Probability Theory and Related Fields
https://doi.org/10.1007/s00440-018-0864-7

The local limit theorem on nilpotent Lie groups

Robert Hough[1]

Received: 19 January 2018 / Revised: 24 July 2018

Abstract
A local limit theorem is proven on connected, simply connected nilpotent Lie groups, for a class of generating measures satisfying a moment condition and a condition on the characteristic function of the abelianization. The result extends an earlier local limit theorem of Alexopoulos which treated absolutely continuous measures with a continuous density of compact support, and also extends local limit theorems of Breuillard and Diaconis–Hough which treated general measures on the Heisenberg group.

Keywords Random walk on a group · Local limit theorem · Nilpotent group

Mathematics Subject Classification Primary 60F05 · 60B15 · 20B25 · 22E25 · 60J10 · 60E10 · 60F25 · 60G42

1 Introduction

Local limit theorems are an important tool in the analysis of random walk on a group, where pointwise or small scale control on the distribution of the walk can be used to study convergence to equilibrium in strong metrics such as the total variation metric, see [8,22,23] where local limit theorems on abelian groups are used. In [10,11] Breuillard proved a remarkable local limit theorem on the Heisenberg group, permitting a general driving measure subject to some technical conditions, and allowing testing the distribution of the measure against a test function translated arbitrarily on the left

Robert Hough is supported by NSF Grant DMS-1712682, "Probabilistic methods in discrete structures and applications."
The author thanks Persi Diaconis for his continued interest in the project, and Terence Tao and Emmanuel Breuillard for helpful comments.

✉ Robert Hough
robert.hough@stonybrook.edu

[1] Department of Mathematics, Stony Brook University, 100 Nicolls Road, Stony Brook, NY 11794, USA

Published online: 31 July 2018

and the right. An earlier local limit theorem of Alexopoulos, which appeared in this journal [3], treated the case of a measure with a continuous density with respect to Haar measure, see also [4] for the lattice case. The main purpose of this article is to generalize Breuillard's theorem to an arbitrary connected, simply connected nilpotent Lie group, with a rate.

Let G be a connected, simply connected nilpotent Lie group. Denote $C_c(G)$ the continuous functions of compact support on G. A function $f \in C_c(G)$ is said to be Lipschitz if it has Lipschitz push-forward to the Lie algebra $\mathfrak{g}$ of G via the logarithm map. Given $f \in C_c(G)$ and $g \in G$, let $L_g f$ and $R_g f$ denote the left and right translation of f by g. For a Borel probability measure μ on G with finite second homogeneous moments, let L_μ be the corresponding sub-Laplacian and let u_t be the fundamental solution to the heat equation $\left(\frac{\partial}{\partial t} + L_\mu\right) u_t = 0$, $t > 0$. On a general nilpotent Lie group, on which there may not be a natural dilation, the heat kernel plays the role of a limiting Gaussian measure. Given a probability measure ν on $\mathbb{R}^n$, write

$$\hat{\nu}(\xi) = \int_{\mathbb{R}^n} e^{-2\pi i \xi \cdot x} d\nu(x). \tag{1}$$

for the characteristic function.

Theorem 1 *Let G be a connected, simply connected nilpotent Lie group of step s and homogeneous degree D. Let μ be a Borel probability measure on G with projection μ_{ab} to the abelianization $G_{\mathrm{ab}} = G/[G, G]$ satisfying*

i. *(Centered) μ_{ab} is mean zero.*
ii. *(Cramér) There is a constant $c > 0$ and a neighborhood U of 0 in $\hat{G}_{\mathrm{ab}}$ containing 0 such that the characteristic function $\hat{\mu}_{\mathrm{ab}}$ satisfies*

$$\sup_{\xi \in U^c} \left|\hat{\mu}_{\mathrm{ab}}(\xi)\right| < 1 - c. \tag{2}$$

For all $A > 0$, there is a constant $C = C(A, D)$ such that, if μ has $C(A, D) \geqslant 3$ finite homogeneous moments, then uniformly in $g, h \in G$, for all Lipschitz $f \in C_c(G)$, as $N \to \infty$,

$$\left\langle L_g R_h f, \mu^{*N} \right\rangle = \left\langle L_g R_h f, u_N \right\rangle + O_\mu\left(\|f\|_1 N^{-\frac{D+1}{2}}\right) + O_{\mu, A, f}\left(N^{-A}\right). \tag{3}$$

See Sect. 3 for the definition of homogeneous moments and other terminology related to nilpotent Lie groups.

The dependence on f in the second error term may be controlled in terms of the maximum distance of $\operatorname{supp} f$ from the identity. The rate is optimal, as may be seen by projecting to the abelianization. The optimal rate does not hold without some decay condition on the characteristic function of the abelianization, although a weaker condition than Cramér would suffice. A different limit is obtained in the lattice case, again with optimal rate, by [4]. In [20] it is shown that on the the Heisenberg group, the limit statement without a rate can be obtained with the Cramér condition replaced with $\left|\hat{\mu}_{\mathrm{ab}}(\xi)\right| \neq 1$ if $\xi \neq 0$; the corresponding statement on a general nilpotent group

is currently open. Without an arbitrary translation on the right and left, the following theorem is still obtained.

Theorem 2 *Let G be a connected, simply connected nilpotent Lie group of step s and homogeneous degree D. There is a constant $C = C(D)$ such that, if μ is a centered Borel probability measure, having $C(D)$ finite homogeneous moments, and satisfying the non-lattice condition, for all $0 \neq \xi \in \hat{G}_{\mathrm{ab}}$, $\left|\hat{\mu}_{\mathrm{ab}}(\xi)\right| < 1$, then for all $f \in C_c(G)$, as $N \to \infty$,*

$$\left\langle f, \mu^{*N} \right\rangle = \langle f, u_N \rangle + o_{\mu,f}\left(N^{-\frac{D}{2}}\right). \tag{4}$$

This recovers a local limit theorem of Breuillard ([12], Corollary 1.3) as a special case.

1.1 Discussion of method, and possible extensions

Theorem 1 is of the type proved by Breuillard [10,11] in which an arbitrary translation to the test function is permitted on the left and right. The proof given there used the representation theory of the real Heisenberg group. In [11] Breuillard writes that he expects his analysis to carry through to general Heisenberg groups, but that new methods would need to be developed to handle the higher step cases treated here. Alexopoulos' theorems hold in the greater generality of groups of polynomial volume growth. The proofs first establish the results in the connected, simply connected nilpotent case using time domain PDE methods. It is of interest to obtain the local theorem for general measures in this full generality.

Theorem 1 is proved via harmonic analysis on the Lie algebra, extending the 'characteristic function approach' from the abelian setting, see e.g. [24]. At phases much larger than the scale of the distribution, the i.i.d. nature of the increments of the walk is used with a rearrangement group action, followed by the Gowers–Cauchy–Schwarz inequality applied to the characteristic function. This has the effect of taking iterated commutators on the group G to reduce to the abelian case; an early form of this argument appears in the author's work with Diaconis in [20] treating random walks on finite nilpotent groups. Green and Tao [21] studied the parallel problem of the distribution of polynomial orbits on nilmanifolds. At frequencies near the scale of the distribution, a Lindeberg replacement scheme is used to replace increments of the walk with those of a continuous compactly supported density with the same heat kernel, thus reducing to Alexopoulos' theorem. It would be possible to make the replacement with increments of the heat kernel directly thus making the argument independent of [3], but the analysis then becomes technically more involved.

1.2 Historical review

Central limit theorems on Lie groups have a long history, with early theorems proven by Wehn [30] and Tutubalin [28], see also [16,26], and [15]. A central limit theorem with an optimal rate on stratified nilpotent groups is obtained under a homogeneous moment condition in [25].

Alexopoulos, and Alexopoulos and Lohoué have made a detailed study of convolution powers of continuous densities, heat kernels and related questions on Lie groups, see [1–6], and [7]. See [17–19] for work of Diaconis and Saloff-Coste in the finite setting.

A famous local limit theorem was proven by Bougerol [9] for convolutions of densities on a semi-simple group, which used the group's representation theory. There are still relatively few local limit theorems on non-abelian Lie groups that treat a measure which is not supported on a discrete group, or is not absolutely continuous with respect to Haar measure, of which [10,11] are an early example. Recently, Varjú has obtained such a local limit theorem for random walks on Euclidean space with transitions by a random isometry [29].

2 Notation and conventions

The connected, simply connected nilpotent Lie group of the theorems is G, with Lie algebra $\mathfrak{g}$ of dimension q. The lower central series of G is

$$G = G_1, \qquad G_{i+1} = [G, G_i], \ i \geqslant 1. \tag{5}$$

The Lie algebra of G_i is $\mathfrak{g}_i$. The Lie algebra $\mathfrak{g}$ is identified with $\mathbb{R}^q$ by choice of basis, which is fixed throughout the argument. Vectors $x \in \mathbb{R}^q$ are written in plain text and sequences of vectors $\underline{x} \in (\mathbb{R}^q)^N$ are written with an underscore. The norm $\|\cdot\|$ refers to the ℓ^2 norm on Euclidean space and is applied to $\mathfrak{g}$ and $\hat{\mathfrak{g}}$ by the fixed choice of basis. The character on $\mathbb{R}^q$ is written $e_\xi(x) = e^{2\pi i \xi \cdot x}$.

A bump function σ on $\mathbb{R}^q$ is a non-negative C^∞ function of compact support with integral 1. Its dilation by $t > 0$ is indicated $\sigma_t(x) = t^q \sigma(tx)$.

Convolution of Borel probability measures μ, ν on G is defined weakly by, for $f \in C_c(G)$,

$$\langle f, \mu * \nu \rangle = \int_G \int_G f(gh) d\mu(g) d\nu(h). \tag{6}$$

For $N \geqslant 1$, the iterated convolution μ^{*N} is defined by

$$\mu^{*1} = \mu, \qquad \mu^{*(i+1)} = \mu * \mu^{*i}, \ (i \geqslant 1). \tag{7}$$

The characteristic function of a probability measure ν on $\mathbb{R}^n$, respectively the Fourier transform of an L^1 function f, is defined to be

$$\hat{\nu}(\xi) = \int_{\mathbb{R}^n} e_{-\xi}(x) d\nu(x), \qquad \hat{f}(\xi) = \int_{\mathbb{R}^n} f(x) e_{-\xi}(x) dx. \tag{8}$$

If $f \in L^1$ is smooth, then the Fourier inversion integral is absolutely convergent, and

$$f(x) = \int_{\mathbb{R}^n} \hat{f}(\xi) e_\xi(x) d\xi. \tag{9}$$

Let $C_2 = \mathbb{Z}/2\mathbb{Z}$ be the group of two elements. For $\tau \in (C_2)^d$, $|\tau|$ is the Hamming or ℓ^1 norm, which counts the number of non-zero coordinates.

The notation $f = O_A(g)$ has the same meaning as $f \ll_A g$. Both indicate that $|f| \leqslant C(A)g$ for some constant $C(A) > 0$ which depends at most on A and possibly the structure of G.

