

THE CENTRAL LIMIT THEOREM ON NILPOTENT LIE GROUPS

BY TIMOTHÉE BÉNARD^{1,a} AND EMMANUEL BREUILLARD^{2,b}

¹Centre for Mathematical Sciences, University of Cambridge, ^atb723@cam.ac.uk

²Mathematical Institute, University of Oxford, ^bbreuillard@maths.ox.ac.uk

We formulate and establish the central limit theorem for products of i.i.d. random variables on arbitrary simply connected nilpotent Lie groups, allowing a possible bias. We find that some interesting new phenomena arise in the presence of a bias: the walk spreads out at a higher rate in the ambient group, while the limiting hypoelliptic diffusion process may not always have full support. We use elementary Fourier analysis to establish our results which include Berry-Esseen bounds under optimal moment assumptions, as well as an analogue of Donsker's invariance principle. Various examples of nilpotent Lie groups are treated in detail showing the variety of different behaviours. We also obtain a characterization of when the limiting distribution is an ordinary gaussian and answer a question of Tutubalin regarding asymptotically close distributions on nilpotent Lie groups.

1. Introduction. Given a Lie group G and a probability measure μ on G , it is a natural problem to investigate the large scale geometry of the convolution powers μ^{*N} for large N , or, in other words, the distribution of the product $X_1 \cdots X_N$ of N independent random variables in G with law μ . We refer the reader to the survey [SC01] for a pleasant introduction to the topic, as well as to the books [Gre63, VSCC92, Lia04, BQ16, Var21].

In the abelian case $G = \mathbb{R}^d$, the answer is provided by the classical central limit theorem: provided $\mathbb{E}(X_1^2) < \infty$, we have the convergence in law:

$$\frac{1}{\sqrt{N}}(X_1 + \cdots + X_N - N\chi) \xrightarrow{N \rightarrow +\infty} \mathcal{N}(0, \sigma^2),$$

where χ and σ^2 denote the expectation and covariance matrix of X_1 respectively, and $\mathcal{N}(0, \sigma^2)$ is the centered Gaussian distribution on \mathbb{R}^d with covariance matrix σ^2 .

What is meant by a central limit theorem in a general Lie group G is less clear, because there rarely is a canonical way to renormalize the random walk. See [Gre63], [ST66, p. 3], [Vir74], [Bre04, Chapter 1] for a discussion. However, if the Lie group G is nilpotent and simply connected, then G is diffeomorphic to its Lie algebra \mathfrak{g} via the exponential map [CG90] and, although there may not exist any dilating automorphisms [Dye70, DL57], there are natural ways to renormalize the space, leading to meaningful limit theorems.

In this paper we shall be interested in the central limit theorem for i.i.d. random walks on an arbitrary simply connected nilpotent Lie group G . These are the simply connected real Lie groups G whose central descending series $G^{[i+1]} := [G, G^{[i]}]$ terminates at the trivial group in a finite number of steps. They can alternatively be described as the connected subgroups of unipotent upper triangular matrices [CG90, Theorem 1.1.1].

In his early work on the subject, Tutubalin [Tut64] handled successfully the case where G is the Heisenberg group. It is the simplest instance of a non-commutative nilpotent Lie group.

MSC2020 subject classifications: 60B15, 60F05, 35H10, 35R03, 22E25, 58J65.

Keywords and phrases: Nilpotent Lie groups, biased random walks, central limit theorem, hypoelliptic diffusion process.

EXAMPLE (Heisenberg group). Let G be the classical 3-dimensional Heisenberg group, defined as $\mathbb{R}^3 = \bigoplus_{i=1}^3 \mathbb{R}e_i$ endowed with the group structure

$$x * y = (x_1 + y_1, x_2 + y_2, x_3 + y_3 + \frac{1}{2}(x_1y_2 - x_2y_1)).$$

Let μ be a probability measure on G satisfying the moment conditions $\mathbb{E}_\mu(|x_i|^2) < \infty$ for $i = 1, 2$ and $\mathbb{E}_\mu(|x_3|) < \infty$. The mean of μ in the abelianization is identified with $X_\mu := \mathbb{E}_\mu(x_1e_1 + x_2e_2)$. Let $r > 0$. If $X_\mu = 0$, set $D_r : x \mapsto (x_1r, x_2r, x_3r^2)$, while if $X_\mu \neq 0$ set $D_r : x \mapsto (x_1r, x_2r, x_3r^3)$.

In the centered case $X_\mu = 0$, Tutubalin [Tut64] proves that the renormalized distribution $D_{\frac{1}{\sqrt{N}}}(\mu^{*N})$ converges to a probability measure on G , which is absolutely continuous with respect to Lebesgue. In fact, it is the law of the value at time $t = 1$ of a hypoelliptic diffusion $(B_t^{(1)}, B_t^{(2)}, L_t)$ on G , where $(B_t^{(1)}, B_t^{(2)})$ is a planar Brownian motion and L_t is its Lévy area (see e.g. [FV10]).

On the other hand, if $X_\mu \neq 0$, Tutubalin [Tut64] shows that the recentered and renormalized distribution $D_{\frac{1}{\sqrt{N}}}(\mu^{*N} * \delta_{-NX_\mu})$, i.e. the distribution of $D_{\frac{1}{\sqrt{N}}}(x * (-NX_\mu))$ when $x \sim \mu^{*N}$, converges towards an ordinary gaussian (normal) distribution on \mathbb{R}^3 . Notice that $\det D_r = r^3$ in the centered case, while $\det D_r = r^5$ in the non-centered case. This means that *the random walk “spreads out” more in space in the presence of a non-trivial bias.*

For an arbitrary simply connected nilpotent Lie group G , the central limit theorem (CLT) for *centered* random walks, i.e. for measures μ whose projection to the abelianization $G/[G, G]$ has zero mean, is by now a classical result and several proofs are known. The case where G is stratified follows easily from Wehn’s infinitesimal CLT [Weh62] or from Stroock-Varadhan [SV73], while the general case is handled by Crépel and Raugi in [CR78]. See also [CCLP99] for a more recent treatment of the centered CLT for stratified nilpotent Lie groups. Berry-Esseen estimates have also been derived by Pap [Pap91] and Bentkus-Pap [BP96] in the stratified centered case. Last but not least, in the case where the driving measure μ has a continuous density, Alexopoulos [Ale02a, Ale02c] obtained very complete results for the centered CLT on an arbitrary nilpotent Lie group including Berry-Esseen bounds and even a full Edgeworth expansion.

Our main goal in this paper will be to formulate and establish the general case of the CLT and the Berry-Esseen estimates, see Theorems 1.1 1.6, allowing the group to be non-stratified and most importantly the walk to be *non-centered* (we also say “biased”). We shall make no assumption of regularity on the measure μ , only necessary moment assumptions. We will also prove an analogue of Donsker’s invariance principle in our context (Theorem 1.7), establishing that the stochastic process obtained by interpolating the random walk does converge in law towards a limiting diffusion process in the topology of Hölder continuous paths (and even in the “rough paths” topology), see Section 5.3.

Unlike the case of commutative groups, *the analysis of non-centered walks does not reduce to that of centered walks.* It is true that certain questions for non-centered walks can be solved by reducing to the centered case, for example the behavior of μ^{*n} in a neighborhood of the origin: as Alexopoulos observes in [Ale02a, 1.11], $d\mu^{*n}(x) = e^{-\beta_\mu n} \chi_\mu(x) d\mu_0^{*n}(x)$ where μ_0 is a centered probability measure, $\beta_\mu > 0$, and χ_μ is a well chosen character of the group. This however yields no information on the kind of questions we are considering in this paper, namely the behavior of μ^{*n} around its average, where most of the mass is located. A delicate issue in establishing a central limit theorem for biased walks is to understand precisely how to renormalize the random walk after recentering, in order to obtain an absolutely continuous probability measure in the limit. The large scale geometry of the distribution μ^{*N} depends in a subtle way on the increment average and attempts to find a suitable renormalization

have been made by Virtser [Vir74] for the full upper-triangular unipotent matrix group, then by Raugi [Rau78] for general nilpotent Lie groups. However neither formulation is fully satisfactory: Virtser’s renormalization is too strong, leading in some cases to non-absolutely continuous distributions in the limit, whereas Raugi’s renormalization may lead to full escape of mass in the limit (see the end of Section 5.1 for a detailed discussion). We shall show the right way to renormalize non-centered random walks on nilpotent Lie groups, so that the limiting distribution is always a smooth probability measure.

We also present a detailed analysis of the limiting measure. In particular, we give decay estimates at infinity (see Proposition 3.8) proving that the limit density is a rapidly decreasing Schwartz function similarly to what is known in the centered case [Heb89, Var88, Sie84]. It came as a surprise to us to discover that a new phenomenon takes place in the biased case, and only in step 3 or more: due to certain inaccessibility constraints [Sus87, AS04, Sac07] appearing in the presence of a bias, *the limiting measure may not have full support* in general and the identity can lie on the boundary of the support, even though it is absolutely continuous. This phenomenon occurs in all free nilpotent Lie groups of step at least 3 whenever there is a non-zero bias (see Theorem 1.2). On the other hand there are Lie groups, such as the upper triangular unipotent groups, where it never happens and the limit measure always has full support (see Theorem 1.3). For such groups we will also establish a lower bound on the density, by exploiting a version of Harnack’s inequality for hypoelliptic operators recently established in [KP16]. Yet there are also examples, such as filiform Lie algebras of dimension at least 4, where both phenomena occur depending on the increment average. Typically the limiting measure is far from being an ordinary Gaussian distribution, but we shall also characterize exactly when this is the case (Theorem 1.4).

In the context of hypoelliptic heat kernel estimates and diffusion processes on manifolds, the consequences of a non-trivial drift for the support and the small time estimates have been studied in several influential papers, starting with those of Ben Arous and Léandre [BeLe91] (see also [AKS93] and more recently [CHT21]). In these papers the drift refers to a first order term in the infinitesimal generator of the diffusion. In our context, the non-centeredness assumption on the driving measure is not reflected by a first order term in the infinitesimal generator of the limiting diffusion. Rather, it forces the generator to be time dependent. This time dependency plays a crucial role here, especially as the generator obtained for a frozen time t is not hypoelliptic. Below, we will present examples of limiting measures with non-full support and also no drift term at all in the (hypoelliptic) infinitesimal generator.

We note that our central limit theorem also applies to random walks on *finitely generated* (discrete) nilpotent groups. Indeed a torsion-free finitely generated nilpotent group can always be realized as a co-compact discrete subgroup of its Malcev closure, which is a connected and simply connected nilpotent Lie group [CG90] (the torsion subgroup is a finite normal subgroup, which does not play a role in the asymptotic behavior of the random walk). We refer the reader to [Ale02b, Bre10, Ish03, IKN20, IKN21, DH21] for related articles where limit theorems for random walks on finitely generated nilpotent groups as well as on nilpotent covers of compact manifolds are obtained.

Contrary to previous work on the nilpotent CLT, which relied on martingale techniques [SV73, Vir74, CCLP99] or convergence theorems for semigroups of contracting operators [Weh62, CR78, Rau78], our proof follows a rather down-to-earth Fourier analytic approach, which allows for greater flexibility and provides directly information on the rate of convergence. Similar techniques are employed by the second named author [Bre05] and by Diaconis and Hough [DH21, Hou19] in their work on the centered local limit theorem. Similarly as in Lindeberg’s proof of the classical CLT on \mathbb{R} (see [Fel68]), we shall use a *replacement scheme*, where at each step one inserts a new random variable distributed according to the limit measure. However, we will first need to replace the Lie group product $*$ by an associated

graded product \ast' , for which the renormalization acts by automorphisms. Coordinatewise, we are led to consider random variables of the form:

$$(1) \quad \sum_{n_1 < \dots < n_t \leq N} M(X_{n_1}, \dots, X_{n_t})$$

where M is a fixed polynomial function on \mathfrak{g} , and the X_i 's represent i.i.d. variables with law μ . Lindeberg's replacement scheme for polynomials has also been exploited recently in a different context in works on noise stability and influences [MOO10, Cha06], see also [Rot75].

We now pass to the more precise description of our results.

