ with $\|h\|_\infty, \|h'\|_\infty < m, h \neq h'$ be arbitrary. By the choice of $K$ and the strict monotonicity of $\psi$,

$$\|g(h) - g(h')\|_\infty \geq \|g(h - h')\|_\infty \geq \max_{1 \leq i \leq r} \|\langle h - h', \alpha_i \rangle\|_{\mathbb{R}/\mathbb{Z}}$$

$$\geq K/\psi(\|h - h'\|_\infty)$$

$$> K/\psi(2m).$$

Therefore $g(h)$ and $g(h')$ cannot lie in the same axis parallel cube with edge length $K/\psi(2m)$. Further, since $g(0) = 0 \in C_0$, no point of $P$ lies in $C_0$.

If $g(h) \in P \cap C_q$ for some $q \in \mathbb{Z}^r$, $q \neq 0$, then

$$\|g(h)\|_\infty \geq \left( \|q\|_\infty - \frac{1}{2} \right) \frac{K}{\psi(2m)} \geq \|q\|_\infty \frac{K}{2\psi(2m)}. $$

Therefore with the constant $a = cK^2/4 > 0$,

$$B_m = \sum_{g(h) \in P} \exp \left( -ck\|g(h)\|_\infty^2 \right) \leq \sum_{q \in \mathbb{Z}^r, q \neq 0} \exp \left( -ak\|q\|_\infty^2/\psi(2m)^2 \right)$$

$$\ll \sum_{\ell = 1}^r \ell^{r-1} \exp \left( -akt^2/\psi(2m)^2 \right),$$

where we used the fact that there are \( \ll \ell^{-1} \) lattice points \( q \in \mathbb{Z}^r \) with \( \|q\|_\infty = \ell \). Note that here the \( \ell = 1 \) term dominates. Indeed, \( \frac{k}{\psi(2m)^2} \geq \frac{k}{\psi(2DH)^2} \geq 1 \) yields

\[
B_m \ll \exp \left( -\frac{ak}{\psi(2m)^2} \right) \sum_{\ell=1}^{\infty} \ell^{-1} \exp \left( -\frac{ak}{2} - \frac{1}{\psi(2m)^2} \right) \\
\leq \exp \left( -\frac{ak}{\psi(2m)^2} \right) \sum_{\ell=1}^{\infty} \ell^{-1} \exp \left( -a \ell^{-1} \right) \\
\ll \exp \left( -\frac{ak}{\psi(2m)^2} \right).
\]

We prove (22) using “multidimensional summation by parts”. Formally, by the definition (23) of \( B_m \) and (24),

\[
\sum_{h \in \mathbb{Z}^d, 0 < \|h\|_\infty < DH} \frac{\exp \left( -ck \max_{1 \leq i \leq r} \|\langle h, \alpha_i \rangle\|_{\mathbb{R}/\mathbb{Z}}^2 \right)}{\|h\|_\infty^p} \\
= \sum_{m=1}^{DH-1} \frac{1}{m^p} \sum_{h \in \mathbb{Z}^d, \|h\|_\infty = m} \exp \left( -ck \max_{1 \leq i \leq r} \|\langle h, \alpha_i \rangle\|_{\mathbb{R}/\mathbb{Z}}^2 \right) \\
= \sum_{m=1}^{DH-1} \frac{1}{m^p} (B_{m+1} - B_m) \\
= \sum_{m=2}^{DH} \left( \frac{1}{m^p} - \frac{1}{(m+1)^p} \right) B_m + \frac{B_{DH}}{(DH+1)^p} \\
\ll \sum_{m=2}^{DH} \frac{\exp \left( -\frac{ak}{\psi(2m)^2} \right)_{m^p+1}}{m^p+1} + \frac{1}{H^{p+1}}.
\]

Here

\[
\frac{k}{\psi(2m)^2} \geq \frac{\psi(2DH)^2}{\psi(2m)^2} \geq \left( \frac{DH}{m} \right)^{2d/r}.
\]