3 Nilpotent Lie groups

A useful reference for the theory of nilpotent Lie groups is [14].

Given G, a connected, simply connected nilpotent Lie group with Lie algebra $\mathfrak{g}$ of dimension q, the exponential map exp is a diffeomorphism which identifies G with $\mathfrak{g}$. Given a probability measure μ on G, denote $\mu_{\mathfrak{g}}$ its push-forward by the logarithm map to a probability measure on the Lie algebra; $\mu_{\mathfrak{g}}^{*k}$ should be read $(\mu^{*k})_{\mathfrak{g}}$ so that convolution is performed on G.

Let the lower central series be defined by $\mathfrak{g}_1 = \mathfrak{g}$ and, for $i \geqslant 1$, $\mathfrak{g}_{i+1} = [\mathfrak{g}_1, \mathfrak{g}_i]$. Since $\mathfrak{g}$ is nilpotent, one has the filtration

$$\mathfrak{g} = \mathfrak{g}_1 \supset \mathfrak{g}_2 \supset \cdots \supset \mathfrak{g}_s \supset \mathfrak{g}_{s+1} = \{0\}, \qquad \mathfrak{g}_s \neq \{0\}; \tag{10}$$

s is called the step of $\mathfrak{g}$. One can check, for $i + j \leqslant s + 1$, $[\mathfrak{g}_i, \mathfrak{g}_j] \subset \mathfrak{g}_{i+j}$. Also, one has $G_i = \exp \mathfrak{g}_i$ is the ith group in the lower central series of G.

Let $\mathfrak{a}_1, \ldots, \mathfrak{a}_s$ be subspaces of $\mathfrak{g}$ such that, for each $1 \leqslant i \leqslant s$,

$$\mathfrak{g}_i = \mathfrak{a}_i \oplus \cdots \oplus \mathfrak{a}_s. \tag{11}$$

Let $d_i = \dim \mathfrak{a}_i$ and $q = \dim \mathfrak{g}$. The homogeneous dimension of G is

$$D = \sum_{i=1}^{s} i d_i. \tag{12}$$

Assume given a basis $\{X_{i,j}\}_{\substack{1 \leqslant i \leqslant s \\ 1 \leqslant j \leqslant d_i}}$ of $\mathfrak{g}$ satisfying $\{X_{i,j}\}_{1 \leqslant j \leqslant d_i}$ is a basis for $\mathfrak{a}_i$. Identify $\mathfrak{g}$ with $\mathbb{R}^q$ via

$$\psi : \begin{pmatrix} x^{(1,1)} \\ \vdots \\ x^{(s,d_s)} \end{pmatrix} \mapsto x^{(1,1)} X_{1,1} + \cdots + x^{(s,d_s)} X_{s,d_s}. \tag{13}$$

Having made this choice of coordinates, Haar measure on G is normalized by pushing forward Lebesgue measure on $\mathfrak{g}$ by the exponential map. Exponential coordinates of the first kind are defined by

$$\phi : \mathbb{R}^q \to G, \qquad \phi : \begin{pmatrix} x^{(1,1)} \\ \vdots \\ x^{(s,d_s)} \end{pmatrix} \mapsto \exp\left(\psi(x)\right). \tag{14}$$

Write for $1 \leqslant n \leqslant s$,

$$x^{(n)} = \begin{pmatrix} x^{(n,1)} \\ \vdots \\ x^{(n,d_n)} \end{pmatrix}, \qquad \xi^{(n)} = \begin{pmatrix} \xi^{(n,1)} \\ \vdots \\ \xi^{(n,d_n)} \end{pmatrix} \tag{15}$$

for the coordinates at level n in the filtration, respectively the corresponding dual frequencies in $\hat{\mathfrak{g}}$. These coordinates are said to have homogeneous degree n. For

$$S \subset \{(i, j) : 1 \leqslant i \leqslant s, 1 \leqslant j \leqslant d_i\}, \tag{16}$$

a monomial $m_\alpha(x) = \prod_{(i,j)\in S} \left(x^{(i,j)}\right)^{\alpha^{(i,j)}}$ with each $\alpha^{(i,j)} \neq 0$ has homogeneous degree

$$\text{hom-deg}(m_\alpha) = \sum_{(i,j)\in S} i\alpha^{(i,j)}. \tag{17}$$

The homogeneous degree of a monomial in several variables is defined to be the sum of the homogeneous degrees in the variables separately. Note that the homogeneous degree is equal to the degree if and only if for every $(i, j) \in S$, $i = 1$. A probability measure μ on G is said to have d finite homogeneous moments if, for all monomials m_α of homogeneous degree at most d,

$$\int_{\mathfrak{g}} |m_\alpha(x)| d\mu_{\mathfrak{g}}(x) < \infty. \tag{18}$$

3.1 Heat kernel and approximating continuous measure

Given a centered measure μ with two finite homogeneous moments on G, define the associated sub-Laplacian

$$L_\mu = -\sum_{1\leqslant i,j\leqslant d_1} a_{ij} X_{1,i} X_{1,j} - \sum_{i\leqslant d_2} a_i X_{2,i} \tag{19}$$

with coefficients

$$\begin{aligned} a_{ij} &= \frac{1}{2}\int x^{(1,i)} x^{(1,j)} d\mu_{\mathfrak{g}}(x), \\ b_i &= \int x^{(2,i)} d\mu_{\mathfrak{g}}(x) \end{aligned} \tag{20}$$

and

$$a_i = b_i - \frac{1}{2} \sum_{1 \leqslant j < k \leqslant d_1} a_{jk} \operatorname{pr}_{2,i}[X_{1,j}, X_{1,k}]. \tag{21}$$

Denote $u_t(x)$ the fundamental solution of the heat equation

$$\left(\frac{\partial}{\partial t} + L_\mu \right) u = 0, \qquad u_0 = \delta_{\mathrm{id}}. \tag{22}$$

The proof of Theorems 1 and 2 is a reduction to the local limit theorem of [3] which treats a continuous compactly supported probability density φ satisfying $\phi(e) > 0$.

Lemma 3 *There exists a function $\varphi = \varphi(\mu)$ satisfying*

(1) $\varphi \geqslant 0$ *and* $\int_G \varphi = 1$
(2) φ *is continuous*
(3) $\varphi(e) > 0$ *and* $\operatorname{supp}(\varphi) = \overline{\{x \in G : \varphi(x) > 0\}}$ *is compact*

which has first three homogeneous moments matching those of μ.

Proof Let $\varphi_{\mathfrak{g}}$ be the push-forward of φ to the Lie algebra, considered to be a continuous function on $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \cdots \times \mathbb{R}^{d_s}$. For the first three homogeneous moments to match, the conditions on $\varphi_{\mathfrak{g}}$ are as follows:

(1) Projected to $\mathbb{R}^{d_1}$, $\varphi_{\mathfrak{g}}$ is mean 0 and has first three mixed moments matching those of $\mu_{\mathfrak{g}}$
(2) On $\mathbb{R}^{d_2}$ the mean and correlation with the $\mathbb{R}^{d_1}$ coordinates are fixed to match those of $\mu_{\mathfrak{g}}$
(3) The mean on $\mathbb{R}^{d_3}$ is fixed.

Evidently the last condition may be fullfilled with a choice of $\varphi_{\mathfrak{g}}$ in which the coordinates at level 3 and higher in the filtration are independent of the first two levels of the filtration, so it suffices to consider $\varphi_{\mathfrak{g}}$ and $\mu_{\mathfrak{g}}$ as measures on $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$.

Note that $\mu_{\mathfrak{g}}$ restricted to $\mathbb{R}^{d_1}$ has a strictly positive definite covariance matrix, since otherwise $\mu_{\mathfrak{g}}$ would be supported on a subspace, and the Cramér condition would not hold. Thus, once the condition on the projection to $\mathbb{R}^{d_1}$ is satisfied, the condition on $\mathbb{R}^{d_2}$ may be attained by making the $\mathbb{R}^{d_2}$ coordinates the sum of two independent random variables $Y_1 + Y_2$ where Y_1 is a linear function of the $\mathbb{R}^{d_1}$ coordinates, and Y_2 is independent of those coordinates, is continuous, compactly supported with positive density at the identity, and with the appropriate mean.

Thus it suffices to consider $\mu_{\mathfrak{g}}$ and $\varphi_{\mathfrak{g}}$ as measures on $\mathbb{R}^{d_1}$. The first three mixed moment condition may be achieved by letting $\varphi_{\mathfrak{g}}$ be the sum of two independent random variables $X_1 + X_2$ where X_1 is a mean zero sum of finitely many point masses satisfying

(1) $\mathbf{Prob}(X_1 = 0) > 0$
(2) $\mathbf{Cov}(\mu_{\mathfrak{g}}) - \mathbf{Cov}(X_1)$ is positive definite
(3) The third mixed moments of $\mu_{\mathfrak{g}}$ match those of X_1

and where X_2 is a continuous density of compact support assigning positive mass to a neighborhood of the identity, which is even in the sense that all odd moments vanish, and such that $\mathbf{Cov}(X_1) + \mathbf{Cov}(X_2) = \mathbf{Cov}(\mu_{\mathfrak{g}})$. Evidently, $\mathbf{Cov}(X_1 + X_2) = \mathbf{Cov}(X_1) + \mathbf{Cov}(X_2)$ and the third mixed moments of $X_1 + X_2$ are equal to those of X_1 by independence, so that the third moment condition is fulfilled. □

Given φ as in Lemma 3, since the sub-Laplacian generated by μ and φ depends only on the first two moments of the abelianized measures and the mean in G_2/G_3, the heat kernels of μ and φ agree. By [3] Theorem 1.9.1, Theorem 1 holds with $\mu_\varphi = \varphi(g)dg$ in place of μ. The argument presented reduces the local limit theorem for μ to that for μ_φ.