1.1. Main results. Let G be a simply connected nilpotent Lie group with Lie algebra \mathfrak{g} . We identify G with \mathfrak{g} via the exponential map. Let s be the step (or nilpotency class) of \mathfrak{g} , that is the smallest integer s such that $\mathfrak{g}^{[s+1]} = \{0\}$, where $\mathfrak{g}^{[i]} := [\mathfrak{g}, \mathfrak{g}^{[i-1]}]$ is the central descending series. We denote by \ast the Lie product on G , which we view as being defined on \mathfrak{g} . By abuse of notation we also denote by \ast the associated convolution product of measures on \mathfrak{g} . Note that \ast is given by a polynomial map on \mathfrak{g} as follows from the Campbell-Hausdorff formula [Dyn47]:

$$x \ast y = x + y + \frac{1}{2}[x, y] + \frac{1}{12}[x, [x, y]] + \frac{1}{12}[y, [y, x]] + \dots$$

We consider a random walk $S_n = X_1 \ast \dots \ast X_n$ associated to a probability measure μ on (\mathfrak{g}, \ast) . We let μ_{ab} be the projection of μ onto the abelianization $\mathfrak{g}_{ab} := \mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]$ and $\bar{X}_\mu = \mathbb{E}(\mu_{ab}) \in \mathfrak{g}_{ab}$ the mean.

It turns out that the large scale behaviour of S_n on the Lie group, is captured by a certain filtration of \mathfrak{g} , namely a certain decreasing sequence of ideals, whose definition depends only on $\bar{X}_\mu \in \mathfrak{g}_{ab}$,

$$\mathfrak{g}^{(1)} \supseteq \mathfrak{g}^{(2)} \supseteq \dots \supseteq \mathfrak{g}^{(t+1)} = \{0\}$$

such that $[\mathfrak{g}^{(i)}, \mathfrak{g}^{(j)}] \subseteq \mathfrak{g}^{(i+j)}$ for all i, j .

In the **centered case**, i.e. when $\mathbb{E}(\mu_{ab}) = 0$, the filtration is simply the central descending series, i.e. $\mathfrak{g}^{(i)} = \mathfrak{g}^{[i]}$ and $t = s$. The projection of the walk to \mathfrak{g}_{ab} exhibits diffusive behavior and will therefore be of order \sqrt{n} as n grows. On the other hand the coordinates of S_n lying deeper in the filtration, i.e. in say $\mathfrak{g}^{[i]} \setminus \mathfrak{g}^{[i+1]}$, grow faster at a rate $n^{i/2}$. Therefore if we renormalize the walk accordingly we may hope that it converges towards a non-degenerate random variable on \mathfrak{g} , analogous say to the Gaussian law in the abelian case. For a precise formulation, let us fix supplementary vector subspaces $\mathfrak{m}^{(i)}$ such that $\mathfrak{g}^{(i)} = \mathfrak{m}^{(i)} \oplus \mathfrak{g}^{(i+1)}$. This induces a direct sum decomposition $\mathfrak{g} = \mathfrak{m}^{(1)} \oplus \dots \oplus \mathfrak{m}^{(s)}$, whose coordinate projections we denote by

$$\begin{aligned} \pi^{(i)} : \mathfrak{g} &\rightarrow \mathfrak{m}^{(i)} \\ x &\mapsto \pi^{(i)}(x) = x^{(i)}. \end{aligned}$$

Let us define a one-parameter family of *dilations* $(D_r)_{r>0}$ by the formula

$$(2) \quad D_r : \mathfrak{g} \rightarrow \mathfrak{g}, \quad D_r(x) = \sum_{i \geq 1} r^i x^{(i)}.$$

Given $r > 0$ and a Borel measure η on \mathfrak{g} , we write $D_r(\eta)$ the push-forward measure of η by D_r , it is characterized by the relation $D_r(\eta)(f) = \eta(f \circ D_r)$ for every measurable function $f : \mathfrak{g} \rightarrow \mathbb{R}^+$. The centered CLT for the μ -walk on (\mathfrak{g}, \ast) claims the convergence in law

$$(3) \quad D_{\frac{1}{\sqrt{N}}}(\mu^{\ast N}) \rightarrow \nu$$

where ν is a certain smooth probability measure on \mathfrak{g} . It also involves a description of ν as the value at time $t = 1$ of a certain explicit one-parameter semigroup of probability measures $(\sigma_t)_{t \in \mathbb{R}^+}$ on \mathfrak{g} . However, one must be careful that the semigroup structure for σ_t refers to a new Lie group product $*'$ on \mathfrak{g} , inherited from the choice of subspaces $\mathfrak{m}^{(i)}$ and defined by

$$(4) \quad a *' b = \lim_{t \rightarrow +\infty} D_{1/t}(D_t(a) * D_t(b)).$$

Equivalently, $*'$ is the Lie group structure induced by the new Lie bracket $[x, y]'$ on \mathfrak{g} , defined on $\mathfrak{m}^{(i)} \times \mathfrak{m}^{(j)}$ as the projection of $[x, y]$ onto $\mathfrak{m}^{(i+j)}$ parallel to the other $\mathfrak{m}^{(k)}$. With this new Lie structure, $(\mathfrak{g}, *')$ is again an s -step nilpotent Lie group, and the dilations D_r become automorphisms. The semigroup $(\sigma_t)_{t > 0}$ can now be defined (as in [Hun56]) by way of its infinitesimal generator

$$(5) \quad L_\mu = \frac{1}{2} \sum_{i \leq \dim \mathfrak{g}_{ab}} E_i^2 + B$$

where E_i, B denote left invariant vectors fields on $(\mathfrak{g}, *')$, with the E_i 's forming a basis of $\mathfrak{m}^{(1)}$ in which the covariance matrix of μ_{ab} is the identity, while $B \in \mathfrak{m}^{(2)}$ is $\mathbb{E}_\mu(x^{(2)})$. The measure $\nu = \sigma_1$ admits a smooth and exponentially decaying density, which may be characterized alternatively as the value at time $t = 1$ of the fundamental solution (heat kernel) $u : \mathbb{R}_{>0} \times \mathfrak{g} \rightarrow \mathbb{R}^+$ to the PDE

$$(6) \quad \partial_t - \frac{1}{2} \sum_{i \leq \dim \mathfrak{g}_{ab}} E_i^2 + B = 0.$$

Here, we identify a vector field on \mathfrak{g} with the corresponding Lie derivative.

A proof of the centered central limit theorem (3) can be found in the work of Tutubalin [Tut64] in the case of the Heisenberg group and Crepel and Raugi [CR78] in general, with roots in the early work of Wehn [Weh62] who proved an infinitesimal CLT on general Lie groups, see also [Gre63, SV73, Bre04]. Note that the limit law ν is entirely determined by the covariance matrix of $\mu_{ab} = \pi^{(1)}(\mu)$ and the mean of $\pi^{(2)}(\mu)$, where $\pi^{(i)}(\mu)$ denotes the push-forward of μ by $\pi^{(i)}$. The convergence (3) holds assuming the existence of a finite second moment, i.e. $d(0, x) \in L^2(\mu)$ for some (any) left invariant geodesic distance d on $(\mathfrak{g}, *)$ or $(\mathfrak{g}, *')$. Equivalently, $\|x^{(i)}\|^{1/i} \in L^2(\mu)$ for each i .

EXAMPLE. Let G be the 3-dimensional Heisenberg group, so

$$\mathfrak{g} = \{xe_1 + ye_2 + ze_3 \mid [e_1, e_2] = e_3, [e_1, e_3] = [e_2, e_3] = 0\},$$

and assume that μ_{ab} is normalized, that is $\mathbb{E}(x^2) = \mathbb{E}(y^2) = 1$, $\mathbb{E}(xy) = \mathbb{E}(x) = \mathbb{E}(y) = \mathbb{E}(z) = 0$. Then ν coincides with the distribution of the random variable $Y = (x(1), y(1), z(1))$, where $(x(t), y(t))$ is a standard Brownian motion on \mathbb{R}^2 and $z(t)$ is the Lévy area (signed area between the cord and the brownian path). The density of ν has been computed by Paul Lévy [Lév50] (see also [Gav77, Théorème 1]):

$$d\nu(x, y, z) = \frac{dxdydz}{2\pi^2} \int_{\mathbb{R}} \cos(2\xi z) \frac{\xi}{\sinh \xi} e^{-\frac{1}{2}(x^2+y^2)\xi / \tanh(\xi)} d\xi.$$

For centered walks on general nilpotent Lie groups there is in general no known closed formula for the limit density $u(1, x)$ of $\nu = u(1, x)dx$, but so-called *Gaussian estimates* that control its decay at infinity are available [Heb89, Var88, VSCC92]. Namely there is $C > 0$ such that for all $x \in \mathfrak{g}$:

$$(7) \quad \frac{1}{C} \exp(-Cd(0, x)^2) \leq u(1, x) \leq C \exp(-d(0, x)^2/C)$$

with a similar upper bound for all partial derivatives of $u(1, \cdot)$. See also [AG16] where certain explicit formulas for $u(1, x)$ are derived in the case of filiform Lie algebras.

Now let us turn to **the non-centered case**, i.e. when $\overline{X}_\mu := \mathbb{E}(\mu_{ab})$ is possibly non-zero. Then, as in the abelian case, the random walk S_n needs to be translated by $-nx$ for some representative x of \overline{X}_μ in \mathfrak{g} , so that its projection to \mathfrak{g}_{ab} has zero mean and we can then attempt to renormalize the resulting random variable. It turns out that the right filtration to be considered for the renormalization is finer and that it depends on the way $[x, \cdot]$ acts on \mathfrak{g} . Indeed, we are looking at the random product:

$$(8) \quad X_1 * \dots * X_n * (-nx) = Y_1 * Y_2^x * \dots * Y_n^{(n-1)x}$$

where $Y_i := X_i * x^{-1}$ and $Y^g := g * Y * g^{-1}$. This is now a product of independent centered variables, though they are no longer identically distributed, due to the iterated adjoint action of x .

The drift of the random walk is linear in the x direction, while it only moves diffusively, at rate $n^{1/2}$, in the other directions of $\mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]$, so in effect x behaves as a bracket of order 2 even though it is not in the commutator subgroup. In order to take this into account, we introduce a new filtration as follows. For $i \leq 2s - 1$, we let $\mathfrak{g}^{(i)}$ be the subspace generated by brackets of length a in which x appears at least b times, where $a + b \geq i$. For $i = 2$ however we set $\mathfrak{g}^{(2)} = [\mathfrak{g}, \mathfrak{g}]$ rather than $[\mathfrak{g}, \mathfrak{g}] + \mathbb{R}x$. Equivalently, $\mathfrak{g}^{(i)}$ is defined recursively by setting $\mathfrak{g}^{(i+1)} = [\mathfrak{g}, \mathfrak{g}^{(i)}] + [x, \mathfrak{g}^{(i-1)}]$ for $i \geq 1$, while $\mathfrak{g}^{(0)} = \mathfrak{g}^{(1)} = \mathfrak{g}$.