Since the function \( \exp \left( -ax^{2d/r} \right) x^{p+1} \) is bounded on \([1, \infty)\),

\[
\frac{\exp \left( -\frac{ak}{\psi(2m)^2} \right)_{m^p+1}}{m^p+1} \leq \frac{\exp \left( -a(DH/m)^{2d/r} \right)_{m^p+1}}{m^p+1} \ll \frac{1}{H^{p+1}}
\]

for all \( 2 \leq m \leq DH \). Applying this estimate to each term in the last sum in (25), the estimate (22) follows. \( \square \)

Finally, we prove Theorem 7. We start with two auxiliary lemmas. A finite set of points \( A \subset \mathbb{R}^d/\mathbb{Z}^d \) is called an \( R \)-net, if the set of closed balls with center in \( A \) and radius \( R \) (in the Euclidean metric) cover all of \( \mathbb{R}^d/\mathbb{Z}^d \).
Lemma 5. Let \( \vartheta \) be a Borel probability measure on \( \mathbb{R}^d/\mathbb{Z}^d \), and let \( 0 < p \leq 1 \). If \( A \) is a finite \( R \)-net of cardinality \( |A| \), then

\[
W_p(\vartheta, \mu) \geq \frac{d}{d+p} (\omega_d |A|)^{-p/d} - R^p (1 - \vartheta(A)),
\]

where \( \omega_d \) is the volume of the unit ball in \( \mathbb{R}^d \).

**Proof.** By Kantorovich duality, it will be enough to find a function \( f : \mathbb{R}^d/\mathbb{Z}^d \to \mathbb{R} \) with \( \|f\|_{p \text{-Höld}} \leq 1 \), for which

\[
\int_{\mathbb{R}^d/\mathbb{Z}^d} f \, d\mu - \int_{\mathbb{R}^d/\mathbb{Z}^d} f \, d\vartheta \geq \frac{d}{d+p} (\omega_d |A|)^{-p/d} - R^p (1 - \vartheta(A)).
\]

We claim that \( f(x) = \text{dist}(x, A)^p \) satisfies (26), where \( \text{dist}(x, A) \) denotes the Euclidean distance from the point \( x \) to the set \( A \). By a trivial estimate, for any \( t \geq 0 \) we have

\[
\mu \left( \{ x \in \mathbb{R}^d/\mathbb{Z}^d : \text{dist}(x, A)^p \leq t \} \right) \leq |A| \omega_d t^d,
\]

and hence

\[
\int_{\mathbb{R}^d/\mathbb{Z}^d} \text{dist}(x, A)^p \, d\mu(x) = \int_0^\infty \mu \left( \{ x \in \mathbb{R}^d/\mathbb{Z}^d : \text{dist}(x, A)^p > t \} \right) \, dt
\]

\[
\geq \int_0^{(\omega_d |A|)^{-p/d}} \left( 1 - |A| \omega_d t^d/p \right) \, dt
\]

\[
= \frac{d}{d+p} (\omega_d |A|)^{-p/d}.
\]

On the other hand,

\[
\int_{\mathbb{R}^d/\mathbb{Z}^d} \text{dist}(x, A)^p \, d\vartheta(x) \leq R^p (1 - \vartheta(A)),
\]

because the integrand \( \text{dist}(x, A)^p \) is zero on the set \( A \), and at most \( R^p \) everywhere by the fact that \( A \) is an \( R \)-net. The last two relations together imply (26), and we are done. \( \square \)

Lemma 6. Assume that the system of vectors \( \alpha_1, \alpha_2, \ldots, \alpha_r \in \mathbb{R}^d \) is badly approximate, and let \( J_1, J_2, \ldots, J_r \subset \mathbb{Z} \) be finite intervals of integers of size \( |J_i| \geq L \geq 1 \) (\( 1 \leq i \leq r \)). Then the set

\[
A = \left\{ \sum_{i=1}^r n_i \alpha_i + \mathbb{Z}^d : n_i \in J_i \text{ for all } 1 \leq i \leq r \right\} \subset \mathbb{R}^d/\mathbb{Z}^d
\]