3.2 The product rule

G is identified with $(\mathbb{R}^q, *)$ with the group law

$$\begin{pmatrix} x^{(1,1)} \\ \vdots \\ x^{(s,d_s)} \end{pmatrix} * \begin{pmatrix} y^{(1,1)} \\ \vdots \\ y^{(s,d_s)} \end{pmatrix} = \begin{pmatrix} z^{(1,1)} \\ \vdots \\ z^{(s,d_s)} \end{pmatrix} = \phi^{-1}\left(\phi \begin{pmatrix} x^{(1,1)} \\ \vdots \\ x^{(s,d_s)} \end{pmatrix} \cdot \phi \begin{pmatrix} y^{(1,1)} \\ \vdots \\ y^{(s,d_s)} \end{pmatrix} \right). \tag{23}$$

Given a sequence of vectors $\underline{x} \in (\mathbb{R}^q)^N$, write

$$\Pi(\underline{x}) = x_1 * x_2 * \cdots * x_N \in \mathbb{R}^q \tag{24}$$

for their product. The basic object of study is the characteristic function, for $\xi \in \hat{\mathfrak{g}} \cong \mathbb{R}^q$,

$$\chi_{N,\mu}(\xi) = \mathbf{E}_{\mu_{\mathfrak{g}}^{\otimes N}}\left[e_\xi(\Pi(\underline{x}))\right]. \tag{25}$$

In the case of a connected, simply connected nilpotent Lie group, the Baker–Campbell–Hausdorff formula is a finite expression that holds for all $X, Y \in \mathfrak{g}$,

$$\log(\exp X \exp Y) = X + Y + \frac{1}{2}[X, Y] + \frac{1}{12}[X, [X, Y]] - \frac{1}{12}[Y, [X, Y]] + \cdots. \tag{26}$$

Using the Baker–Campbell–Hausdorff formula, the product rule for a sequence of group elements may be expressed as a polynomial map on the Lie algebra. To describe this, given a sequence of elements $\underline{x} = (x_k)_{k=1}^N$ of elements from $\mathbb{R}^q$, sort a list of triples $\{(k_t, i_t, j_t)\}_{t=1}^\ell$, $1 \leqslant k_t \leqslant N$, $1 \leqslant i_t \leqslant s$, $1 \leqslant j_t \leqslant d_{i_t}$ lexicographically. Say that the monomial

$$m_\alpha(\underline{x}) = \prod_{t=1}^{\ell} \left(x_{k_t}^{(i_t, j_t)}\right)^{\alpha_{k_t}^{(i_t, j_t)}}, \tag{27}$$

is *initial* if it has the form

$$m_\alpha(\underline{x}) = \prod_{k=1}^{r} \prod_{(i,j) \in S_k} \left(x_k^{(i,j)}\right)^{\alpha_k^{(i,j)}} \tag{28}$$

where, for each k, S_k is a non-empty subset of $\{(i,\, j) : 1 \leqslant i \leqslant s, 1 \leqslant j \leqslant d_i\}$. If m_α is initial, say that monomial m'_α is of *type* m_α if for some $\ell_1 < \ell_2 < \cdots < \ell_r$,

$$m'_\alpha(\underline{x}) = \prod_{k=1}^{r} \prod_{(i,j)\in S_k} \left(x_{\ell_k}^{(i,j)}\right)^{\alpha_k^{(i,j)}}. \tag{29}$$

Thus m'_α is of type m_α if it is obtained by mapping the indices appearing in m_α forward by an increasing function.

Lemma 4 *Let $\underline{x} = (x_1, \ldots, x_N)$ be a sequence of vectors from $\mathbb{R}^q$ identified with coordinates on the Lie algebra $\mathfrak{g}$. There are polynomials $\left\{P_N^{(i,j)}\right\}_{\substack{1\leqslant i\leqslant s\\ 1\leqslant j\leqslant d_i}}$ on $(\mathbb{R}^q)^N$ satisfying the following conditions*

(1) *(Degree bound) Each monomial m_α in $P_N^{(i,j)}$ satisfies* $\text{hom-deg}(m_\alpha) \leqslant i$
(2) *(Stability) If m_α appears in $P_N^{(i,j)}$ and if $M > N$ then m_α appears in $P_M^{(i,j)}$ with the same leading coefficient*
(3) *(Invariance) If m'_α is of type m_α, and if the maximum index of m'_α is at most N, then m'_α appears in $P_N^{(i,j)}$ with the same leading coefficient as m_α*

such that the multiplication is given in coordinates by

$$\Pi(\underline{x}) = \begin{pmatrix} \sum_{k=1}^N x_k^{(1,1)} + P_N^{(1,1)}(\underline{x}) \\ \vdots \\ \sum_{k=1}^N x_k^{(s,d_s)} + P_N^{(s,d_s)}(\underline{x}) \end{pmatrix}. \tag{30}$$

Proof This follows from the Baker–Campbell–Hausdorff formula and induction. Write $\underline{x}'$ for the string $\underline{x}$ with x_N deleted, so that

$$\Pi(\underline{x}) = \Pi(\underline{x}') * x_N. \tag{31}$$

To obtain the degree bound, use that for $i + j \leqslant s + 1$, $[\mathfrak{g}_i, \mathfrak{g}_j] \subset \mathfrak{g}_{i+j}$, so that, when taking commutators, the step in the filtration always increases at least as quickly as the homogeneous degree of the coefficient. To obtain stability, note that if a contribution is made with a commutator involving $\psi(x_N)$ then the resulting monomial has an x_N dependence, so that monomials without an x_N dependence arise in $\Pi(\underline{x})$ only from the linear term in the Baker–Campbell–Hausdorff formula. To prove invariance, let m'_α be a monomial appearing in P_N which depends on x_N. Let the type of m'_α be m_α. Let $\tilde{\Pi}(\underline{x}')$ be obtained from $\Pi(\underline{x}')$ by setting to 0 all x_j that do not appear in m'_α and write the remaining indices in order $\ell_1 < \ell_2 < \cdots < \ell_{r-1}$. By induction, $\tilde{\Pi}(\underline{x}') = \Pi(x_{\ell_1}, \ldots, x_{\ell_{r-1}})$ and thus the coefficients of m_α and m'_α are equal. □

Let

$$m_\alpha = \prod_{k=1}^{r} \prod_{(i,j)\in S_k} \left(x_k^{(i,j)}\right)^{\alpha_k^{(i,j)}} \tag{32}$$

be an initial monomial of homogeneous degree n. Given $\underline{x} \in (\mathbb{R}^q)^N$, define the generalized U-statistic

$$U_\alpha(\underline{x}) = \sum_{1 \leqslant \ell_1 < \ell_2 < \cdots < \ell_r} \prod_{k=1}^{r} \prod_{(i,j) \in S_k} \left(x_{\ell_k}^{(i,j)} \right)^{\alpha_k^{(i,j)}}. \tag{33}$$

Lemma 4 may be summarized as stating that

$$\Pi^{(i,j)}(\underline{x}) = \sum_{k=1}^{N} x_k^{(i,j)} + P_N^{(i,j)}(\underline{x}) \tag{34}$$

where $P_N^{(i,j)}$ is a linear combination of generalized U-statistics of homogeneous degree at most i, with the ℓ^1 norm of the coefficients in the linear combination bounded by a constant depending on the fixed choice of basis.

3.3 Manipulations regarding the test function

The test function of the theorem takes the form, for $x \in \mathbb{R}^q \cong \mathfrak{g}$,

$$L_g R_h f(\phi(x)) = f_{\mathfrak{g}}(\log g * x * \log h). \tag{35}$$

Applying the Baker–Campbell–Hausdorff formula, there are polynomials $p_{g,h}$ and $q_{g,h}$ satisfying for $1 \leqslant n \leqslant s$, $p_{g,h}^{(n)}, q_{g,h}^{(n)}$ are of homogeneous degree at most n, such that

$$x' = \log g * x * \log h = p_{g,h}(x), \qquad x = q_{g,h}(x'). \tag{36}$$

The relationship between $p_{g,h}^{(n)}$ and $q_{g,h}^{(n)}$ is linear in $x^{(n)}$ and $x'^{(n)}$ and polynomial in the lower degree coordinates. In particular, p can be obtained from q by a polynomial change, and vice-versa, see [21], Appendix A for a further discussion.

Define the (naive) height $\mathrm{ht}(p)$ of a polynomial p to be the sup norm on its coefficients. In particular,

$$\mathrm{ht}(p_{g,h}) \ll_D 1 + \mathrm{ht}(q_{g,h})^{O_s(1)}, \qquad \mathrm{ht}(q_{g,h}) \ll_D 1 + \mathrm{ht}(p_{g,h})^{O_s(1)}. \tag{37}$$

Let $\sigma \in C_c^\infty(\mathbb{R}^q)$ be a smooth bump function with dilation, for $t > 0$, $\sigma_t(x) = t^q \sigma(tx)$. Let $f_{\mathfrak{g},t} = f_{\mathfrak{g}} * \sigma_t$ be the Euclidean convolution. Since f is assumed Lipschitz,

$$\left\| f_{\mathfrak{g}} - f_{\mathfrak{g},t} \right\|_\infty = O\left(\frac{1}{t} \right) \tag{38}$$

as $t \to \infty$. Also, $\left\| f_{\mathfrak{g},t} \right\|_1 \leqslant \left\| f_{\mathfrak{g}} \right\|_1$.

Lemma 5 *For each $n \geqslant 1$ and for $\xi \in \mathbb{R}^q$, $\|\xi\| \geqslant 1$ and $t > 1$, the Fourier transform*

$$\widehat{L_g R_h f_{\mathfrak{g},t}}(\xi) = \int_{\mathbb{R}^q} f_{\mathfrak{g},t}(\log g * x * \log h) e_{-\xi}(x) dx \tag{39}$$

satisfies

$$\left|\widehat{L_g R_h f_{\mathfrak{g},t}}(\xi)\right| \leqslant O_{n,D,f}(1)\left(1 + \mathrm{ht}(p_{g,h})\right)^{O_s(1)} \left(\frac{t}{\|\xi\|}\right)^n \|f\|_1. \tag{40}$$

Proof Let $\xi_0 = \frac{\xi}{\|\xi\|}$ and integrate by parts n times in the ξ_0 direction to obtain

$$\widehat{L_g R_h f_{\mathfrak{g},t}}(\xi) = \left(\frac{1}{2\pi i \|\xi\|}\right)^n \int_{\mathbb{R}^q} D_{\xi_0}^n \left[f_{\mathfrak{g},t}(\log g * x * \log h)\right] e_{-\xi}(x) dx. \tag{41}$$

Write $\log g * x * \log h = p_{g,h}(x)$ and note that for $0 \leqslant j \leqslant n$,

$$\left|D_{\xi_0}^j p_{g,h}(x)\right| \leqslant O_{n,D}(1)\,\mathrm{ht}(p_{g,h})\|x\|^{s-j} \tag{42}$$

By the compact support of $f_{\mathfrak{g}}$, restrict to $x' = p_{g,h}(x) \in \mathrm{supp}\, f$ which is $O_f(1)$. Thus

$$\|x\| = \left\|q_{g,h}(x')\right\| = O_{D,f}(\mathrm{ht}(q_{g,h})) = O_{D,f}\left((1 + \mathrm{ht}(p_{g,h}))^{O_s(1)}\right). \tag{43}$$

Meanwhile $D_{\xi_0}^j f_{\mathfrak{g},t} = f_{\mathfrak{g}} * D_{\xi_0}^j \sigma_t$, and $\left\|D_{\xi_0}^j \sigma_t\right\|_1 \ll_j t^j$. Hence $\left\|D_{\xi_0}^j f_{\mathfrak{g},t}\right\|_1 \ll_j t^j \|f\|_1$. The conclusion now follows on applying the chain rule and bounding the integral in L^1. □

The following lemma based on [13] Theorem 2 is used to restrict the translations g, h in Theorem 1 to those for which $p_{g,h}$ has controlled height.