We thus obtain a nested sequence of ideals:

$$\underbrace{\mathfrak{g}^{(1)}}_{\mathfrak{g}} \supseteq \underbrace{\mathfrak{g}^{(2)}}_{[\mathfrak{g}, \mathfrak{g}]} \supseteq \underbrace{\mathfrak{g}^{(3)}}_{[\mathfrak{g}, x] + [[\mathfrak{g}, \mathfrak{g}], \mathfrak{g}]} \supseteq \dots \supseteq \mathfrak{g}^{(2s)} = 0.$$

We call the filtration $(\mathfrak{g}^{(i)})_{i \leq 2s-1}$ the *weight filtration* of \mathfrak{g} associated to \overline{X}_μ . Indeed, note that it does not depend on the choice of representative $x \in \mathfrak{g}$ of $\overline{X}_\mu \in \mathfrak{g}_{ab}$. Furthermore $\mathfrak{g}^{[i]} \subseteq \mathfrak{g}^{(i)}$ for each i with equality if $\overline{X}_\mu = 0$. We then choose supplementary subspaces $\mathfrak{m}^{(i)}$ as before so that $\mathfrak{g}^{(i)} = \mathfrak{m}^{(i)} \oplus \mathfrak{g}^{(i+1)}$ and denote by X_μ the representative in $\mathfrak{m}^{(1)}$:

$$(9) \quad \overline{X}_\mu = X_\mu + [\mathfrak{g}, \mathfrak{g}], \quad X_\mu \in \mathfrak{m}^{(1)}$$

We keep the notation $y^{(i)} = \pi^{(i)}(y)$ to denote the projection of $y \in \mathfrak{g}$ onto $\mathfrak{m}^{(i)}$. We also define the dilations D_r as in (2) above, and the $*$ Lie product as in (4). We warn the reader that, unlike in the centered case, $\mathfrak{m}^{(1)}$ may no longer be bracket-generating in $(\mathfrak{g}, *)$ as happens already in the Heisenberg group case. Finally we define the nilpotent operator $a_{X_\mu} : \mathfrak{g} \rightarrow \mathfrak{g}$ by setting

$$(10) \quad a_{X_\mu}(y) = \pi^{(i+2)}([X_\mu, y]) \text{ when } y \in \mathfrak{m}^{(i)}.$$

Given a probability measure μ on G , we say that it has a finite moment of order $m \in (0, +\infty)$ for the weight filtration if $y \mapsto \|y^{(i)}\|^{m/i}$ is in $L^1(\mu)$ for each i (where we take some norm, say Euclidean, on each $\mathfrak{m}^{(i)}$). Since $\mathfrak{g}^{[i]} \subseteq \mathfrak{g}^{(i)}$ and the inclusion can be strict when $\overline{X}_\mu \neq 0$, this is a slightly weaker constraint on μ than the ordinary notion of moment on nilpotent Lie groups.

THEOREM 1.1 (Non-centered CLT). *Let G be a simply connected nilpotent Lie group with Lie algebra \mathfrak{g} . Let μ be a probability measure on G with finite second moment for the weight filtration induced by $\overline{X}_\mu := \mathbb{E}(\mu_{ab}) \in \mathfrak{g}_{ab}$. Assume that the support of μ_{ab} is not contained in a proper affine subspace of \mathfrak{g}_{ab} . Then*

$$(11) \quad D_{\frac{1}{\sqrt{N}}}(\mu^{*N} * \delta_{-NX_\mu}) \xrightarrow{N \rightarrow +\infty} \nu$$

where ν is a smooth probability measure on \mathfrak{g} . More precisely, writing $\nu_t := D_{\sqrt{t}}(\nu)$, we have $\nu_t = u(t, x)dx$ where the density $u(t, x)$ is the (smooth) fundamental solution to the following hypoelliptic time-dependent PDE:

$$(12) \quad \partial_t v = \frac{1}{2} \sum_{i \leq \dim \mathfrak{g}_{ab}} (\exp(ta_{X_\mu})E_i)^2 v - (\exp(ta_{X_\mu})B_\mu)v.$$

Here the E_i 's form a basis of $\mathfrak{m}^{(1)}$ in which $\pi^{(1)}(\mu)$ has identity covariance matrix. And $B_\mu := \mathbb{E}_\mu(x^{(2)})$ is the ‘‘commutator mean’’ of μ .

In (12), a vector $X \in \mathfrak{g}$ is seen as a left invariant vector field on $(\mathfrak{g}, *')$, with the corresponding Lie derivative

$$Xf = \lim_{\varepsilon \rightarrow 0} \frac{f(x *' \varepsilon X) - f(x)}{\varepsilon}.$$

Remark. It follows from the theorem that the limit measure ν depends only on $\overline{X}_\mu = \mathbb{E}(\mu_{ab})$, $\text{Cov}(\mu_{ab})$ the covariance matrix of μ_{ab} , the mean of μ in projection to $\mathfrak{g}/\mathfrak{g}^{(3)}$, and the choice of weight decomposition $\mathfrak{g} = \bigoplus_i \mathfrak{m}^{(i)}$.

Remark. The result remains valid verbatim if we replace the translation δ_{-NX_μ} in (11) by an arbitrary one $\delta_{-NX_\mu + g_N}$, provided the sequence $g_N \in \mathfrak{g}$ is such that $\|g_N^{(i)}\| = o(N^{i/2})$ for each $i \geq 1$. We may also replace the multiplicative recentering by an additive recentering, up to adjusting the limiting smooth probability measure, see Lemma 5.5.

We shall also establish upper estimates on the density $\nu = u(1, x)dx$, namely the existence of constants $C, c > 0$ such that

$$(13) \quad u(1, x) \leq C \exp(-d'(0, x)^c)$$

where d' is a left-invariant geodesic (e.g. Riemannian) metric on $(\mathfrak{g}, *')$, and similar estimates hold for all partial derivatives of $u(1, \cdot)$ (see Proposition 3.8), showing that $u(1, \cdot)$ is a Schwartz function on \mathfrak{g} .

The question of a lower bound is more subtle. In the centered case, Gaussian lower estimates have been known for a while [Var88, VSCC92]. However, when one authorizes the walk to be non-centered, the picture changes drastically as the support of the limiting distribution ν may not even coincide with \mathfrak{g} . In this case, the support of ν is a proper closed subset with non-empty (dense) interior. Describing the support is a question related to *control theory* on Lie groups ([Sac07]). We compute the support in an explicit example in the case of the filiform Lie algebra of step 3 (Lemma 3.23). This leads to the following:

THEOREM 1.2 (Never full support). *Assume G is a free nilpotent Lie group of step at least 3. Then for any non-centered probability measure μ on G satisfying the assumptions of Theorem 1.1, the support of the limiting distribution ν does not contain the identity in its interior.*

At the extreme opposite, we identify a class of nilpotent Lie groups for which every limiting distribution must have full support (Section 3.3.1), and give a satisfactory lower bound (Corollary 3.32). This class contains the upper triangular unipotent groups

$$Upp_d(\mathbb{R}) := \{a_{ij} \in \text{SL}_d(\mathbb{R}); a_{ii} = 1, i > j \Rightarrow a_{ij} = 0\},$$

or any nilpotent Lie group of step at most 2. It is defined by the *double cancellation condition* **(DC)**: there exists a bracket-generating family (v_i) of elements in $(\mathfrak{g}, [\cdot, \cdot])$, such that any bracket of the form $[v_i, [v, [\dots [v, v_i] \dots]]$ vanishes.

THEOREM 1.3 (Full support and lower bound). *Assume $(\mathfrak{g}, [\cdot, \cdot])$ satisfies **(DC)**. Then for any probability measure μ on G as in Theorem 1.1, the support of the limiting measure ν is all of \mathfrak{g} and its density $u(1, x)$ satisfies: $\exists \alpha, A > 0, \forall x \in \mathfrak{g}$,*

$$u(1, x) \geq \alpha \exp(-A d^l(0, x)^2).$$

In Section 3.4 we discuss a general lower bound valid in general even when the limiting distribution is not of full support. Using the recent work [KP16], we provide an answer in terms *Harnack chains* whose increments are not too close to the boundary of the support, see Theorem 3.28. This analysis allows us to establish that $u(1, x)$ does not vanish on the interior of the support. As a by-product, our lower bound allows us to characterize the case where ν is an ordinary Gaussian distribution (in the ordinary Euclidean sense). The implication i) \implies iii) below is due to Crepel-Raugi [CR78].

THEOREM 1.4 (Characterization of the Gaussian case). *Keep the assumptions of Theorem 1.1. The following are equivalent.*

i) *For all $a \in \{1, \dots, s\}$, we have*

$$[X_\mu, \mathfrak{g}^{[a]}] = \mathfrak{g}^{[a+1]}.$$

ii) *The eigenvalues of $D_{\frac{1}{\sqrt{N}}}$ are exactly $\{N^{-(2k-1)/2}, 1 \leq k \leq s\}$.*

iii) *The limit measure ν is Gaussian (in the classical Euclidean sense) on \mathfrak{g} .*

In [Tut64] (see also [ST66, §5.4]) Tutubalin asks for necessary and sufficient conditions for a pair of probability measures μ_1 and μ_2 on a nilpotent Lie group to be *asymptotically close* in the sense that

$$\sup_{A \in \mathcal{C}} |\mu_1^{*n}(A) - \mu_2^{*n}(A)| \xrightarrow{n \rightarrow +\infty} 0,$$

where \mathcal{C} is a certain class of Borel subsets of \mathfrak{g} , which for definiteness we will take to be the class of all convex subsets. Solutions were provided in the case of the upper triangular unipotent $n \times n$ matrix group, first by Tutubalin [Tut64, §6] for $n = 3$, then by Virtser [Vir74, Theorem 5] for arbitrary n and only certain values of \bar{X}_μ . Thanks to Theorem 1.1 we can answer Tutubalin's question in any simply connected nilpotent Lie group.

COROLLARY 1.5 (Characterization of asymptotically close measures). *Let G be a simply connected nilpotent Lie group with Lie algebra \mathfrak{g} . Consider two probability measures μ_1, μ_2 on G with a finite second moment for their respective weight filtrations and non singular covariance matrix in the abelianization. The measures μ_1 and μ_2 are asymptotically close if and only if their projections to $\mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]$ have the same mean and covariance matrix, and their projections to $\mathfrak{g}/\mathfrak{g}^{(3)}$ have the same mean.*

We recall that $\mathfrak{g}^{(3)} = [\mathfrak{g}, [\mathfrak{g}, \mathfrak{g}]] + [\bar{X}_\mu, \mathfrak{g}]$ and $\bar{X}_\mu = \mathbb{E}(\mu_{ab})$. In particular the equivalence classes of asymptotically close measures are determined by a finite number of parameters. This is in sharp contrast with what happens in semisimple Lie groups, see [ST66], [Bou81].

While previous approaches for proving the CLT were based either on probabilistic methods (martingale convergence [SV73, Vir74]) or functional analytic methods (perturbation of operator semigroups [Weh62, CR78, Rau78]), we use characteristic functions. This approach is motivated by [DH21, Hou19]. This has a pay-off as it allows us to obtain Berry-Esseen bounds, yielding information on the rate of convergence. **It is formulated in terms of the moments of μ . Given any $r > 0$ and a choice of norm on \mathfrak{g} , the r -th moment of μ (for the weight filtration) is defined by:**

$$m_r(\mu) := \sum_{1 \leq i \leq 2s-1} \int_{\mathfrak{g}} \|x^{(i)}\|^{r/i} d\mu(x).$$

THEOREM 1.6 (Berry-Esseen estimate). *Keep the assumptions of Theorem 1.1. Then for every smooth function of compact support f on \mathfrak{g} we have for all $N \geq 1$*

$$|D_{\frac{1}{\sqrt{N}}}(\mu^{*N} * \delta_{-NX_\mu})(f) - \nu(f)| \leq \varepsilon_N \max_{|\alpha| \leq \dim \mathfrak{g} + 4} \|\partial^\alpha f\|_{L^1}$$

where $\varepsilon_N = o_\mu(1)$. *Furthermore, if μ has a finite third moment for the weight filtration, then we can take*

$$\varepsilon_N \leq C(1 + m_2(\mu)^{c_s})(1 + m_3(\mu))N^{-1/2}$$

where $s \geq 1$ is the nilpotency class of \mathfrak{g} , and $c_s = 6s^2$ while $C > 0$ is a constant depending only on $(\mathfrak{g}, [\cdot, \cdot])$, a choice of norm and Haar measure, the line $\mathbb{R}\bar{X}_\mu \subseteq \mathfrak{g}_{ab}$, and the weight decomposition $(\mathfrak{m}^{(i)})_i$.

Here, ∂^α refers to partial derivatives of f (in the classical sense on a vector space), and $|\alpha|$ is the order of the derivation.

Finally, we shall prove that the previous CLT can be upgraded to the convergence of the full interpolated process in the spirit of Donsker's theorem [RY99, Theorem XIII.1.9]. We consider:

$$W^{(N)}(t) := D_{\frac{1}{\sqrt{N}}}(X_1 * \cdots * X_{[tN]} * (tN - [tN])X_{[tN]+1} * -tNX_\mu)$$

where the X_i 's are i.i.d. with law μ . Note that $W^{(N)}$ defines a random variable with values in the space of locally Lipschitz paths on \mathfrak{g} .