is an \( R \)-net for some \( R \ll L^{-r/d} \), with an implied constant depending only on \( \alpha_1, \alpha_2, \ldots, \alpha_r \).
Proof. We may assume that $L \geq 3$ is an odd integer. According to a classical transference principle in the theory of Diophantine approximation [13, p. 84], a system of vectors $\alpha_1, \alpha_2, \ldots, \alpha_r \in \mathbb{R}^d$ is badly approximable if and only if there exists a constant $c > 0$ such that for any integer $N \geq 1$ and any point $x \in \mathbb{R}^d$ the system of inequalities

\[
\left\| \sum_{i=1}^r n_i \alpha_i - x \right\|_{\mathbb{R}^d/\mathbb{Z}^d} \leq c N^{-r/d},
\]

\[
|n_1|, |n_2|, \ldots, |n_r| \leq N
\]

has an integer solution $n_1, n_2, \ldots, n_r$. In particular, the set

\[
\left\{ \sum_{i=1}^r n_i \alpha_i + \mathbb{Z}^d : n_i \in \mathbb{Z}, |n_i| \leq \frac{L-1}{2} \text{ for all } 1 \leq i \leq r \right\}
\]

is an $R$-net with $R = c((L-1)/2)^{-r/d} \ll L^{-r/d}$, being a translate of this set, so is $A$. \qed

**Proof of Theorem 7.** In this proof constants and implied constants depend only on $\alpha_1, \alpha_2, \ldots, \alpha_r, B$ and $p$. Consider a sequence of i.i.d. $\mathbb{Z}^r$-valued random variables $V_1, V_2, \ldots$, where $V_1 = \xi_1 e_1$; here $e_1, e_2, \ldots, e_r$ denote the standard basis vectors in $\mathbb{Z}^r$. Observe that the $i$th coordinate of $V_1$ has expected value $E_i := \Pr(I = i) \mathbb{E}_{\xi_i}$, and variance

\[
\Pr(I = i) \mathbb{E}_{\xi_i}^2 - \Pr(I = i)^2 (\mathbb{E}_{\xi_i})^2 \leq \Pr(I = i) \mathbb{E}_{\xi_i}^2 \leq B.
\]

Applying the Chebyshev inequality in each coordinate, for any $\lambda > 0$,

\[
\Pr \left( \sum_{j=1}^k V_j \notin \prod_{i=1}^r \left[ E_i k - \lambda \sqrt{k}, E_i k + \lambda \sqrt{k} \right] \right) \leq \frac{B}{\lambda^2} \ll \frac{1}{\lambda^2}.
\]

The random walk $\sum_{j=1}^k V_j$ is mapped to our random walk $S_k = \sum_{j=1}^k X_j$ by the function $g : \mathbb{Z}^r \to \mathbb{R}^d/\mathbb{Z}^d$, $g(n_1, n_2, \ldots, n_r) = \sum_{i=1}^r n_i \alpha_i + \mathbb{Z}^d$. Therefore the set

\[
A := \left\{ \sum_{i=1}^r n_i \alpha_i + \mathbb{Z}^d : n_i \in \left[ E_i k - \lambda \sqrt{k}, E_i k + \lambda \sqrt{k} \right] \cap \mathbb{Z} \text{ for all } 1 \leq i \leq r \right\}
\]

satisfies $\Pr(S_k \notin A) \ll 1/\lambda^2$ as well; in other words, $1 - \nu^k(A) \ll 1/\lambda^2$. On the other hand, by Lemma 6 the set $A$ is an $R$-net with some $R \ll \lambda^{-r/d}k^{-r/(2d)}$ of cardinality $|A| \ll \lambda^r k^{r/2}$. Choosing $\lambda > 0$ to be a large enough constant, Lemma 5 thus gives

\[
W_p(\nu^k, \mu) \geq \frac{d}{d+p} (\omega_d |A|)^{-p/d} - R^p (1 - \nu^k(A)) \gg k^{-pr/(2d)},
\]

as claimed. \qed
Equidistribution of random walks on compact groups II.

Acknowledgments

The author is supported by the Austrian Science Fund (FWF), project Y-901.

References

30