Lemma 6 *Let $p : \mathbb{R}^q \to \mathbb{R}^q$ be a polynomial of degree at most s. There is a constant $C = C(q,s) > 0$ such that, for any $\alpha > 0$,*

$$\mathrm{meas}\left\{x \in \left[-\frac{1}{2}, \frac{1}{2}\right]^q : \|p(x)\| \leqslant \alpha\right\} \leqslant \frac{C\alpha^{\frac{1}{s}}}{\mathrm{ht}(p)^{\frac{1}{s}}}. \tag{44}$$

Proof The statement

$$\mathrm{meas}\left\{x \in \left[-\frac{1}{2}, \frac{1}{2}\right]^q : \|p(x)\| \leqslant \alpha\right\} \leqslant \frac{C\alpha^{\frac{1}{s}}}{\left(\int_{\left[-\frac{1}{2},\frac{1}{2}\right]^q} \|p\|^2\right)^{\frac{1}{2s}}}. \tag{45}$$

is a specialization of [13] Theorem 2. The conclusion follows since all norms on the space of degree s polynomials on $\mathbb{R}^q$ are equivalent. □

4 Rearrangement group action

In [20] Diaconis and the author used the following group action on strings. The group C_2^{n-1} acts on strings of length kn with the jth factor exchanging the relative order of the $j+1$st block of length k with the previous jk elements. For instance, in the case $n=4$, if $x_1,\ldots,x_4$ each represent a block of k indices, the action is illustrated in

$$\begin{aligned}(1,0,0)\cdot\underline{x} &= x_2x_1x_3x_4\\ (1,1,0)\cdot\underline{x} &= x_3x_2x_1x_4\\ (0,1,1)\cdot\underline{x} &= x_4x_3x_1x_2.\end{aligned} \tag{46}$$

The relative order within the segments of length k in each x_i remains unchanged.

For $n\geqslant 2$, $k\geqslant 1$ and $1\leqslant N'\leqslant \left\lfloor \frac{N}{kn}\right\rfloor$ let

$$A_{k,n}^{N'}=\left(C_2^{n-1}\right)^{N'} \tag{47}$$

act on strings of length knN' with, for $j\geqslant 1$, the jth factor of C_2^{n-1} in $A_{k,n}^{N'}$ acting as above on the contiguous subsequence of indices of length kn ending at jkn. The argument below considers $A_{k,n}^{N'}$ acting on substrings of length knN' within a string of length N.

4.1 The Gowers–Cauchy–Schwarz inequality

Given two elements

$$\underline{\tau}_0,\underline{\tau}_1\in A_{k,n}^{N'}=\left(C_2^{n-1}\right)^{N'}=\left(C_2^{N'}\right)^{n-1} \tag{48}$$

write $\underline{\tau}=\left(\tau^{(1)},\ldots,\tau^{(n-1)}\right)$. Thus $\tau_{0,i}^{(j)}$ is the element in C_2 which belongs to the ith factor of (C_2^{n-1}) in $A_{k,n}^{N'}$ and within this factor, the jth factor of C_2. Given $s\in\{0,1\}^{n-1}\cong C_2^{n-1}$, define $\underline{\tau}_s=\left(\tau_{s_1}^{(1)},\ldots,\tau_{s_{n-1}}^{(n-1)}\right)\in A_{k,n}^{N'}$.

Since the increments of $\mu_{\mathfrak{g}}$ in the characteristic function $\chi_{N,\mu}$ are i.i.d., a further averaging may be introduced in which the group $A_{k,n}^{N'}$ acts on a substring of the product measure. In general, let $P(\underline{x})$ be a continuous function of $\underline{x}$ and let its characteristic function be

$$\chi(\xi)=\mathbf{E}_{\mu_{\mathfrak{g}}^{\otimes N}}\left[e_\xi(P(\underline{x}))\right]. \tag{49}$$

Then

$$\chi(\xi)=\mathbf{E}_{\mu_{\mathfrak{g}}^{\otimes N}}\left[\mathbf{E}_{\underline{\tau}\in A_{k,n}^{N'}}\left[e_\xi\left(P(\underline{\tau}\cdot\underline{x})\right)\right]\right]. \tag{50}$$

Writing $\underline{\tau} = (\underline{\tau}^{(1)}, \underline{\tau}^{(2)}, \ldots, \underline{\tau}^{(n-1)})$,

$$\chi(\xi) = \mathbf{E}_{\mu_{\mathfrak{g}}^{\otimes N}} \left[\mathbf{E}_{\underline{\tau}^{(1)} \in C_2^{N'}} \cdots \mathbf{E}_{\underline{\tau}^{(n-1)} \in C_2^{N'}} \left[e_\xi \left(P(\underline{\tau} \cdot \underline{x}) \right) \right] \right]. \tag{51}$$

Moving one $\underline{\tau}^{(i)}$ to the inside at a time and applying Cauchy–Schwarz to the inner expectation recovers the Gowers–Cauchy–Schwarz inequality:

$$|\chi(\xi)|^{2^{n-1}} \leqslant \mathbf{E}_{\mu_{\mathfrak{g}}^{\otimes N}} \left[\mathbf{E}_{\underline{\tau}_0, \underline{\tau}_1 \in A_{k,n}^{N'}} \left[e_\xi \left(\sum_{s \in \{0,1\}^{n-1}} (-1)^{|s|} P(\underline{\tau}_s \cdot \underline{x}) \right) \right] \right]. \tag{52}$$

In the case $\chi(\xi) = \chi_{N,\mu}(\xi)$, denote the right hand side $F\left(\xi, \mu; A_{k,n}^{N'}\right)$.

A basic lemma, which generalizes Lemma 24 of [20], is as follows.

Lemma 7 *Let $N, N' \geqslant 1$, let $k \geqslant 1$ and $n \geqslant 2$ be such that $knN' \leqslant N$. Let $\underline{x} \in (\mathbb{R}^q)^N$ and let $A_{k,n}^{N'}$ act on the substring of $\underline{x}$ with indices in range, for some offset $o \geqslant 0$, $[o+1, o+knN']$. Define for $1 \leqslant i \leqslant nN'$,*

$$\omega_i = \sum_{j=(i-1)k+1}^{ik} x_{o+j}. \tag{53}$$

For any $\underline{\tau}_0, \underline{\tau}_1 \in A_{k,n}^{N'}$ the summation formula holds,

$$\begin{aligned} &\sum_{s \subset \{0,1\}^{n-1}} (-1)^{|s|} \Pi^{(n)}(\underline{\tau}_s \cdot \underline{x}) \\ &= \sum_{i=1}^{N'} \left(\sum_{s \in \{0,1\}^{n-1}} (-1)^{|s|} \Pi^{(n)}(\tau_{s,i} \cdot (\omega_{n(i-1)+1}, \ldots, \omega_{ni})) \right) \end{aligned} \tag{54}$$

while for all $n' < n$,

$$\sum_{s \subset \{0,1\}^{n-1}} (-1)^{|s|} \Pi^{(n')}(\underline{\tau}_s \cdot \underline{x}) = 0. \tag{55}$$

Moreover,

$$\begin{aligned} &\sum_{s \in \{0,1\}^{n-1}} (-1)^{|s|} \Pi^{(n)}(\tau_{s,i} \cdot (\omega_1, \ldots, \omega_n)) \\ &= \begin{cases} (-1)^{|\tau_{0,i}|} \sum_{\tau \in C_2^{n-1}} (-1)^{|\tau|} \Pi^{(n)}(\tau \cdot (\omega_1, \ldots, \omega_n)) & \tau_{0,i} + \tau_{1,i} = (1)^{n-1} \\ 0 & \textit{otherwise} \end{cases}. \end{aligned} \tag{56}$$

In the $(1)^{n-1}$ *case, the sum is a vector whose coordinates are non-zero multilinear polynomials in* $\left(\omega_1^{(1)}, \omega_2^{(1)}, \ldots, \omega_n^{(1)}\right)$.

Proof By the degree bound, all monomials appearing in $\Pi^{(n,j)}$ have homogeneous degree at most n, and hence degree at most n. Given any collection of $m \leqslant n$ indices $k_1 < k_2 < \cdots < k_m$, if there is any bit $b \in C_2^{n-1}$ of Hamming weight 1 such that $\underline{\tau}_s$ and $\underline{\tau}_{s+b}$ act on $k_1, \ldots, k_m$ leaving them in the same relative order, then by the invariance principle any monomials associated to these indices in the alternating sum

$$\sum_{s \subset \{0,1\}^{n-1}} (-1)^{|s|} \Pi^{(n,j)}(\underline{\tau}_s \cdot \underline{x}) \tag{57}$$

cancel. In particular, this occurs if $m < n$, or if the indices $k_1 < k_2 < \cdots < k_m$ are not acted on by the same factor of C_2^{n-1} in $A_{k,n}^{N'}$, or if $k_1, k_2, \ldots, k_n$ do not appear in distinct blocks in the action, or if the corresponding factor of $\underline{\tau}_0$ and $\underline{\tau}_1$ do not add to the all 1's element. In particular this proves (55).

Since the only surviving monomials have degree n and homogeneous degree n, the surviving variables are all from the first level of the filtration $x^{(1)}$ and all of the monomials are linear in each variable. By the invariance principle, the surviving polynomial is in fact a polynomial on the sums $\omega_1, \omega_2, \ldots, \omega_n$.

The formula (56) is immediate. To prove that the $(1)^{n-1}$ case of (56) is non-vanishing, in the case $k = 1$ let $g_i = \exp(\omega_i) \in G$. The sum

$$\sum_{\tau \in C_2^{n-1}} (-1)^{|\tau|} \Pi^{(n,j)} (\tau \cdot (\omega_1, \ldots, \omega_n)) \tag{58}$$

is equal to the $X^{(n,j)}$ coordinate in the logarithm of the iterated commutator

$$[\cdots[[[g_1, g_2], g_3], g_4], \ldots, g_n]. \tag{59}$$

To verify this by induction, note that the commutator may be calculated in G_n/G_{n+1}, which is abelian, and depends only on g_n in G_1/G_2, so that the calculation may be performed by truncating the Baker–Campbell–Hausdorff formula at the first commutator.