In view of (12), the natural candidate to be the limit of $W^{(N)}$ is the continuous left-invariant diffusion process on $(\mathfrak{g}, *')$ whose infinitesimal generator at time $t \geq 0$ (see Section 3.1) is

$$(14) \quad \mathcal{L}_t = \frac{1}{2} \sum_{i=1}^q (\exp(ta_{X_\mu})E_i)^2 + \exp(ta_{X_\mu})B_\mu.$$

Denoting by $\sigma = \{\sigma_t\}_{t \in \mathbb{R}^+}$ the law of the diffusion on the space of paths, Theorem 1.1 establishes the convergence in law of $W^{(N)}(1)$ towards σ_1 .

The process associated to $(\mathcal{L}_t)_t$ will have α -Hölder sample paths for any $\alpha \in [0, 1/2)$. We endow the space $C^{0,\alpha}(\mathbb{R}^+, \mathfrak{g})$ of \mathfrak{g} -valued α -Hölder paths with the topology of α -Hölder convergence on compact intervals. It is induced by the (separable complete) metric $d_\alpha = \sum_{n \geq 0} 2^{-n} d_{\alpha,n}$, where:

$$d_{\alpha,n}((x(t))_t, (y(t))_t) := \|x(0) - y(0)\| + \sup_{t \neq s \in [0, n]} \frac{\|x(s)^{-1} * x(t) - y(s)^{-1} * y(t)\|}{|t - s|^\alpha}.$$

The $*$ operation in this formula could be replaced with $*'$ or $+$ without changing the induced topology on $C^{0,\alpha}(\mathbb{R}^+, \mathfrak{g})$. Note that convergence in the α -Hölder topology is more and more restrictive as α grows, and implies uniform convergence on compact intervals (case $\alpha = 0$).

THEOREM 1.7 (Central limit theorem for processes). *Keep the assumptions of Theorem 1.1 and assume further that μ has finite moments of all orders. Then for each $\alpha \in [0, \frac{1}{2})$, as N tends to $+\infty$, the distribution of $W^{(N)}$ converges to σ in the space $C^{0,\alpha}(\mathbb{R}^+, \mathfrak{g})$ endowed with the topology of α -Hölder convergence on compact intervals.*

Given a fixed $\alpha \in [0, \frac{1}{2})$, the proofs require only a finite moment of order $2mb_{\max}$ for the weight filtration, where m is any integer $m > 1/(1 - 2\alpha)$, and $b_{\max} \leq 2s - 1$ is the length of the weight filtration $\mathfrak{g}^{(b)}$ defined by $\mathfrak{g}^{(b_{\max})} \supsetneq \mathfrak{g}^{(b_{\max}+1)} = \{0\}$.

Such invariance principles have been studied previously in the special case of centered walks. See in particular [CCLP99], which treats the $\alpha = 0$ case, and [BFH09], where Theorem 1.7 is proven for centered walks on the free nilpotent Lie group. We remark in passing that in the centered case the limiting diffusion is not time-dependent and even admits an explicit representation in terms iterated integrals of Brownian motion, see [Kun80, Theorem 6.1] and [Ben89, Théorème 17]. We shall also briefly discuss convergence in the sense of rough paths.

Remark. In a sequel to this paper [BB23], we pursue our study further and establish the local limit theorem for biased walks on arbitrary simply connected nilpotent Lie groups, thus extending the recent work of Diaconis and Hough [DH21, Hou19] to the biased case.

1.2. Organization of the paper. In Section 2, we set up the notations to be used in the rest of the paper. In Section 3, we study the class of diffusion processes on simply connected nilpotent Lie groups that arise as limits of rescaled right random walks with i.i.d. increments. Section 4 is devoted to various techniques of truncation, graded replacement and Lindeberg replacement, that help in estimating the Fourier transform of the random walk. In Section 5, we prove the central limit theorem, the Berry-Esseen estimate, and the analogue of Donsker's invariance principle for nilpotent random walks. We also characterize pairs of asymptotically close measures.

CONTENTS

2	Set-up and notation	12
2.1	Weight filtration on \mathfrak{g}	12
2.2	The bias extension $\widetilde{\mathfrak{g}}$	12
2.3	Grading	13
2.4	Dilations	13
2.5	Bi-grading on the free Lie algebra	13
2.6	Norm, Lebesgue measure, and dual	15
2.7	Polynomial functions and degree	15
2.8	The driving measure μ	15
2.9	The bias extension $\widetilde{\mu}$	15
2.10	Moderate deviations	16
2.11	Asymptotic notations	16
3	Limiting diffusion on a nilpotent Lie group	16
3.1	Diffusions on a connected Lie group	16
3.2	Invariance and regularity of the limiting diffusion	19

3.3	Support of the limiting diffusion	24
3.4	Discussion on a lower bound	32
4	The three reductions	37
4.1	Truncation and moment estimates	37
4.2	Graded replacement	40
4.3	Lindeberg replacement	41
5	Global limit theorems	43
5.1	Central limit theorem and convergence rate	43
5.2	Asymptotically close measures	48
5.3	Convergence of processes	48

2. Set-up and notation. In this section we set up the terminology and notations used throughout the paper.

2.1. *Weight filtration on \mathfrak{g} .* If \mathfrak{g} is a nilpotent Lie algebra and $\overline{X} \in \mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]$, we define recursively the following nested sequence of ideals: $\mathfrak{g}^{(0)} = \mathfrak{g}^{(1)} = \mathfrak{g}$ and for $i \geq 1$

$$\mathfrak{g}^{(i+1)} = [\mathfrak{g}, \mathfrak{g}^{(i)}] + [\overline{X}, \mathfrak{g}^{(i-1)}].$$

We note that it is well-defined (independently of the choice of representative of \overline{X} in \mathfrak{g}), and satisfies for $i, j \geq 1$,

$$(15) \quad [\mathfrak{g}^{(i)}, \mathfrak{g}^{(j)}] \subseteq \mathfrak{g}^{(i+j)}$$

We call $(\mathfrak{g}^{(i)})_{i \geq 1}$ the *weight filtration* of \mathfrak{g} associated to \overline{X} . When $\overline{X} = 0 \pmod{[\mathfrak{g}, \mathfrak{g}]}$, this is the familiar descending central series, which we denote by $(\mathfrak{g}^{[i]})_{i \geq 1}$. In general we have $\mathfrak{g}^{[i]} \subseteq \mathfrak{g}^{(i)}$, and $\mathfrak{g}^{(2i)} \subseteq \mathfrak{g}^{[i+1]}$. In particular, if \mathfrak{g} has nilpotency class at most s , then $\mathfrak{g}^{[s+1]} = \mathfrak{g}^{(2s)} = 0$. The weight filtration $(\mathfrak{g}^{(i)})_{i \geq 1}$ on \mathfrak{g} allows to define a *graded Lie algebra*:

$$(16) \quad gr_{\overline{X}}(\mathfrak{g}) := \bigoplus_{i \geq 1} \mathfrak{g}^{(i)} / \mathfrak{g}^{(i+1)}$$

whose Lie bracket is inherited from that of \mathfrak{g} using (15).

EXAMPLE. In the case of the Heisenberg group

$$\mathfrak{g} = \langle e_1, e_2, e_3 \mid [e_1, e_2] = e_3 \rangle \quad \text{with} \quad \overline{X} = e_2 \pmod{[\mathfrak{g}, \mathfrak{g}]}$$

we have $\mathfrak{g}^{(0)} = \mathfrak{g}^{(1)} = \mathfrak{g}$, $\mathfrak{g}^{(2)} = \{0\}$, $\mathfrak{g}^{(3)} = \mathbb{R}e_3$, $(\mathfrak{g}^{(i)})_{i \geq 4} = \{0\}$, and $gr_{\overline{X}}(\mathfrak{g})$ is abelian.

2.2. *The bias extension $\tilde{\mathfrak{g}}$.* We define a new Lie algebra $\tilde{\mathfrak{g}}$ by adding to \mathfrak{g} an independent copy χ of the bias $\overline{X} \in \mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]$. This extension is a convenient technical device to dissociate the contributions of the drift incarnated by \overline{X} , which ought to have weight 2, from those of the fluctuations of the random walk in the direction of \overline{X} , which ought to have weight 1. Many intermediate results will therefore be formulated in $\tilde{\mathfrak{g}}$, before being translated back to \mathfrak{g} .

If $\overline{X} = 0$, we just set $\tilde{\mathfrak{g}} = \mathfrak{g}$, $\chi = 0$. Otherwise, we define $\tilde{\mathfrak{g}}$ as the Lie algebra direct sum:

$$\tilde{\mathfrak{g}} := \mathfrak{g} \oplus \mathbb{R}$$

and given a lift X of \overline{X} to \mathfrak{g} , we set $\chi := (X, 1)$. In this way, $\tilde{\mathfrak{g}} = \mathfrak{g} + \mathbb{R}\chi$ where χ is linearly independent from \mathfrak{g} and satisfies $[\chi, x] = [X, x]$ for all $x \in \mathfrak{g}$. We call $\tilde{\mathfrak{g}}$ the *bias extension* of \mathfrak{g} .

We also extend the weight filtration from \mathfrak{g} to $\tilde{\mathfrak{g}}$ by setting $\tilde{\mathfrak{g}}^{(i)} = \mathfrak{g}^{(i)}$ if $i \neq 2$, and $\tilde{\mathfrak{g}}^{(2)} = \mathfrak{g}^{(2)} + \mathbb{R}\chi$. It is straightforward to check that this is indeed a filtration, i.e. $[\tilde{\mathfrak{g}}^{(i)}, \tilde{\mathfrak{g}}^{(j)}] \subseteq \tilde{\mathfrak{g}}^{(i+j)}$, and is independent of the choice of lift. The filtration $(\tilde{\mathfrak{g}}^{(i)})_{i \geq 1}$ determines as in (16) a graded Lie algebra $gr_{\overline{X}}(\tilde{\mathfrak{g}})$, which we call the *graded bias extension*.

Observe that $gr_{\overline{X}}(\tilde{\mathfrak{g}}) = gr_{\overline{X}}(\mathfrak{g})$ if $\overline{X} = 0$. If $\overline{X} \neq 0$, $gr_{\overline{X}}(\tilde{\mathfrak{g}})$ is isomorphic to the Lie algebra *semi-direct* sum

$$gr_{\overline{X}}(\mathfrak{g}) \oplus \mathbb{R},$$

where the \mathbb{R} factor acts on the ideal $gr_{\overline{X}}(\mathfrak{g})$ via the nilpotent endomorphisms $a_X(t)$ defined on $\mathfrak{g}^{(i)} / \mathfrak{g}^{(i+1)}$ by the formula

$$a_X(t)(y) = [tX, y] \pmod{\mathfrak{g}^{(i+3)}}.$$

Different choices of representative of X in \mathfrak{g} lead to isomorphic extensions of $gr_{\overline{X}}(\mathfrak{g})$.

Remark. The limit measure ν_t appearing in the CLT (Theorem 1.1) will arise as the projection of the value at t of a one-parameter semigroup of probability measures λ_t defined on $gr_{\overline{X}}(\tilde{\mathfrak{g}})$ and left-invariant for that Lie product. See Section 5.