Since commutators of the type (59) generate G_n, it follows that (58) is non-zero. □

Given probability measure μ on G, let μ_n be the probability measure on G_n/G_{n+1} with distribution

$$[\cdots[[g_1, g_2], g_3], \ldots, g_n], \qquad g_i \text{ i.i.d. } \mu. \tag{60}$$

Thus $\mu_1 = \mu_{\text{ab}}$, and for $n \geqslant 2$, μ_n has distribution given by

$$\sum_{\tau \in C_2^{n-1}} (-1)^{|\tau|} \Pi^{(n)} (\tau \cdot (\omega_1, \ldots, \omega_n)) \tag{61}$$

in which the ω_i are drawn i.i.d. from $\mu_{\mathfrak{g}}$. Given $\xi^{(n)} \in \widehat{G_n/G_{n+1}} \cong \widehat{\mathfrak{g}_n/\mathfrak{g}_{n+1}}$ denote the characteristic function of μ_n by

$$F_{n,\mu}\left(\xi^{(n)}\right) = \hat{\mu}_n\left(\xi^{(n)}\right). \tag{62}$$

Lemma 8 *Let $2 \leqslant n \leqslant s$, $k, N' \geqslant 1$, and let $N \geqslant knN'$. Let $A_{k,n}^{N'}$ act on a substring of $\underline{x} \in (\mathbb{R}^q)^N$ as above. Let $\xi \in \hat{\mathfrak{g}}$ satisfy $\xi^{(j)} = 0$ for all $j > n$. Then*

$$F\left(\xi, \mu; A_{k,n}^{N'}\right) = \left(1 - \frac{1}{2^{n-1}} + \frac{\mathrm{Re}\left[F_{n,\mu^{*k}}\left(\xi^{(n)}\right)\right]}{2^{n-1}}\right)^{N'}. \tag{63}$$

Proof The expectation factors through the product structure of $A_{k,n}^{N'}$ due to the summation formula (54). In the expectation over $A_{k,n}^{N'}$ the probability that $\tau_{0,j} + \tau_{1,j} = (1)^{n-1}$ is $\frac{1}{2^{n-1}}$, and conditioned on this, the expectation over the corresponding block is $(-1)^{|\tau_{0,i}|} F_{n,\mu^{*k}}\left(\xi^{(n)}\right)$. The real part occurs since conditionally, $|\tau_{0,i}|$ has parity 0 and 1 with equal probability. □

4.2 The Cramér condition

A probability measure ν on $\mathbb{R}^m$ has characteristic function $\hat{\nu}$ satisfying the Cramér condition if there exists $0 < \epsilon < 1$ such that

$$\sup_{\xi \in \mathbb{R}^m, \|\xi\| > 1} |\hat{\nu}(\xi)| \leqslant 1 - \epsilon. \tag{64}$$

The condition is equivalent to the statement, for all $r > 0$ there exists $0 < \epsilon(r) < 1$ such that

$$\sup_{\xi \in \mathbb{R}^m, \|\xi\| > r} |\hat{\nu}(\xi)| \leqslant 1 - \epsilon(r). \tag{65}$$

The equivalence may be checked by noting $1 - |\hat{\nu}(\xi_1 + \xi_2)| \leqslant 2(2 - |\hat{\nu}(\xi_1)| - |\hat{\nu}(\xi_2)|)$, see [27] p. 183, where the proof does not use that the probability measure has finite support.

Lemma 9 *Let μ be a centered probability measure on G satisfying $G = \overline{\langle \mathrm{supp}\, \mu \rangle}$, whose abelianization μ_{ab} has characteristic function satisfying the Cramér condition. For each $2 \leqslant n \leqslant s$ the measure μ_n on G_n/G_{n+1} has characteristic function satisfying the Cramér condition.*

Proof Write

$$[\cdots[[g_1, g_2], g_3], \ldots, g_n] \bmod G_{n+1} = \langle \lambda(g_1, g_2, \ldots, g_{n-1}), g_n \rangle \tag{66}$$

in which $\lambda(g_1, g_2, \ldots, g_{n-1})$ is a linear map from G_{ab} to G_n/G_{n+1}. Recall that λ itself is multilinear in $g_1, g_2, \ldots, g_{n-1} \bmod G_2$. Given $\xi \in \widehat{G_n/G_{n+1}}$, $\|\xi\| \geqslant 1$ one has

$$\xi\left([\cdots[[g_1,g_2],g_3],\ldots,g_n]\right)=\left(\xi\cdot\lambda(g_1,g_2,\ldots,g_{n-1})\right)(g_n)\,. \tag{67}$$

Since the semigroup generated by $\operatorname{supp}\mu_{\mathrm{ab}}$ is dense in G_{ab}, and since $G_n=[G_{n-1},G_1]$ is equal to the set of commutators of the stated type on G, and since λ is multilinear, it follows that ξ does not annihilate $\lambda(g_1,\ldots,g_{n-1})$ with positive probability, and hence for some $r>0$, $\|\xi\cdot\lambda(g_1,g_2,\ldots,g_{n-1})\|>r$ with positive probability. Integrating, this suffices to obtain the Cramér condition. □

In the case that $\hat{\mu}_{\mathrm{ab}}$ satisfies the non-lattice condition, for all $0\neq\xi\in\hat{G}_{\mathrm{ab}}$, $\left|\hat{\mu}_{\mathrm{ab}}(\xi)\right|<1$, the same proof yields that $\hat{\mu}_n$ is non-lattice.

Lemma 10 *There is a constant $c=c(\mu)>0$ such that, for each $1\leqslant n\leqslant s$, for all $\xi^{(n)}\neq 0$, when k is assigned by the rule*

$$k=\begin{cases}\left\lfloor\frac{1}{\left\|\xi^{(n)}\right\|^{\frac{2}{n}}}\right\rfloor, & \left\|\xi^{(n)}\right\|\leqslant 1\\ 1, & \left\|\xi^{(n)}\right\|>1\end{cases}, \tag{68}$$

one has $\left|F_{n,\mu^{*k}}\left(\xi^{(n)}\right)\right|\leqslant 1-c$.

Proof For any fixed $r>0$, for $\left\|\xi^{(n)}\right\|\geqslant r$ this follows from the Cramér condition. Otherwise, using the description (56), it follows from the functional central limit theorem that when $\underline{x}$ is drawn from $(\mu^{*k})^{\otimes n}$,

$$\frac{1}{k^{\frac{n}{2}}}\left(\sum_{\tau\in C_2^{n-1}}(-1)^{|\tau|}\tau\cdot\Pi^{(n)}(\underline{x})\right) \tag{69}$$

converges to a continuous probability density. Since $\left\|\xi^{(n)}\right\|\asymp\frac{1}{k^{\frac{n}{2}}}$, the claim follows. □

In the case that $\hat{\mu}_{\mathrm{ab}}$ is non-lattice, the same conclusion holds when $\xi^{(n)}$ is restricted to any compact set, with a constant c which possibly depends upon the compact set.

5 Estimates of moments

Throughout this section μ is a centered probability measure on G.

Lemma 11 *Let $m,n,N\geqslant 1$ and suppose that μ has $2mn$ finite homogeneous moments. For all generalized U-statistics U_α of homogeneous degree n,*

$$\mathbf{E}_{\mu_{\mathfrak{g}}^{\otimes N}}\left[\left|U_\alpha(\underline{x})\right|^{2m}\right]\leqslant O_{\mu,mn}(1)N^{mn}. \tag{70}$$

Proof Let a be the number of indices in m_α of homogeneous degree 1, and note that $n\geqslant a+2(r-a)=2r-a$. On expanding $\left|U_\alpha(\underline{x})\right|^{2m}$ and performing expectation,

any monomials that have indices which appear with homogeneous degree 1 have expectation 0. Those remaining monomials have homogeneous degree at least 2 in every coordinate upon which the expectation depends, and hence have expectation $O_{\mu,mn}(1)$ by the moment condition. The total number of indices which may appear in such a monomial is at most $(2m)(r-a)+am \leqslant nm$. Counting the number of monomials with non-vanishing expectation by letting L be the number of indices appearing, the expectation is bounded by

$$\mathbf{E}\left[\left|U_\alpha(\underline{x})\right|^{2m}\right] \leqslant O_{\mu,mn}(1)\sum_{L=1}^{nm}\binom{N}{L}L^{2rm} \leqslant O_{\mu,mn}(1)N^{nm}. \tag{71}$$

□

Lemma 12 *For each $A,\delta>0$ there is $C(A,\delta)>0$ such that, if μ has $C(A,\delta)$ finite homogeneous moments then*

$$\mu_{\mathfrak{g}}^{\otimes N}\left\{\max_n \frac{1}{N^{\frac{n}{2}}}\left\|\Pi^{(n)}(\underline{x})\right\| > N^{\delta}\right\} = O_{\mu,A,\delta}\left(N^{-A}\right). \tag{72}$$

Proof If μ has $2mn$ homogeneous moments then the estimate

$$\mathbf{E}_{\mu_{\mathfrak{g}}^{\otimes N}}\left[\left\|\Pi^{(n)}(\underline{x})\right\|^{2m}\right] \leqslant O_{\mu,mn}(1)N^{mn} \tag{73}$$

follows by repeatedly applying the power mean inequality to first estimate $\mathbf{E}[\|\Pi^{(n)}\|^{2m}]$ in terms of moments of the individual coordinates $\mathbf{E}[\|\Pi^{(n,j)}(\underline{x})\|^{2m}]$ and then in terms of the moments of individual U statistics of homogeneous degree at most n, to which Lemma 11 applies.

The claim now follows by taking a high enough moment and applying Markov's inequality. □

Lemma 13 *Let $m \geqslant 1$ and $1 \leqslant n \leqslant s$, and assume that μ has $2mn$ finite homogeneous moments. For all $N' \leqslant N$, when $\underline{x} = \underline{x}_0 \oplus \underline{x}_t$ is the concatenation of strings of length N' and $N-N'$*

$$\mathbf{E}_{\mu_{\mathfrak{g}}^{\otimes N}}\left[\left\|\Pi^{(n)}(\underline{x}) - \Pi^{(n)}(\underline{x}_t)\right\|^{2m}\right] \leqslant O_{\mu,mn}(1)N^{mn}\left(\frac{N'}{N}\right)^m. \tag{74}$$

Proof By repeatedly applying the power mean inequality it suffices to prove, for any generalized U-statistic

$$U_\alpha(\underline{x}) = \sum_{1\leqslant \ell_1<\ell_2<\cdots<\ell_r}\prod_{k=1}^{r}\prod_{(i,j)\in S_k}\left(x_{\ell_k}^{(i,j)}\right)^{\alpha_k^{(i,j)}} \tag{75}$$

of homogeneous degree at most n, the estimate

$$\mathbf{E}_{\mu_{\mathfrak{g}}^{\otimes N}}\left[\left|U_\alpha(\underline{x})-U_\alpha(\underline{x}_t)\right|^{2m}\right]\leqslant O_{\mu,mn}(1)N^{mn}\left(\frac{N'}{N}\right)^m. \tag{76}$$

Define polynomials, for $1\leqslant a\leqslant r$, $U_\alpha^{0,a}$, $U_\alpha^{t,a}$,

$$\begin{aligned}U_\alpha^{0,a}(\underline{x})&=\sum_{1\leqslant\ell_1<\ell_2<\cdots<\ell_a}\prod_{k=1}^{a}\prod_{(i,j)\in S_k}\left(x_{\ell_k}^{(i,j)}\right)^{\alpha_k^{(i,j)}}\\ U_\alpha^{t,a}(\underline{x})&=\sum_{1\leqslant\ell_1<\ell_2<\cdots<\ell_a}\prod_{k=1}^{a}\prod_{(i,j)\in S_{r-a+k}}\left(x_{\ell_k}^{(i,j)}\right)^{\alpha_k^{(i,j)}},\end{aligned} \tag{77}$$

and also, make the convention that $U_\alpha^{0,0}=U_\alpha^{t,0}=1$. Hence,

$$U_\alpha(\underline{x})-U_\alpha(\underline{x}_t)=\sum_{a=1}^{r}U_\alpha^{0,a}(\underline{x}_0)U_\alpha^{t,r-a}(\underline{x}_t). \tag{78}$$

Applying the power mean inequality one further time, it suffices to prove the estimate, for each $1\leqslant a\leqslant r$,

$$\mathbf{E}_{\mu_{\mathfrak{g}}^{\otimes N}}\left[\left|U_\alpha^{0,a}(\underline{x}_0)U_\alpha^{t,r-a}(\underline{x}_t)\right|^{2m}\right]\leqslant O_{\mu,mn}(1)N^{mn}\left(\frac{N'}{N}\right)^m. \tag{79}$$

Since

$$\mathbf{E}_{\mu_{\mathfrak{g}}^{\otimes N}}\left[\left|U_\alpha^{0,a}(\underline{x}_0)U_\alpha^{t,r-a}(\underline{x}_t)\right|^{2m}\right]=\mathbf{E}_{\mu_{\mathfrak{g}}^{\otimes N'}}\left[\left|U_\alpha^{0,a}(\underline{x}_0)\right|^{2m}\right]\mathbf{E}_{\mu_{\mathfrak{g}}^{\otimes(N-N')}}\left[\left|U_\alpha^{t,r-a}(\underline{x}_t)\right|^{2m}\right] \tag{80}$$

the claim follows from Lemma 11, since, for each $1\leqslant a\leqslant r$,

$$\text{hom-deg}(U_\alpha^{0,a})+\text{hom-deg}(U_\alpha^{t,r-a})\leqslant n,\qquad \text{hom-deg}(U_\alpha^{t,r-a})\leqslant n-1. \tag{81}$$

□

Denote $\Pi_j(\underline{x})$ the part of $\Pi(\underline{x})$ which depends on x_j. Set

$$\Pi_j^{\leqslant 3}(\underline{x}),\qquad \Pi_j^{>3}(\underline{x}) \tag{82}$$

the part of $\Pi_j(\underline{x})$ which is of homogeneous degree $\leqslant 3$, respectively >3 in x_j, and for $d=1,2,3$, $\Pi_j^d(\underline{x})$ the part of $\Pi_j(\underline{x})$ which is of homogeneous degree d in x_j. $\Pi_j^{*,(n)}$ denotes the part of Π_j^* at level n. Use the same notation with Π replaced with a U-statistic U_α.

Recall that μ_φ is the measure of the continuous density φ.

Lemma 14 *Assume that μ has at least $6s$ homogeneous moments. Let $N \geqslant 1$, $1 \leqslant j \leqslant N$ and $\underline{\mu} = \mu^{\otimes(j-1)} \otimes \mu_\varphi^{\otimes(N-j+1)}$ or $\underline{\mu} = \mu^{\otimes j} \otimes \mu_\varphi^{\otimes(N-j)}$. For $k = 1, 2, 3$, and $m \leqslant 3$,*

$$\mathbf{E}_{\underline{\mu}}\left[\left|\xi \cdot \Pi_j^k(\underline{x})\right|^{2m}\right] = O_\mu\left(\sum_{\ell=k}^{s} \|\xi_\ell\|^{2m} N^{(\ell-k)m}\right) \tag{83}$$

and

$$\mathbf{E}_{\underline{\mu}}\left[\left|\xi \cdot \Pi_j^{>3}(\underline{x})\right|^{2}\right] = O_\mu\left(\sum_{\ell=4}^{s} \|\xi_\ell\|^{2} N^{\ell-4}\right). \tag{84}$$

Proof By the power mean inequality, then Cauchy–Schwarz,

$$\begin{aligned}\mathbf{E}_{\underline{\mu}}\left[\left|\xi \cdot \Pi_j^k(\underline{x})\right|^{2m}\right] &\ll_s \sum_{\ell=k}^{s} \mathbf{E}_{\underline{\mu}}\left[\left|\xi^{(\ell)} \cdot \Pi_j^{k,(\ell)}(\underline{x})\right|^{2m}\right] \\ &\leqslant \sum_{\ell=k}^{s} \left\|\xi^{(\ell)}\right\|^{2m} \mathbf{E}_{\underline{\mu}}\left[\left\|\Pi_j^{k,(\ell)}(\underline{x})\right\|^{2m}\right].\end{aligned} \tag{85}$$

Applying the power mean inequality several further times to first replace $\Pi_j^{k,(\ell)}$ with its individual coordinates, then with an individual generalized U-statistic reduces to proving the bound for a U-statistic U_α of homogeneous degree $\ell \geqslant k$,

$$\mathbf{E}_{\underline{\mu}}\left[\left|U_{\alpha,j}^k(\underline{x})\right|^{2m}\right] = O_\mu\left(N^{(\ell-k)m}\right). \tag{86}$$

Let $\underline{x}_0$, $\underline{x}_t$ denote the substrings of $\underline{x}$ prior to j and after j respectively. The claim follows from Lemma 11 after factoring

$$U_{\alpha,j}^k(\underline{x}) = U_1(\underline{x}_0) m(x_j) U_2(\underline{x}_t) \tag{87}$$

where m is a monomial of homogeneous degree k and U_1 and U_2 are U-statistics satisfying $\hom\text{-}\deg(U_1) + \hom\text{-}\deg(U_2) = \ell - k$.

The proof of (84) is similar. □

6 Proof of Theorems

The following lemma is used to truncate in frequency space to the scale of the distribution.

Lemma 15 *Let $N \geqslant 1$, $A > 0$ and let $1 > \epsilon_1 > \epsilon_2 > \cdots > \epsilon_s > \epsilon_{s+1} = 0$ be a collection of constants satisfying for all $1 \leqslant n < s$, $\epsilon_n > n\epsilon_{n+1}$. Suppose that μ has $C(A, \epsilon)$ finite homogeneous moments for some constant $C(A, \epsilon) > 0$. If*

$$\max\left\{\left\|\xi^{(n)}\right\| N^{\frac{n}{2}-\epsilon_n} : 1 \leqslant n \leqslant s\right\} > 1, \tag{88}$$

then $|\chi_{N,\mu}(\xi)| = O_{\mu,A,\epsilon}\left(N^{-A}\right)$.

Proof Let n be maximal such that $\left\|\xi^{(n)}\right\| N^{\frac{n}{2}-\epsilon_n} > 1$. If $n = s$, set $N' = N$, otherwise, set $N' = \left\lfloor N^{1-\frac{\epsilon_n}{n}-\epsilon_{n+1}} \right\rfloor$. Let $\underline{x}_0$ and $\underline{x}_t$ be strings of vectors from $\mathbb{R}^q$ of lengths N' and $N - N'$ and let $\underline{x} = \underline{x}_0 \oplus \underline{x}_t$ be the concatenation. Let

$$\Xi_{n+1}(\underline{x}) = \sum_{j=n+1}^{s} \xi^{(j)} \cdot \Pi^{(j)}(\underline{x}). \tag{89}$$

Denote $T_{2m-1}(x) = \sum_{j=0}^{2m-1} \frac{(2\pi i x)^j}{j!}$ the degree $2m-1$ Taylor expansion of $e^{2\pi i x}$ and recall that Taylor's theorem with remainder gives

$$\left| T_{2m-1}(x) - e^{2\pi i x} \right| \leqslant \frac{(2\pi x)^{2m}}{(2m)!}. \tag{90}$$

It follows that

$$\begin{aligned}
&\chi_{N,\mu}(\xi) \\
&= \mathbf{E}_{\mu_{\mathfrak{g}}^{\otimes N}}\left[\prod_{j=1}^{n} e_{\xi^{(j)}}(\Pi^{(j)}(\underline{x})) \prod_{\ell=n+1}^{s} e_{\xi^{(\ell)}}(\Pi^{(\ell)}(\underline{x}_t)) T_{2m-1}\left(\Xi_{n+1}(\underline{x}) - \Xi_{n+1}(\underline{x}_t)\right) \right] \\
&\quad + O_m\left(\mathbf{E}_{\mu_{\mathfrak{g}}^{\otimes N}}\left[\left| \Xi_{n+1}(\underline{x}) - \Xi_{n+1}(\underline{x}_t) \right|^{2m} \right] \right).
\end{aligned} \tag{91}$$

By Lemma 13 and Hölder's inequality, if μ has sufficiently many homogeneous moments,

$$\begin{aligned}
\mathbf{E}_{\mu_{\mathfrak{g}}^{\otimes N}}\left[\left| \Xi_{n+1}(\underline{x}) - \Xi_{n+1}(\underline{x}_t) \right|^{2m} \right] &\ll_{\mu,m} \sum_{j=n+1}^{s} \left\| \xi^{(j)} \right\|_{2m}^{2m} N^{m(j-1)} (N')^m \\
&\leqslant \sum_{j=n+1}^{s} \left(N^{-\frac{j}{2}+\epsilon_j} \right)^{2m} N^{m(j-1)} (N')^m \\
&\leqslant \sum_{j=n+1}^{s} N^{m\left(2\epsilon_j - \frac{\epsilon_n}{n} - \epsilon_{n+1}\right)}.
\end{aligned} \tag{92}$$

Since each exponent is negative, the sum may be made $O_{\mu,A,\epsilon}\left(N^{-A}\right)$ by choosing m sufficiently large in terms of A and ϵ.

Expand $T_{2m-1}\left(\Xi_{n+1}(\underline{x}) - \Xi_{n+1}(\underline{x}_t)\right)$ into monomials of degree bounded by $(2m-1)s$ with coefficients of ℓ^1 norm bounded by $\ll_B N^B$. Set $N_1 = \left\lfloor \frac{N'}{2ms} \right\rfloor$. Given a typical monomial M, let $[J+1, J+N_1]$ be a set of indices from $[1, N']$ which does not meet M.