2.3. *Grading.* We choose for all $i \leq 2s - 1$ a vector subspace $\mathfrak{m}^{(i)}$ in \mathfrak{g} such that

$$\mathfrak{g}^{(i)} = \mathfrak{m}^{(i)} \oplus \mathfrak{g}^{(i+1)}$$

It follows that \mathfrak{g} is the (vector space) direct sum

$$\mathfrak{g} = \mathfrak{m}^{(1)} \oplus \dots \oplus \mathfrak{m}^{(2s-1)}$$

and we may identify $\mathfrak{m}^{(1)}$ with $\mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]$. From now on, the notation X will indicate the unique element of $\mathfrak{m}^{(1)}$ lifting X_μ . We call $(\mathfrak{m}^{(i)})_{i \leq 2s-1}$ a *weight decomposition* of \mathfrak{g} . It naturally extends to a weight decomposition of $\tilde{\mathfrak{g}}$:

$$\tilde{\mathfrak{g}} = \tilde{\mathfrak{m}}^{(1)} \oplus \dots \oplus \tilde{\mathfrak{m}}^{(2s-1)}$$

by setting $\tilde{\mathfrak{m}}^{(i)} = \mathfrak{m}^{(i)}$ if $i \neq 2$ and $\tilde{\mathfrak{m}}^{(2)} = \mathfrak{m}^{(2)} \oplus \mathbb{R}\chi$. These splittings allow to define new compatible Lie brackets on $\mathfrak{g}, \tilde{\mathfrak{g}}$ by setting

$$(17) \quad [x, y]' = \pi^{(i+j)}([x, y])$$

if $x \in \tilde{\mathfrak{m}}^{(i)}, y \in \tilde{\mathfrak{m}}^{(j)}$ and $\pi^{(k)} : \tilde{\mathfrak{g}} \rightarrow \tilde{\mathfrak{m}}^{(k)}$ denotes the linear projection modulo the other $\tilde{\mathfrak{m}}^{(k')}$, $k' \neq k$. Notice that the vector space \mathfrak{g} is an ideal of $(\tilde{\mathfrak{g}}, [.,.]')$. The Lie product associated to $[.,.]'$ will be denoted by $*'$. With this new Lie bracket, the vector space \mathfrak{g} (resp. $\tilde{\mathfrak{g}}$) becomes a Lie algebra naturally isomorphic to the graded Lie algebra $gr_{\overline{X}}(\mathfrak{g})$ (resp. $gr_{\overline{X}}(\tilde{\mathfrak{g}})$) and the $\mathfrak{m}^{(i)}$ (resp. $\tilde{\mathfrak{m}}^{(i)}$) form a *grading*.

We fix a choice of weight grading $(\mathfrak{m}^{(i)})_{i \leq 2s-1}$ and write $X \in \mathfrak{m}^{(1)}$ the lift of \overline{X} .

Remark. Comparing 15 and 17, we see that X lies in the center of $(\tilde{\mathfrak{g}}, *')$.

EXAMPLE. In the case of the biased Heisenberg group mentioned in 2.1, we can choose $\mathfrak{m}^{(1)} = \mathbb{R}e_1 \oplus \mathbb{R}e_2$, $\mathfrak{m}^{(2)} = \{0\}$, $\mathfrak{m}^{(3)} = \mathbb{R}e_3$. The bracket $[.,.]'$ is such that $[e_i, e_j]' = 0$ and $[\chi, e_j]' = [e_2, e_j]$ for $i, j \in \{1, 2, 3\}$.

2.4. *Dilations.* The weight decompositions on \mathfrak{g} and $\tilde{\mathfrak{g}}$ induce a one-parameter subgroup of dilations D_r , which are the linear maps that act on $\tilde{\mathfrak{m}}^{(i)}$ by multiplication by r^i . Note that with the $*'$ product, D_r becomes an automorphism of $\tilde{\mathfrak{g}}$ (and \mathfrak{g}): $\forall x, y \in \tilde{\mathfrak{g}}$,

$$D_r(x *' y) = D_r x *' D_r y.$$

The determinant of the dilations D_r on \mathfrak{g} is $\det D_r = r^{d_{\overline{X}}}$, where

$$(18) \quad d_{\overline{X}} = \sum_{i \geq 1} \dim \mathfrak{g}^{(i)}.$$

We call $d_{\overline{X}}$ the *homogeneous dimension* of \mathfrak{g} with respect to the weight filtration induced by \overline{X} . When $\overline{X} = 0$, $d_{\overline{X}}$ coincides with the homogeneous dimension of \mathfrak{g} , namely $d_0 = \sum_{i \geq 1} \dim \mathfrak{g}^{(i)}$, which controls the volume growth of balls in G by the Bass-Guivarc'h formula [Gui73, Bre14]. Note that $d_{\overline{X}} \geq d_0$ in general.

2.5. *Bi-grading on the free Lie algebra.* For basics on the free Lie algebra we refer the reader to [Ser06, chapter 2] and [Reu93]. We denote by $L^{[r]}$ the *r-bracket*, which is defined recursively as $L^{[1]}(x) = x$ and

$$L^{[r]}(x_1, \dots, x_r) = [L^{[r-1]}(x_1, \dots, x_{r-1}), x_r],$$

an element of the free Lie algebra on r letters.

Throughout we will consider formal expressions involving Lie brackets of N indeterminates x_1, \dots, x_N corresponding to $\tilde{\mathfrak{g}}$ -valued random variables. In particular the product of $\Pi(\underline{x}) := x_1 * x_2 * \dots * x_N$ can be expanded via the Campbell-Hausdorff formula (see [Dyn47]) into an element of the s -step free Lie algebra on N variables of the form:

$$\Pi(\underline{x}) = \sum_{r=1}^s \sum_{\substack{r_1 + \dots + r_t = r \\ 1 \leq n_1 < \dots < n_t < \infty}, r_i \neq 0} L_{\underline{r}}(x_{n_1}^{\otimes r_1}, \dots, x_{n_t}^{\otimes r_t})$$

where $\underline{r} = (r_1, \dots, r_t)$ and $L_{\underline{r}}$ is a linear combination with rational coefficients of terms, each of which is an r -bracket $L^{[r]}$ of some permutation of the r variables involved.

EXAMPLE. The Campbell-Hausdorff formula allows to compute $L_1(x_1) = x_1$, $L_{1,1} = \frac{1}{2}L^{[2]}$, $L_{2,1} = L_{1,2} = \frac{1}{12}L^{[3]}$,

$$L_{1,1,1}(x_1, x_2, x_3) = \frac{1}{6}L^{[3]}(x_1, x_2, x_3) + \frac{1}{6}L^{[3]}(x_3, x_2, x_1)$$

Note that the free Lie algebra comes with a natural grading given by bracket length. To account for the weight structure discussed above, we will need to consider a further grading on the s -step free Lie algebra. To each indeterminate x_j we associate $2s - 1$ extra indeterminates $x_j^{(i)}$ for $i = 1, \dots, 2s - 1$ so that $x_j = x_j^{(1)} + \dots + x_j^{(2s-1)}$ and expand the product $\Pi(x_1, \dots, x_N)$ as an element of the s -step free Lie algebra \mathcal{F}_s on the set of $(2s - 1)N$ indeterminates $\{x_j^{(i)}\}_{1 \leq j \leq N, 1 \leq i \leq 2s-1}$. We define a grading on this free Lie algebra by assigning weight i to each $x_j^{(i)}$ and defining the weight of an r -bracket of $x_j^{(i)}$'s to be the sum of the weights of its r variables. On the other hand, we have the ordinary grading of the free Lie algebra, where the degree of an r -bracket equals its length r . This yields a bi-grading

$$(19) \quad \mathcal{F}_s = \bigoplus_{1 \leq a \leq s, 1 \leq b \leq 2s-1} \mathcal{F}_s^{[a,b]}$$

where the homogeneous part $\mathcal{F}_s^{[a,b]}$ is spanned by all a -brackets of weight b in the variables $x_j^{(i)}$. We thus get a decomposition:

$$\Pi(\underline{x}) = \bigoplus_{a,b} \Pi^{[a,b]}(\underline{x}).$$

We also set $\Pi^{[a]}(\underline{x}) := \bigoplus_b \Pi^{[a,b]}(\underline{x})$ and $\Pi^{(b)}(\underline{x}) := \bigoplus_a \Pi^{[a,b]}(\underline{x})$.

Coming back to our Lie algebra $\tilde{\mathfrak{g}}$ of nilpotency class at most s (we use the words step and nilpotency class interchangeably) endowed with the grading $\tilde{\mathfrak{g}} = \bigoplus_{1 \leq b \leq 2s-1} \tilde{\mathfrak{m}}^{(b)}$, the element $\Pi^{[a,b]}(\underline{x}) \in \mathcal{F}_s$ naturally yields an evaluation mapping on $\tilde{\mathfrak{g}}^N$ with values in $\tilde{\mathfrak{g}}^{(b)} \cap \tilde{\mathfrak{g}}^{[a]}$, which, by abuse of notation, we continue to denote by $\Pi^{[a,b]}(\underline{x})$:

$$(20) \quad \Pi^{[a,b]} : \tilde{\mathfrak{g}}^N \rightarrow \tilde{\mathfrak{g}}^{(b)} \cap \tilde{\mathfrak{g}}^{[a]}$$

In fact, when the evaluation is taken in $\tilde{\mathfrak{g}}$ with the $*'$ product, we will denote the evaluation mapping by $\Pi'^{[a,b]}$. We note that $\Pi'^{[a,b]}$ is homogeneous of degree b with respect to the dilations D_r , namely for any $\underline{x} \in \tilde{\mathfrak{g}}^N$:

$$\Pi'^{[a,b]}(D_r \underline{x}) = r^b \Pi'^{[a,b]}(\underline{x})$$

2.6. *Norm, Lebesgue measure, and dual.* Throughout we shall fix a basis $(e_j^{(i)})_{1 \leq j \leq q_i}$ of each $\mathfrak{m}^{(i)}$, adding $e_0^{(2)} = \chi$ to $\tilde{\mathfrak{m}}^{(2)}$. We shall consider the induced Euclidean norm $\|\cdot\|$ on \mathfrak{g} and $\tilde{\mathfrak{g}}$, thus making the subspaces $\tilde{\mathfrak{m}}^{(i)}$ mutually orthogonal. The associated Lebesgue measure will be denoted by dx . Notice that dx is also a left and right invariant Haar measure on the associated Lie groups (for both $*$ and $*'$).

We also fix a choice of left invariant Riemannian metrics on the various group structures $(\mathfrak{g}, *)$, $(\mathfrak{g}, *')$, $(\tilde{\mathfrak{g}}, *')$ and denote them respectively by $|\cdot|_{\mathfrak{g}}$, $|\cdot|'_{\mathfrak{g}}$, $|\cdot|'_{\tilde{\mathfrak{g}}}$.

We will often consider the dual $\hat{\mathfrak{g}}$ of linear forms on \mathfrak{g} . It is endowed with operator norm and the Lebesgue measure $d\xi$ associated to the dual basis of the $(e_j^{(i)})_{i,j}$. Linear forms $\xi \in \hat{\mathfrak{g}}$ will be identified with their extension to $\tilde{\mathfrak{g}}$ by setting $\xi(\chi) = 0$.

2.7. *Polynomial functions and degree.* With the above basis, we obtain a well-defined notion of degree of a polynomial function by assigning degree i to each member $e_j^{(i)*}$ of the dual basis of $\tilde{\mathfrak{m}}^{(i)}$. For example, composing $\Pi^{[a,b]}(\underline{x})$ with any linear form on $\tilde{\mathfrak{m}}^{(b)}$ yields a polynomial function of degree b . This notion of degree depends on the filtration $(\tilde{\mathfrak{g}}^{(b)})_{b \geq 1}$, but not on the choice of weight decomposition and nor on the choice of basis.

A function on \mathfrak{g} or $\tilde{\mathfrak{g}}$ is Schwartz if it is smooth with all derivatives decaying faster than any polynomial. The class of Schwartz functions is solely determined by the vector space structure of \mathfrak{g} or $\tilde{\mathfrak{g}}$.

2.8. *The driving measure μ .* Let μ be a probability measure on \mathfrak{g} . For $r > 0$, we say that μ has *finite r -th moment* for the weight filtration if one has

$$m_r(\mu) := \sum_{1 \leq b \leq 2s-1} \int_{\mathfrak{g}} \|x^{(b)}\|^{r/b} d\mu(x) < \infty$$

In particular, for every polynomial $P : \mathfrak{g} \rightarrow \mathbb{R}$ of degree at most m for the weight filtration

$$\int_{\mathfrak{g}} |P(x)| d\mu(x) < \infty$$

Two measures μ, η with finite m -th moment are said to *coincide up to order m* if $\mu(P) = \eta(P)$ for all P of degree at most m .