Define,

$$k = \begin{cases} \left\lfloor \frac{1}{\|\xi^{(n)}\|^{\frac{2}{n}}} \right\rfloor, & \|\xi^{(n)}\| \leqslant 1 \\ 1, & \|\xi^{(n)}\| > 1 \end{cases}, \qquad N_1' = \left\lfloor \frac{N_1}{kn} \right\rfloor. \tag{93}$$

Let $A_{k,n}^{N_1'} = (C_2^{n-1})^{N_1'}$ act on the substring $[J+1, J+N_1]$ as described in Sect. 4. Since the monomial M is invariant under the group action, its contribution to the expectation (91) is given by

$$\mathbf{E}_{\underline{x}_t \sim \mu_{\mathfrak{g}}^{\otimes(N-N')}} \left[M \prod_{j=n+1}^{s} e_{\xi^{(j)}}(\Pi^{(j)}(\underline{x}_t)) \mathbf{E}_{\underline{x}_0 \sim \mu_{\mathfrak{g}}^{\otimes N'}} \left[\mathbf{E}_{\underline{\tau} \in A_{k,n}} \left[\prod_{j=1}^{n} e_{\xi^{(j)}}(\Pi^{(j)}(\underline{\tau} \cdot \underline{x})) \right] \right] \right]. \tag{94}$$

By Cauchy–Schwarz,

$$\begin{aligned} |(94)|^2 \leqslant & \mathbf{E}_{\underline{x}_t \sim \mu_{\mathfrak{g}}^{\otimes(N-N')}} \left[|M|^2 \right] \\ & \times \mathbf{E}_{\underline{x}_t \sim \mu_{\mathfrak{g}}^{\otimes(N-N')}} \left[\left| \mathbf{E}_{\underline{x}_0 \sim \mu_{\mathfrak{g}}^{\otimes N'}} \left[\mathbf{E}_{\underline{\tau} \in A_{k,n}} \left[\prod_{j=1}^{n} e_{\xi^{(j)}}(\Pi^{(j)}(\underline{\tau} \cdot \underline{x})) \right] \right] \right|^2 \right]. \end{aligned}$$

Bound the first expectation by a constant. In the case $n \geqslant 2$, apply Gowers–Cauchy–Schwarz to bound the second expectation, using Lemma 8 to evaluate the expectation that results. In either the case $n = 1$ or $n \geqslant 2$, it follows from Lemma 10 that

$$|(94)|^{2^{n-1}} \ll_\mu \left(1 - \frac{1}{2^{n-1}} + \frac{\left| F_{n,\mu_{\mathfrak{g}}^{*k}}(\xi^{(n)}) \right|}{2^{n-1}} \right)^{N_1'} \ll_\mu \exp\left(-C N_1' \right). \tag{95}$$

Since

$$N_1' \gg \frac{N_1}{k} \gg \min\left(1, \left\| \xi^{(n)} \right\|^{\frac{2}{n}} \right) N^{1-\frac{\epsilon_n}{n}-\epsilon_{n+1}} \gg N^{\frac{\epsilon_n}{n}-\epsilon_{n+1}} \tag{96}$$

tends to infinity with N like a small power of N, the exponential savings dominates the polynomial bound on the coefficients of the monomials, which proves the lemma. □

Recall that $\varphi = \varphi(\mu)$ has been fixed as in Lemma 3, such that φ is a continuous, compactly supported density, $\varphi(e) > 0$ with first three homogeneous moments matching those of μ. The following Lindeberg exchange lemma approximates the distribution of μ^{*N} with that of μ_φ^{*N}.

Lemma 16 *Let $N \geqslant 1$. Let $1 > \epsilon_1 > \epsilon_2 > \cdots > \epsilon_s > \epsilon_{s+1} = 0$ be a collection of constants satisfying for all $1 \leqslant n < s$, $\epsilon_n > n\epsilon_{n+1}$. Assume that*

$$\max\left\{ \left\| \xi^{(n)} \right\| N^{\frac{n}{2}-\epsilon_n} : 1 \leqslant n \leqslant s \right\} \leqslant 1. \tag{97}$$

Then

$$\left|\chi_{N,\mu}(\xi) - \chi_{N,\mu_\varphi}(\xi)\right| = O_\mu\left(N^{-1+O(\epsilon_1)}\right). \tag{98}$$

Proof Define

$$\Delta_j = \mathbf{E}_{\mu_{\mathfrak{g}}^{\otimes j}\otimes\mu_{\varphi,\mathfrak{g}}^{\otimes(N-j)}}\left[e_\xi(\Pi(\underline{x}))\right] - \mathbf{E}_{\mu_{\mathfrak{g}}^{\otimes(j-1)}\otimes\mu_{\varphi,\mathfrak{g}}^{\otimes(N-j+1)}}\left[e_\xi(\Pi(\underline{x}))\right] \tag{99}$$

so that, by the triangle inequality,

$$\left|\chi_{N,\mu}(\xi) - \chi_{N,\mu_\varphi}(\xi)\right| \leqslant \sum_{j=1}^{N} |\Delta_j|. \tag{100}$$

For $1 \leqslant j \leqslant N$ bound, moving expectation against x_j to the inside and using the triangle inequality,

$$|\Delta_j| \leqslant \mathbf{E}_{\mu_{\mathfrak{g}}^{\otimes(j-1)}\otimes\mu_{\varphi,\mathfrak{g}}^{\otimes(N-j)}}\left[\left|\int_{\mathfrak{g}} e_\xi(\Pi_j(\underline{x}))d\mu_{\mathfrak{g}}(x_j) - \int_{\mathfrak{g}} e_\xi(\Pi_j(\underline{x}))\varphi_{\mathfrak{g}}(x_j)dx_j\right|\right]. \tag{101}$$

Using $|e(x) - e(y)| \leqslant 2\pi|x - y|$ and the triangle inequality, the right hand side is bounded by a constant times

$$\begin{aligned}&\mathbf{E}_{\mu_{\mathfrak{g}}^{\otimes j}\otimes\mu_{\varphi,\mathfrak{g}}^{\otimes(N-j)}}\left[\left|\xi\cdot\Pi_j^{>3}(\underline{x})\right|\right] + \mathbf{E}_{\mu_{\mathfrak{g}}^{\otimes(j-1)}\otimes\mu_{\varphi,\mathfrak{g}}^{\otimes(N-j+1)}}\left[\left|\xi\cdot\Pi_j^{>3}(\underline{x})\right|\right]\\ &\quad + \mathbf{E}_{\mu_{\mathfrak{g}}^{\otimes(j-1)}\otimes\mu_{\varphi,\mathfrak{g}}^{\otimes(N-j)}}\left[\left|\int_{\mathfrak{g}} e_\xi\left(\Pi_j^{\leqslant 3}(\underline{x})\right)d\mu_{\mathfrak{g}}(x_j) - \int_{\mathfrak{g}} e_\xi\left(\Pi_j^{\leqslant 3}(\underline{x})\right)\varphi_{\mathfrak{g}}(x_j)dx_j\right|\right].\end{aligned} \tag{102}$$

Note that $\Pi_j^{(n),>3}$ is of homogeneous degree $\leqslant n-4$ in the variables other than x_j, while $\|\xi^{(n)}\| \leqslant N^{-\frac{n}{2}+\epsilon_n}$. Thus, by Cauchy–Schwarz and (84) of Lemma 14, the top line is $O_\mu\left(N^{-2+O(\epsilon_1)}\right)$.

By Taylor expansion,

$$\begin{aligned}e_\xi\left(\Pi_j^{\leqslant 3}(\underline{x})\right) &= 1 + i2\pi\xi\cdot\Pi_j^1(\underline{x}) - \frac{1}{2}\left(2\pi\xi\cdot\Pi_j^1(\underline{x})\right)^2 - \frac{i}{6}\left(2\pi\xi\cdot\Pi_j^1(\underline{x})\right)^3\\ &\quad + i2\pi\xi\cdot\Pi_j^2(\underline{x}) - 4\pi^2(\xi\cdot\Pi_j^1(\underline{x}))(\xi\cdot\Pi_j^2(\underline{x})) + i2\pi\xi\cdot\Pi_j^3(\underline{x})\\ &\quad + O\left(\left|\xi\cdot\Pi_j^1(\underline{x})\right|^4 + \left(1+|\xi\cdot\Pi_j^1(\underline{x})|^3\right)\left(|\xi\cdot\Pi_j^2(\underline{x})|^2 + |\xi\cdot\Pi_j^3(\underline{x})|^2\right)\right).\end{aligned} \tag{103}$$

Since the main term has homogeneous degree at most 3 in x_j, and since the first three homogeneous moments of μ and μ_φ agree, the integral of these terms cancel. In the error term, separate $|\xi\cdot\Pi_j^1(\underline{x})|^3$ from $|\xi\cdot\Pi_j^2(\underline{x})|^2$ and $|\xi\cdot\Pi_j^3(\underline{x})|^2$ with Cauchy–Schwarz. Now applying (83) of Lemma 14, the error term is bounded in expectation by $O_\mu\left(N^{-2+O(\epsilon_1)}\right)$ as before. □

Proof of Theorem 1 Let f be the Lipschitz, compactly supported test function of the theorem, let $f_{\mathfrak{g}}$ be the push-forward by the logarithm map to the Lie algebra, and identify $f_{\mathfrak{g}}$ as a Lipschitz function on $\mathbb{R}^q$. By [3] Theorem 1.9.1, integrating the pointwise approximation to the heat kernel u_N,

$$\left\langle L_g R_h f, \mu_\varphi^{*N} \right\rangle = \left\langle L_g R_h f, u_N \right\rangle + O\left(\|f\|_1 N^{-\frac{D+1}{2}} \right). \tag{104}$$

Thus it suffices to show that

$$\left| \left\langle L_g R_h f, \mu^{*N} \right\rangle - \left\langle L_g R_h f, \mu_\varphi^{*N} \right\rangle \right| = O_\mu\left(\|f\|_1 N^{-\frac{D+1}{2}} \right) + O_{\mu,A,f}\left(N^{-A} \right). \tag{105}$$

As in Sect. 3.3, let $p_{g,h}$ and $q_{g,h}$ be polynomials such that

$$x' = \log g * x * \log h = p_{g,h}(x), \qquad x = q_{g,h}(x'). \tag{106}$$

Thus

$$\begin{aligned} \left\langle L_g R_h f, \mu^{*N} \right\rangle &= \int_G f(gxh) d\mu^{*N}(x) \\ &= \int_{\mathfrak{g}^N} f_{\mathfrak{g}}\left(\log g * \Pi(\underline{x}) * \log h \right) d\mu_{\mathfrak{g}}^{\otimes N} = \int_{\mathfrak{g}^N} f_{\mathfrak{g}}(p_{g,h}(\Pi(\underline{x}))) d\mu_{\mathfrak{g}}^{\otimes N}. \end{aligned} \tag{107}$$

First consider the case that $\mathrm{ht}(q_{g,h}) \geqslant N^C$ for a constant C depending at most on D, and A. Let for some $B > 0$, $\mathrm{supp}\, f_{\mathfrak{g}} \subset \left[-\frac{B}{2}, \frac{B}{2}\right]^q$, let $\delta > 0$ and let

$$S_{\mathrm{small}} = \left\{ x' \in \left[-\frac{B}{2}, \frac{B}{2}\right]^q : \max_n \left\{ \left\| q_{g,h}^{(n)}(x') \right\| N^{-\frac{n}{2}} \right\} \leqslant N^\delta \right\}. \tag{108}$$

If C is sufficiently large then Lemma 6 implies that

$$\mathrm{meas}(S_{\mathrm{small}}) = O_{A,D,f}\left(N^{-A} \right), \tag{109}$$

with the dependence on f arising from dilating the box $\left[-\frac{B}{2}, \frac{B}{2}\right]^q$ to $\left[-\frac{1}{2}, \frac{1}{2}\right]^q$. Let $f = f_1 + f_2$, with $f_{\mathfrak{g},1} = f_{\mathfrak{g}}|_{S_{\mathrm{small}}}$. Use

$$\|f_1\|_1 \leqslant \|f\|_\infty \,\mathrm{meas}(S_{\mathrm{small}}) \tag{110}$$

and hence

$$\langle L_g R_h f_1, \mu^{*N} \rangle,\ \langle L_g R_h f_1, \mu_\varphi^{*N} \rangle = O_{A,f}\left(N^{-A} \right). \tag{111}$$

Meanwhile, by Lemma 12,

$$\mu^{\otimes N}\left\{ \max_n \left\{ \left\| \Pi^{(n)}(x) \right\| N^{-\frac{n}{2}} \right\} > N^\delta \right\} = O_{\mu,A}\left(N^{-A} \right), \tag{112}$$

and similarly for μ_φ. Since $x' \in \operatorname{supp} f \backslash S_{\text{small}}$ implies that

$$\max_n \left\{ \left\| x^{(n)} \right\| N^{-\frac{n}{2}} \right\} > N^{\delta} \tag{113}$$

it follows that

$$\langle L_g R_h f_2, \mu^{*N} \rangle, \ \langle L_g R_h f_2, \mu_\varphi^{*N} \rangle = O_{\mu,A,f}\left(N^{-A}\right). \tag{114}$$

Together (111) and (114) imply Theorem 1 in this case.