We will estimate the μ -average (or μ^{*N} -average) of an integrable test function $f \in L^1(\mathfrak{g}, dx)$ using Fourier analysis. Given $\xi \in \hat{\mathfrak{g}}$, we set

$$e_{\xi} : \mathfrak{g} \rightarrow \mathbb{C}, x \mapsto e^{-2i\pi\xi(x)}$$

and we define the *Fourier transform* of μ or f by

$$\hat{\mu}(\xi) = \int_{\mathfrak{g}} e_{\xi}(x) d\mu(x) \quad \hat{f}(\xi) = \int_{\mathfrak{g}} e_{-\xi}(x) f(x) dx$$

If f is continuous and \hat{f} is integrable, the Fourier inversion formula states that

$$\mu(f) = \int_{\hat{\mathfrak{g}}} \hat{f}(\xi) \hat{\mu}(\xi) d\xi$$

2.9. *The bias extension $\tilde{\mu}$.* We will consider a measure μ with finite second moment for the weight filtration and whose projection μ_{ab} to $\mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]$ has expectation $\bar{X} \pmod{[\mathfrak{g}, \mathfrak{g}]}$. We define the *bias extension* $\tilde{\mu}$ of μ as the image of μ by $\mathfrak{g} \rightarrow \tilde{\mathfrak{g}}, x \mapsto (x, 1)$. Given a weight

decomposition $(\mathfrak{m}^{(b)})$ (thus lifting \overline{X} into $X \in \mathfrak{m}^{(1)}$ and defining $\chi = (X, 1) \in \widetilde{\mathfrak{g}}$), one may write for every $N \geq 1$,

$$\widetilde{\mu}^{*N} = \mu^{*N} - NX + N\chi = \mu^{*N} * -NX * N\chi$$

Here, the notation $\mu^{*N} - NX + N\chi$ represents the image of μ^{*N} by the map $\widetilde{\mathfrak{g}} \rightarrow \widetilde{\mathfrak{g}}, x \mapsto x - NX + N\chi$, and similarly for $\mu^{*N} * -NX * N\chi$. The measure $\widetilde{\mu}$ reports the drift $X \in \mathfrak{m}^{(1)}$ into $\chi \in \widetilde{\mathfrak{m}}^{(2)}$ while keeping the fluctuations around the drift in $\mathbb{R}X \subseteq \mathfrak{m}^{(1)}$.

2.10. Moderate deviations. Given $N \geq 1, g, h \in \widetilde{\mathfrak{g}}$, we denote by $g * \widetilde{\mu}^{*N} * h$, or $\widetilde{\mu}_{g,h}^{*N}$ for short, the image of $\widetilde{\mu}^{*N}$ by $\widetilde{\mathfrak{g}} \rightarrow \widetilde{\mathfrak{g}}, x \mapsto g * x * h$. A similar notation holds for deviations of μ on \mathfrak{g} .

We shall prove estimates for $\widetilde{\mu}_{g,h}^{*N}$ uniformly for g, h in a certain range of *moderate deviations*: given $\delta_0 \in [0, 1)$, we let

$$\mathcal{D}_N(\delta_0) = \{x \in \widetilde{\mathfrak{g}} : \forall i \leq 2s - 1, \|x^{(i)}\| \leq N^{i/2+s^{-1}\delta_0}\},$$

where $x^{(i)} = \pi^{(i)}(x)$ is the coordinate projection with respect to the grading $(\widetilde{\mathfrak{m}}^{(b)})_b$ of $\widetilde{\mathfrak{g}}$. It is worth pointing out that \mathcal{D}_N behaves well with respect to the product $*$: There exists $C > 0$ such that for all $\delta_0 > 0$, for all $N \geq 1$ large enough,

$$x, y \in \mathcal{D}_N(\delta_0) \implies x * y \in \mathcal{D}_N(C\delta_0)$$

2.11. Asymptotic notations. In the rest of the text, we use the standard Vinogradov notation \ll and Landau notations $O(\cdot), o(\cdot)$. Unless otherwise stated, the implied constants will depend only on the initial data $\mathfrak{g}, \overline{X}, \mu, (\mathfrak{m}^{(b)})_b, (e_j^{(i)})_{i,j}, |\cdot|_{\mathfrak{g}}, |\cdot|'_{\mathfrak{g}}, |\cdot|'_{\widetilde{\mathfrak{g}}}$. In some statements, we will also use the notation $r_1 \lll r_2$ to mean that the conclusions that follow are true up to choosing $r_1 \leq cr_2$ where $c > 0$ is a (small) constant depending only the initial data. Subscripts are added to indicate any additional dependency (for instance, on the parameter δ_0 controlling the range of deviations). Furthermore \mathbb{N} denotes the non-negative integers and \mathbb{R}^+ the non-negative reals.

3. Limiting diffusion on a nilpotent Lie group. In this section, we study the diffusion processes on a nilpotent Lie group that arise as limits of rescaled i.i.d. right random walks. We start by recalling background material on the general theory of diffusion processes on Lie groups. Then we focus on the case of limiting diffusions on nilpotent Lie groups, and establish hypoellipticity as well as a general Gaussian upper bound. We continue with a description of the support of the time-1 distribution, and give several examples. Finally, we give a general lower bound, and several corollaries, notably characterizing the case where the limiting process is Gaussian in the ordinary Euclidean sense.

3.1. Diffusions on a connected Lie group. We start by recalling some standard facts about diffusion processes on a connected real Lie group. We refer the reader to the books [Lia04] and [HS01] as well as Hunt's original article [Hun56] and the papers by Stroock and Varadhan [SV72, SV73] for more background material on diffusion processes. The support of invariant diffusions was also studied by Siebert, see [Sie82, Theorem 2]. Connections with sub-Riemannian geometry are discussed in [VSCC92, Kup97] and references therein.

Let H be a connected real Lie group and denote by 1_H the identity element of H . A *continuous left-invariant diffusion process* on H is a stochastic process $(W(t))_{t \in \mathbb{R}^+}$ satisfying the following:

- $W(0) = 1_H$ with probability 1.

- The random map $W : \mathbb{R}^+ \rightarrow H, t \mapsto W(t)$ is continuous with probability 1.
- For any $k \in \mathbb{N}$ and $0 = t_0 < \dots < t_k < +\infty$, the increments $(W(t_i)^{-1}W(t_{i+1}))_{i=0, \dots, k-1}$ are mutually independent.

In addition, if for any $0 \leq s < t$, the distribution of the increment $W(s)^{-1}W(t)$ only depends on $t - s$, then the process is called *homogeneous*.

Denote by $\Omega(H)$ the set of continuous paths $c : \mathbb{R}^+ \rightarrow H$, set \mathcal{F}_t the σ -algebra generated by the projections $s \mapsto c(s)$ where $s \leq t$, and $\mathcal{F}_\infty = \cup_{t \in \mathbb{R}^+} \mathcal{F}_t$. The law of a diffusion process W on H is a certain distribution σ on $(\Omega(H), \mathcal{F}_\infty)$. We denote by σ_s the image of σ by $c \mapsto c(s)$, or in other terms the law of $W(s)$. We denote by \mathcal{Q} the collection of probability measures on $(\Omega(H), \mathcal{F}_\infty)$ which arise from a diffusion, and call them diffusion laws on H .

A left-invariant second order differential operator on H defines a diffusion law on H as follows. Denote by \mathfrak{h} the Lie algebra of H , identify any vector $X \in \mathfrak{h}$ with its associated left-invariant vector field on H and with the corresponding Lie derivative. In particular, for any $f \in C^1(H)$,

$$X.f : x \mapsto \lim_{\varepsilon \rightarrow 0} \frac{f(xe^{\varepsilon X}) - f(x)}{\varepsilon}$$

Consider a time-dependent left-invariant second order differential operator $\mathcal{L} = (\mathcal{L}_t)_{t \in \mathbb{R}^+}$ of the form

$$\mathcal{L}_t = \frac{1}{2} \sum_{i=1}^q E_i(t)^2 + B(t)$$

where $E_i : \mathbb{R}^+ \rightarrow \mathfrak{h}$ for $i = 1, \dots, q$ and $B : \mathbb{R}^+ \rightarrow \mathfrak{h}$ are smooth maps.

FACT 3.1 (Infinitesimal generator [Lia04]). There exists a unique diffusion law $\sigma \in \mathcal{Q}$ such that for every $s \geq 0$ and every \mathcal{F}_s -measurable variable $f \in C_c^\infty(H)$, the process

$$(t \in \mathbb{R}^+) \quad f(c(t)) - f(c(s)) - \int_s^t (\mathcal{L}_r f)(c(r)) dr$$

is a martingale on $(\Omega(H), (\mathcal{F}_t)_{t \geq s}, \sigma)$.

Note that σ characterizes \mathcal{L} because for any fixed $f \in C_c^\infty(H)$,

$$\frac{1}{\varepsilon} \mathbb{E}_\sigma (f [c(s)^{-1}c(s+\varepsilon)] - f(1_H) | \mathcal{F}_s) \xrightarrow{\varepsilon \rightarrow 0^+} \mathcal{L}_s f(1_H)$$

where the arrow stands for the convergence in probability.

A diffusion process on H with infinitesimal generator \mathcal{L} is by definition a stochastic process W on H with law σ . Observe then that for any $s > 0$, the process $(W(s)^{-1}W(s+t))_{t \geq 0}$ is a diffusion with infinitesimal generator $(\mathcal{L}_{s+t})_{t \in \mathbb{R}^+}$. It follows that $(W(t))_{t \geq 0}$ is homogeneous if and only if \mathcal{L} is independent of the time parameter, and in this case the distributions σ_t form a semigroup: $\forall s, t \in \mathbb{R}^+$,

$$\sigma_s * \sigma_t = \sigma_{s+t}.$$

We now describe the support of the time- t distribution σ_t associated to \mathcal{L} . Given an open interval I , a C^1 path $\gamma : I \rightarrow H$, and $s \in I$, we define the multiplicative derivative $\partial\gamma(s) \in \mathfrak{h}$ as the derivative at $t = 0$ (in the classical sense) of the path $t \mapsto \gamma(s)^{-1}\gamma(s+t)$. We say that a

continuous, piecewise C^1 path $\gamma : [0, t] \rightarrow H$ is \mathcal{L} -horizontal if its multiplicative derivative satisfies

$$\partial\gamma(s) \in \sum_{i=1}^q \mathbb{R}E_i(s) + B(s)$$

for every s in the complement of a finite subset of $[0, t]$.

FACT 3.2 (Support [SV72, SV73]). The support of σ_t is the closure of the set of arrival points $\gamma(t)$, where $\gamma : [0, t] \rightarrow H$ is a continuous piecewise C^1 \mathcal{L} -horizontal path.

This fact is a consequence of Stroock-Varadhan's support theorem [SV72] for diffusion processes on \mathbb{R}^n , applied to W locally in charts of H as in [SV73, page 284], where it is shown that invariant diffusions on Lie groups never explode in finite time.

Next we give a smoothness criterion for the distribution σ_t . **Denote by \mathcal{H} the Lie algebra of vector fields on the product manifold $\mathbb{R}_{>0} \times H$ generated by $(E_i)_{i \leq q}$ and $\partial_t + B$** , where ∂_t stands for the constant vector field $(1, 0)$. We see E_i, B as functions $\mathbb{R}_{>0} \times H \rightarrow TH$, where TH is the tangent bundle of H . Recall Hörmander's criterion for hypoellipticity:

FACT 3.3 (Hypoellipticity [Hör67]). Assume the Hörmander condition: for every point $(t, x) \in \mathbb{R}_{>0} \times H$, every vector $v \in TH_x$ tangent to x , there exists $Y \in \mathcal{H}$ such that $Y(t, x) = v$. Then the differential equation on $\mathbb{R}_{>0} \times H$ given by

$$(21) \quad \partial_t - \frac{1}{2} \sum_{i=1}^q E_i^2 + B = 0$$

is hypoelliptic, i.e. every solution (in the sense of distributions) is smooth.

A direct computation shows that the distribution $u = \int_{\mathbb{R}_{>0}} \delta_t \otimes \sigma_t dt$ is solution of (21). Assuming (21) is hypoelliptic, one may then write for $t > 0$,

$$\sigma_t = u(t, x) dx$$

where $u : \mathbb{R}_{>0} \times H \rightarrow \mathbb{R}^+$ is a smooth function, and dx stands for a left-invariant Haar measure on H .

Hypoellipticity also yields the following domination principle [VSCC92, Corollary III.1.3]: for any open set $\Omega \subseteq \mathbb{R}_{>0} \times H$, any compact subset $K \subseteq \Omega$, $p \in \mathbb{N}$, there exists $C_1 > 0$ such that for every solution $v(t, x)$ of (21), one has

$$(22) \quad \|v\|_{C^p(K)} \leq C_1 \int_{\Omega} |v| dt dx$$

where $\|v\|_{C^p(K)}$ is the supremum on K of the absolute value of the derivatives of v up to order p . Later, we will use this inequality to convert a moment estimate on σ into a bound on the derivatives of the density u .

We also record a Harnack inequality, namely a statement to the effect that non-negative solutions of (21) do not decrease too much while going along a small horizontal path. We rely here on the recent work of Kogoj and Polidoro [KP16] addressing the setting of a time-dependent hypoelliptic equation. Let $1_H \in U \subseteq H$ be an open neighborhood of the identity, whose closure is compact and included in a chart of H . Let $0 < \varepsilon_0 < \varepsilon_1$ be two parameters and define \mathcal{P} as the set of couples $(\tau, x) \in (\varepsilon_0, \varepsilon_1) \times U$ such that there exists a U -valued horizontal path $\gamma : [\varepsilon_0, \tau] \rightarrow U$ such that $\gamma(\varepsilon_0) = 1_H$, $\gamma(\tau) = x$.

FACT 3.4 (Harnack inequality [KP16]). Assume the Hörmander condition from Fact 3.3. Let $U, \varepsilon_0, \varepsilon_1, \mathcal{P}$ as above. Let L be a compact set in the interior of the closure of \mathcal{P} in $[\varepsilon_0, \varepsilon_1] \times \bar{U}$. Then there exists a constant $C > 1$ such that for any non negative solution $v : (\frac{\varepsilon_0}{2}, 2\varepsilon_1) \times U \rightarrow \mathbb{R}^+$ of (21), one has

$$v(\varepsilon_0, 1_H) \leq C \inf_{(t,x) \in L} v(t, x)$$

This fact is a consequence of [KP16, Corollary 2], applied in the bounded chart U to the differential operator $\partial_t - \frac{1}{2} \sum_{i=1}^q (E'_i)^2 + B'$ where $E'_i(t) = E_i(-t), B'(t) = -B(-t)$. The time reversal we use here allows to convert the supremum on the left-hand side in the Harnack inequality from [KP16] into an infimum on the right-hand side, as displayed above.

Finally, we put aside any assumption of hypoellipticity on σ and record a moment estimate due to Kisynski [Kis79, Theorem 2] (see also [Jør75, Proposition 3.1]) dealing with homogeneous diffusions. Denote by d_H the distance induced by a left-invariant Riemmanian metric on H .

FACT 3.5 (Quadratic exponential moment [Kis79]). Assume \mathcal{L} is independent of the time parameter (but not necessarily hypoelliptic). Then there exists $\varepsilon > 0$ such that for $x \in H$, all $T > 0$

$$\sup_{t \in [0, T]} \int_H \exp(\varepsilon d_H(x, y)^2) d\sigma_t(y) < +\infty$$

3.2. Invariance and regularity of the limiting diffusion. In this subsection we restrict our attention to a certain family of diffusion processes on nilpotent Lie groups. They are those that arise as limiting processes for right random walks with i.i.d. increments. Using Section 3.1, we prove that every such diffusion is smooth, and give a Gaussian upper bound on its density and each of its derivatives (Proposition 3.8). In the biased case, the diffusion process we have to deal with can be seen in two ways, either as a time-dependent left-invariant diffusion process on \mathfrak{g} with absolutely continuous increments (with generator (23) below), or as the restriction to the affine hyperplane $\mathfrak{g} + t\chi$ of a bona fide left-invariant time-homogeneous diffusion process on the bias extension $\tilde{\mathfrak{g}}$ (Lemma 3.11). Both points of view are good to bear in mind.

Let \mathfrak{g} be a nilpotent Lie algebra, $\bar{X} \in \mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]$ a bias, and $\mathfrak{g} = \bigoplus_b \mathfrak{m}^{(b)}$ an adapted weight decomposition for the weight filtration induced by \bar{X} . We saw in Section 2 that these data define a bias extension $\tilde{\mathfrak{g}}$, a variable χ for which $\tilde{\mathfrak{g}} = \mathfrak{g} + \mathbb{R}\chi$, and graded Lie bracket or Lie product $[\cdot, \cdot]', *'$ on $\tilde{\mathfrak{g}}$.

For the rest of the section, we fix vectors $E_1, \dots, E_q, B, \mathcal{Y} \in \tilde{\mathfrak{g}}$ such that the linear span of $\{E_1, \dots, E_q\}$ is $\mathfrak{m}^{(1)}$, $B \in \mathfrak{m}^{(2)}$, and $\mathcal{Y} \in \chi + \mathfrak{m}^{(2)}$. Our goal is to study the diffusion law σ on $(\mathfrak{g}, *')$ whose infinitesimal generator is given by $\mathcal{L} = (\mathcal{L}_t)_{t \in \mathbb{R}^+}$ where

$$(23) \quad \mathcal{L}_t = \frac{1}{2} \sum_{i=1}^q (\text{Ad}(t\mathcal{Y})E_i)^2 + \text{Ad}(t\mathcal{Y})B$$

Here the notation Ad refers to the adjoint representation for $*'$, namely for every X, Y in $\tilde{\mathfrak{g}}$, $\text{Ad}(Y)X = Y *' X *' (-Y)$. We will occasionally use the notation W to refer to a diffusion process with law σ .

If the random walk on $(\mathfrak{g}, *)$ is driven by a measure μ with finite second moment and mean \bar{X} in the abelianization, then we shall prove (Section 5) that the limiting process has

infinitesimal generator $\mathcal{L} = \mathcal{L}(\mu, (\mathfrak{m}^{(b)}))$ such that E_1, \dots, E_q is a basis of $\mathfrak{m}^{(1)}$ in which the covariance matrix of $\mu^{(1)}$ is the identity, $B = \mathbb{E}_\mu(x^{(2)})$, and $\mathcal{Y} = \chi$. We allow here \mathcal{Y} to be more generally in $\chi + \mathfrak{m}^{(2)}$ in order to consider a class of processes that is stable by change of weight decomposition. This flexibility will be useful later for explicit computations of the support of the time- t distribution σ_t .

LEMMA 3.6 (Changing the weight decomposition). *Let $\mathfrak{g} = \bigoplus_b \mathfrak{m}^{(b)}$ be some other choice of weight decomposition for the weight filtration $(\mathfrak{g}^{(b)})$ induced by \overline{X} . Call $\mathring{\chi}, \mathring{*}'$ the associated objects on $\tilde{\mathfrak{g}}$.*

Let $\phi : \tilde{\mathfrak{g}} \rightarrow \tilde{\mathfrak{g}}$ be the unique linear map such that $\phi : \tilde{\mathfrak{m}}^{(b)} \rightarrow \tilde{\mathfrak{m}}^{(b)}$ is the isomorphism that factorizes via the identity on the vector space $\tilde{\mathfrak{g}}^{(b)}/\tilde{\mathfrak{g}}^{(b+1)}$.

Then ϕ is a group isomorphism from $(\tilde{\mathfrak{g}}, \mathring{}')$ to $(\tilde{\mathfrak{g}}, \mathring{*}')$. It satisfies $\phi(\mathfrak{m}^{(2)}) = \mathring{\mathfrak{m}}^{(2)}$ and $\phi(\chi) \in \mathring{\chi} + \mathring{\mathfrak{m}}^{(2)}$. Moreover, $\phi \circ \sigma$ is the diffusion law on $(\tilde{\mathfrak{g}}, \mathring{*}')$ determined by the parameters $\phi(E_i), \phi(B), \phi(\mathcal{Y})$.*

Remark. $\phi(\chi)$ may not be equal to $\mathring{\chi}$. It follows from the proof that the equality case occurs exactly when the respective lifts of \overline{X} to $\mathfrak{m}^{(1)}$ and $\mathring{\mathfrak{m}}^{(1)}$ have the same projection in the quotient vector space $\mathfrak{g}/\mathfrak{g}^{(3)}$.

PROOF. For the first claim, recall the graded Lie algebra $gr_{\overline{X}}(\tilde{\mathfrak{g}})$ from Section 2. A choice of weight decomposition for \mathfrak{g} yields a vector space isomorphism by $\tilde{\mathfrak{g}}$ to $gr_{\overline{X}}(\tilde{\mathfrak{g}})$, by sending via the identity the subspace of weight b to $\tilde{\mathfrak{g}}^{(b)}/\tilde{\mathfrak{g}}^{(b+1)}$. For the choices $\mathfrak{g} = \bigoplus \mathfrak{m}^{(b)}$, $\mathring{\mathfrak{g}} = \bigoplus \mathring{\mathfrak{m}}^{(b)}$, call those maps pr, \mathring{pr} . The graded bracket $[\cdot, \cdot]'$ is defined so that $pr : (\tilde{\mathfrak{g}}, [\cdot, \cdot]') \rightarrow gr_{\overline{X}}(\tilde{\mathfrak{g}})$ is a morphism of Lie algebras, and the same holds for $[\cdot, \cdot]'$, \mathring{pr} . As $\phi = \mathring{pr}^{-1} \circ pr$, we deduce that ϕ is an isomorphism of Lie algebras from $(\tilde{\mathfrak{g}}, [\cdot, \cdot]')$ to $(\tilde{\mathfrak{g}}, [\cdot, \cdot]')$, which implies the claim.

For the second claim, we observe that $pr(\mathfrak{m}^{(2)}) = \mathfrak{g}^{(2)}/\mathfrak{g}^{(3)} = \mathring{pr}(\mathring{\mathfrak{m}}^{(2)})$, so in particular $\phi(\mathfrak{m}^{(2)}) = \mathring{\mathfrak{m}}^{(2)}$. Now, slightly abusing notations, denote by $X \in \mathfrak{m}^{(1)}$, $\mathring{X} \in \mathring{\mathfrak{m}}^{(1)}$ the respective lifts of $\overline{X} \pmod{[\mathfrak{g}, \mathfrak{g}]}$. One can write $X = \mathring{X} + \mathring{Y} + \mathring{Z}$ where $\mathring{Y} \in \mathring{\mathfrak{m}}^{(2)}$, $\mathring{Z} \in \mathring{\mathfrak{g}}^{(3)}$. This equality can be rewritten as $\chi = \mathring{\chi} + \mathring{Y} + \mathring{Z}$, and by definition of ϕ , we get

$$\phi(\chi) = \mathring{\chi} + \mathring{Y}$$

The third claim is a straightforward consequence of the first claim. The second claim ensures that the type of diffusions we consider is indeed preserved by ϕ . \square

We now stick with the given weight decomposition $\mathfrak{g} = \bigoplus_b \mathfrak{m}^{(b)}$ and present invariance properties for the diffusion σ . Recall that D_r denotes the dilation $D_r = \bigoplus_b r^b \text{Id}_{\mathfrak{m}^{(b)}}$.