Now suppose that $\operatorname{ht}(q_{g,h}) \leqslant N^C$ so that $\operatorname{ht}(p_{g,h}) \ll N^{C'}$ for some $C' > 0$. Let σ be a compactly supported bump function on $\mathbb{R}^q$ with dilation, for $t > 0, \sigma_t(x) = t^q \sigma(tx)$. Let $f_{\mathfrak{g},t} = f_{\mathfrak{g}} * \sigma_t$ be the Euclidean convolution. Choose $t \gg \|f\|_1^{-1} N^{\frac{D+1}{2}}$ so that $\left\| f_{\mathfrak{g}} - f_{\mathfrak{g},t} \right\|_\infty \ll \|f\|_1 N^{-\frac{D+1}{2}}$. It thus suffices to prove (105) with f replaced by f_t.

Expand, using the Fourier transform,

$$\begin{aligned} \left\langle L_g R_h f_t, \mu^{*N} \right\rangle &= \int_{\mathfrak{g}^N} f_{\mathfrak{g},t} \left(\log g * \Pi(\underline{x}) * \log h \right) d\mu_{\mathfrak{g}}^{\otimes N}(\underline{x}) \\ &= \int_{\hat{\mathfrak{g}}} \widehat{L_g R_h f}_{\mathfrak{g},t}(\xi) \mathbf{E}_{\mu_{\mathfrak{g}}^{\otimes N}} \left[e_\xi(\Pi(\underline{x})) \right] d\xi. \end{aligned} \tag{115}$$

Since the test function $f_{\mathfrak{g},t}$ is smooth, the integral converges absolutely.

Let $\epsilon_1 > \epsilon_2 > \cdots > \epsilon_s > 0$ be a collection of constants as in Lemmas 15 and 16. Define

$$E_{\text{small}} = \left\{ \xi \in \hat{\mathfrak{g}} : \max \left\{ \left\| \xi^{(n)} \right\| N^{\frac{n}{2} - \epsilon_n} : 1 \leqslant n \leqslant s \right\} \leqslant 1 \right\}. \tag{116}$$

Thus $\operatorname{meas}(E_{\text{small}}) \ll N^{-\frac{D}{2} + O(\epsilon_1)}$. Apply Lemma 15, and Lemma 5 with $n = q + 2$, to obtain

$$\begin{aligned} &\left| \int_{E_{\text{small}}^c} \widehat{L_g R_h f}_{\mathfrak{g},t}(\xi) \chi_{N,\mu}(\xi) d\xi \right| \\ &\leqslant \left\| \chi_{N,\mu}|_{E_{\text{small}}^c} \right\|_\infty \left\| \widehat{L_g R_h f}_{\mathfrak{g},t}|_{E_{\text{small}}^c} \right\|_1 = O_{\mu,A,\epsilon,f}\left(N^{-A}\right). \end{aligned} \tag{117}$$

Thus

$$\left\langle L_g R_h f_t, \mu^{*N} \right\rangle = O_{\mu,A,\epsilon,f}\left(N^{-A}\right) + \int_{E_{\text{small}}} \widehat{L_g R_h f}_{\mathfrak{g},t}(\xi) \chi_{N,\mu}(\xi) d\xi, \tag{118}$$

and similarly with μ replaced by μ_φ.

On the remainder of the integral, apply Lemma 16 to obtain

$$\begin{aligned} &\left| \int_{E_{\text{small}}} \widehat{L_g R_h f}_{\mathfrak{g},t}(\xi) \left(\chi_{N,\mu}(\xi) - \chi_{N,\mu_\varphi}(\xi) \right) d\xi \right| \\ &\leqslant \|f\|_1 \left\| \left(\chi_{N,\mu}(\xi) - \chi_{N,\mu_\varphi}(\xi) \right)|_{E_{\text{small}}} \right\|_\infty \operatorname{meas}(E_{\text{small}}) \end{aligned}$$

$$\ll_\mu \|f\|_1 N^{-\frac{D+2}{2}+O(\epsilon_1)}. \tag{119}$$

Choose ϵ_1 sufficiently small but fixed so that the error term is $O_\mu\left(\|f\|_1 N^{-\frac{D+1}{2}}\right)$, which proves Theorem 1 in the remaining case.

Evidently the argument presented requires only finitely many moments of the measure μ, but how many? To gain convergence in the Fourier integral (117) it was necessary to integrate by parts $n = q+2$ times, which costs a factor of $\mathrm{ht}(p_{g,h})^{O_{s,q}(1)}$. Hence the number of moments depends on A, the dimension q and the step s, and hence is controlled by A and the homogeneous dimension D. □

To obtain Theorem 2, approximate $f_{\mathfrak{g}}$ from above and below in L^1 by functions having Fourier transforms of compact support. The argument in the small height case then goes through as before.

References

1. Alexopoulos, G.: Parabolic Harnack inequalities and Riesz transforms on Lie groups of polynomial growth. Harmonic Analysis (Sendai, 1990), pp. 1–14, ICM-90 Satell, Conf. Proc, Springer, Tokyo (1991)
2. Alexopoulos, G.: An application of homogenization theory to harmonic analysis on solvable Lie groups of polynomial growth. Pac. J. Math. **159**(1), 19–45 (1993)
3. Alexopoulos, G.K.: Centered densities on Lie groups of polynomial volume growth. Probab. Theory Relat. Fields **124**(1), 112–150 (2002)
4. Alexopoulos, G.K.: Random walks on discrete groups of polynomial volume growth. Ann. Probab. **30**(2), 723–801 (2002)
5. Alexopoulos, G.K.: Sub-Laplacians with drift on Lie groups of polynomial volume growth. Mem. Am. Math. Soc. **155**(739) (2002)
6. Alexopoulos, G., Lohoué, N.: On the large time behavior of heat kernels on Lie groups. Duke Math. J. **120**(2), 311–351 (2003)
7. Alexopoulos, G., Lohoué, N.: On the asymptotic behavior of convolution powers and heat kernels on semisimple Lie groups. C. R. Math. Acad. Sci. Paris **350**(5–6), 255–257 (2012)
8. Bayer, D., Diaconis, P.: Trailing the dovetail shuffle to its Lair. Ann. Appl. Probab. **2**, 294–313 (1992)
9. Bougerol, P.: Théoreme central limite local sur certains groupes de Lie. Ann. Sci. Ecole Norm. Supérieure **14**(4), 403–432 (1981)
10. Breuillard, E.F.: Equidistribution of Random walks on Nilpotent Lie Groups and Homogeneous Spaces. ProQuest LLC, Ann Arbor (2004)
11. Breuillard, E.: Local limit theorems and equidistribution of random walks on the Heisenberg group. Geom. Funct. Anal. **15**(1), 35–82 (2005)
12. Breuillard, E.: Equidistribution of dense subgroups on nilpotent Lie groups. Ergod. Theory Dyn. Syst. **30**(1), 131–150 (2010)
13. Carbery, A., Wright, J.: Distributional and L^q norm inequalities for polynomials over convex bodies in $\mathbb{R}^n$. Math. Res. Lett. **8**(3), 233–248 (2001)
14. Corwin, L.J., Greenleaf, F.P.: Representations of Nilpotent Lie Groups and Their Applications. Part I. Basic Theory and Examples. Cambridge Studies in Advanced Mathematics, vol. 18. Cambridge University Press, Cambridge (1990)
15. Coulhon, T., Saloff-Coste, L., Varopoulos, N.: Analysis and Geometry on Groups. Cambridge Tracts on Mathematics. Cambridge University Press, Cambridge (1992)
16. Crépel, P., Raugi, A.: Théorème central limite sur les groupes nilpotents. Ann. Inst. H. Poincaré Sect. B Probab. Stat. **24**(2), 145–164 (1978)
17. Diaconis, P., Saloff-Coste, L.: Moderate growth and random walk on finite groups. Geom. Funct. Anal. **4**(1), 1–36 (1994)

18. Diaconis, P., Saloff-Coste, L.: An application of Harnack inequalities to random walk on nilpotent quotients. J. Fourier Anal. Appl. Spec. Issue 189–207 (1995)
19. Diaconis, P., Saloff-Coste, L.: Nash inequalities for finite Markov chains. J. Theor. Probab. **9**(2), 459–510 (1996)
20. Diaconis, P., Hough, R.: Random walk on unipotent matrix groups. Preprint arXiv:1512.06304
21. Green, B., Tao, T.: The quantitative behaviour of polynomial orbits on nilmanifolds. Ann. Math. **175**(2), 465–540 (2012). (2)
22. Hough, R.: Mixing and cut-off in cycle walks. Electron. J. Probab. **22**(90), 1–49 (2017)
23. Hough, R., Jerison, D., Levine, L.: Sandpiles on the square lattice. Commun. Math. Phys. (to appear)
24. Lawler, G.F., Limic, V.: Random Walk: A Modern Introduction, vol. 123. Cambridge University Press, Cambridge (2010)
25. Pap, G.: Rate of convergence in CLT on stratified groups. J. Multivar. Anal. **38**(2), 333–365 (1991)
26. Stroock, D.W., Varadhan, S.R.S.: Limit theorems for random walks on Lie groups. Sankhyā Ser. A **35**(3), 277–294 (1973)
27. Tao, T., Van Vu, H.: Additive Combinatorics, vol. 105. Cambridge University Press, Cambridge (2006)
28. Tutubalin, V.N.: Compositions of measures on the simplest nilpotent group. Teor. Verojatnost. i Primenen. **9**, 531–539 (1964). (Russian)
29. Varjú, P.P.: Random walks in Euclidean space. Ann. Math. **181**(1), 243–301 (2015). (2)
30. Wehn, D.: Probabilities on Lie groups. Proc. Nat. Acad. Sci. USA **48**, 791–795 (1962)