PROPOSITION 3.7 (Invariance). *The diffusion law σ associated to \mathcal{L} (defined in eq. (23)) satisfies for all $r, s, t > 0$,*

$$\sigma_s \mathring{*}' \text{Ad}(s\mathcal{Y})\sigma_t = \sigma_{s+t} \qquad D_{\sqrt{r}}\sigma_t = \sigma_{rt}$$

PROOF. For the first identity on the left hand side, it is sufficient to check that the process $(\text{Ad}(s\mathcal{Y})W(t))_{t \in \mathbb{R}^+}$ is the diffusion with infinitesimal generator $(\mathcal{L}_{s+t})_{t \in \mathbb{R}^+}$. This follows from definitions and the straightforward computation

$$\mathcal{L}_t(f \circ \text{Ad}(s\mathcal{Y})) = \mathcal{L}_{s+t}(f) \circ \text{Ad}(s\mathcal{Y})$$

For the identity on the right hand side, it is sufficient to check that the diffusion processes $(W(rt))_{t \in \mathbb{R}^+}$ and $(D_{\sqrt{r}}W(t))_{t \in \mathbb{R}^+}$ have the same infinitesimal generator $(r\mathcal{L}_{rt})_{t \in \mathbb{R}^+}$. This

follows on the one hand by a simple change of variables, and on the other hand from the straightforward computation

$$\mathcal{L}_t(f \circ D_{\sqrt{r}}) = r \mathcal{L}_{rt}(f) \circ D_{\sqrt{r}}$$

□

We now study the regularity of σ_t . We denote by $|x|'_{\mathfrak{g}}$ the distance between 0 and x for a fixed left-invariant Riemannian metric on $(\tilde{\mathfrak{g}}, *')$. For $\alpha = (\alpha_j^i)_{i,j} \in \mathbb{N}^d$, we write ∂^α the operator differentiating (for the addition) α_j^i times in each direction $e_j^{(i)}$ (see Section 2.6 for this notation). We also set $\|\alpha\| = \sum_{i,j} i \alpha_j^i$ and recall that $d_{\overline{X}}$ is the *homogeneous dimension* of the weight filtration induced by \overline{X} (Equation (18)).

PROPOSITION 3.8 (Regularity). *The diffusion law σ associated to \mathcal{L} can be written*

$$\sigma_t = u(t, x) dx$$

where $u : \mathbb{R}_{>0} \times \mathfrak{g} \rightarrow \mathbb{R}^+$ is a smooth function, and for each $\alpha \in \mathbb{N}^d$, there exists $C > 0$ such that for all $t > 0$, $x \in \mathfrak{g}$,

$$(24) \quad \partial^\alpha u(t, x) \leq C t^{-\frac{1}{2}(d_{\overline{X}} + \|\alpha\|)} e^{-\frac{1}{C} |D_{\frac{1}{\sqrt{t}}} x|'_{\mathfrak{g}}{}^2}.$$

Remarks.

1) The distance $| \cdot |'_{\mathfrak{g}}$ depends on the choice of grading $\mathfrak{m}^{(b)}$. However any two left invariant geodesic metrics on a connected Lie group are coarsely equivalent and in our setting (*ball box principle*, see e.g. [Bel96]) for every Euclidean norm $\| \cdot \|$ on $\tilde{\mathfrak{g}}$, there exists $\varepsilon \in (0, 1]$ such that

$$\varepsilon \| \cdot \|^\varepsilon \leq | \cdot |'_{\mathfrak{g}} \leq \varepsilon^{-1} \| \cdot \|$$

outside of a compact neighbourhood of the origin in \mathfrak{g} . In particular, (24) implies that $u(t, x)$ is a *Schwartz function* in x and eq. (13) holds.

2) We will see in the proof of Proposition 3.8 that (24) continues to hold for iterated right derivatives with respect to the product $*'$ in place of additive derivatives.

3) The exponential rate of decay in the upper bound is probably not sharp in general. We expect that a $*'$ -left invariant Riemannian metric on \mathfrak{g} can be taken in place of the distance $| \cdot |'_{\mathfrak{g}}$ in eq. (24). This is what happens in the Heisenberg group case or whenever one is in the situation of Theorem 3.33. Not that it would be futile to obtain a similar looking lower bound in general, because the support of σ_t may not be full in general as we shall see.

Plan of the proof. The argument to prove Proposition 3.8 will be in two parts. First, we use Fact 3.3 to show that the diffusion σ is hypoelliptic, justifying the smoothness of σ_t and yielding also a principle of domination derivative-integral, that bounds the value of any fixed derivative of u at a point by the integral of u on a small neighborhood. Second, we show that σ_t has quadratic exponential moment for $| \cdot |'_{\mathfrak{g}}$, by observing that σ is the projection to \mathfrak{g} of some homogeneous (though degenerate) diffusion law on $(\tilde{\mathfrak{g}}, *')$ and applying Fact 3.5.

The hypoellipticity of σ relies on the following geometrical statement.

LEMMA 3.9. *We have the equality $[\tilde{\mathfrak{g}}, \tilde{\mathfrak{g}}]' = \mathfrak{g}^{(2)}$. In particular, $(\tilde{\mathfrak{g}}, [\cdot, \cdot]')$ is generated as a Lie algebra by $\mathfrak{m}^{(1)} \oplus \mathbb{R}\mathcal{Y}$.*

PROOF. The inclusion $[\tilde{\mathfrak{g}}, \tilde{\mathfrak{g}}]' \subseteq \mathfrak{g}^{(2)}$ follows directly from the definition of $[\cdot, \cdot]'$. For the converse inclusion, we show by induction that $\mathfrak{g}^{(b)} \subseteq [\tilde{\mathfrak{g}}, \tilde{\mathfrak{g}}]'$ for all $b \in \{2, \dots, 2s\}$. The initial step $\{0\} = \mathfrak{g}^{(2s)} \subseteq [\tilde{\mathfrak{g}}, \tilde{\mathfrak{g}}]'$ is clear. Now assume $\mathfrak{g}^{(b+1)} \subseteq [\tilde{\mathfrak{g}}, \tilde{\mathfrak{g}}]'$. We consider $x \in \mathfrak{g}^{(b)}$ and show $x \in [\tilde{\mathfrak{g}}, \tilde{\mathfrak{g}}]'$. By definition of $\mathfrak{g}^{(b)}$, we may suppose that x is a (non-graded) bracket $x = [v_1, [v_2, [\dots, [v_{a-1}, v_a] \dots]]]$ ($a \geq 2$) where the v_i 's belong to $\mathfrak{m}^{(1)}$, and if j denotes the number of occurrences of X among them, then $a + j \geq b$. As $[\chi, \cdot] = [X, \cdot]$, we may replace each v_i by $w_i = v_i \mathbb{1}_{v_i \neq X} + \chi \mathbb{1}_{v_i = X}$ without altering the value of the total bracket. By induction hypothesis, we can further assume $a + j = b$. Now $x = [w_1, [w_2, [\dots, [w_{a-1}, w_a] \dots]]]$ coincides with $[w_1, [w_2, [\dots, [w_{a-1}, w_a]' \dots]]']'$ modulo $\mathfrak{g}^{(b+1)}$. As $\mathfrak{g}^{(b+1)} \subseteq [\tilde{\mathfrak{g}}, \tilde{\mathfrak{g}}]'$ by induction hypothesis, it follows that $x \in [\tilde{\mathfrak{g}}, \tilde{\mathfrak{g}}]'$. This proves the induction step, and the claim $[\tilde{\mathfrak{g}}, \tilde{\mathfrak{g}}]' = \mathfrak{g}^{(2)}$.

The second claim follows because $\mathfrak{m}^1 \oplus \mathbb{R}\mathcal{Y}$ projects surjectively to $\tilde{\mathfrak{g}}/\mathfrak{g}^{(2)}$, which we just established to be the abelianization of $(\tilde{\mathfrak{g}}, [\cdot, \cdot]')$. \square

We combine Lemma 3.9 and Fact 3.3 to obtain

LEMMA 3.10. *The vector fields $\text{Ad}(t\mathcal{Y})E_1, \dots, \text{Ad}(t\mathcal{Y})E_q, \partial_t + \text{Ad}(t\mathcal{Y})B$ satisfy the Hörmander condition from Fact 3.3. In particular, the differential operator*

$$\partial_t - \frac{1}{2} \sum_{i=1}^q (\text{Ad}(t\mathcal{Y})E_i)^2 + \text{Ad}(t\mathcal{Y})B$$

is hypoelliptic on $\mathbb{R}_{>0} \times \mathfrak{g}$.

Remember that for $X \in \mathfrak{g}$, the notation $\text{Ad}(t\mathcal{Y})X$ refers to the vector field $\mathbb{R}_{>0} \times \mathfrak{g} \rightarrow T\mathfrak{g}, (t, x) \mapsto TL_x(\text{Ad}(t\mathcal{Y})X)$ where $L_x : y \mapsto x *' y$. It is usually not left-invariant on the product Lie group $\mathbb{R}_{>0} \times \mathfrak{g}$.

PROOF. We start with a few useful observations concerning the Lie brackets of vector fields on the manifold $\mathbb{R}_{>0} \times \mathfrak{g}$. Such brackets are denoted by $[\cdot, \cdot]$. The notation $[\cdot, \cdot]'$ is reserved for the graded bracket of elements in $\tilde{\mathfrak{g}}$, seen as vectors and not vector fields. First, note that

$$[\partial_t, \text{Ad}(t\mathcal{Y})X] = \sum_{k=1}^s [\partial_t, \frac{t^k}{k!} \text{ad}(\mathcal{Y})^k X]$$

where $\text{ad}(\mathcal{Y}) := [\mathcal{Y}, \cdot]'$. Combined with the general formula $[Y, \varphi Z] = \varphi[Y, Z] + Y.\varphi Z$, and the equality $[\partial_t, Z] = 0$ for every vector field Z on \mathfrak{g} extended to $\mathbb{R}_{>0} \times \mathfrak{g}$ by \mathbb{R} -invariance, we get

$$[\partial_t, \text{Ad}(t\mathcal{Y})X] = \sum_{k=1}^s \frac{t^{k-1}}{(k-1)!} \text{ad}(\mathcal{Y})^k X = \text{Ad}(t\mathcal{Y})[\mathcal{Y}, X]'$$

Write $\mathfrak{g}^{[i]}$ the descending central filtration of $(\mathfrak{g}, [\cdot, \cdot]')$, and similarly for $\tilde{\mathfrak{g}}$. It follows that if X belongs to $\mathfrak{g}^{[i]}$, noting that $B \in [\tilde{\mathfrak{g}}, \tilde{\mathfrak{g}}]'$ (Lemma 3.9), we can rewrite the bracket vector field

$$[\partial_t + \text{Ad}(t\mathcal{Y})B, \text{Ad}(t\mathcal{Y})X] = \text{Ad}(t\mathcal{Y})([\mathcal{Y}, X]' + Z)$$

where $Z := [B, X]' \in [\tilde{\mathfrak{g}}^{[2]}, X]'$ is a vector in \mathfrak{g} .

We now check Hörmander's condition from Fact 3.3. Call \mathcal{H} the Lie algebra generated by the vector fields $\text{Ad}(t\mathcal{Y})E_1, \dots, \text{Ad}(t\mathcal{Y})E_q, \partial_t + \text{Ad}(t\mathcal{Y})B$ on $\mathbb{R}_{>0} \times \mathfrak{g}$. Write $\mathfrak{h} \subseteq \mathfrak{g}$ the subset of vectors $X \in \mathfrak{g}$ such that the vector field corresponding to $\text{Ad}(t\mathcal{Y})X$ is in \mathcal{H} . It is

enough to check that $\mathfrak{h} = \mathfrak{g}$. Note that \mathfrak{h} contains E_1, \dots, E_q by definition. \mathfrak{h} is also a sub Lie algebra of $(\mathfrak{g}, [\cdot, \cdot]')$, and in view of the previous paragraph, if $X \in \mathfrak{h}$ then $[\mathcal{Y}, X]'$ is in $\mathfrak{h} + [\widetilde{\mathfrak{g}}^{[2]'}, X]'$. By Lemma 3.9, this forces $\mathfrak{h} = \mathfrak{g}$ and concludes the proof. \square

LEMMA 3.11 (Homogenisation). *Let λ be the homogeneous diffusion law on $(\widetilde{\mathfrak{g}}, *')$ associated to the differential operator*

$$\widetilde{\mathcal{L}} = \frac{1}{2} \sum_{i=1}^q E_i^2 + (B + \mathcal{Y})$$

If W is a diffusion process on \mathfrak{g} with law σ , then $Z(t) := W(t) *' t\mathcal{Y}$ is a diffusion process on $\widetilde{\mathfrak{g}}$ with law λ .

Remarks. 1) The diffusion λ is not hypoelliptic.

2) Writing $\mathcal{L}_+ := \frac{1}{2} \sum_{i=1}^q \text{Ad}(t(B + \mathcal{Y})) E_i^2$, we see that $\widetilde{\mathcal{L}}_+ = \widetilde{\mathcal{L}}$. In particular Lemma 3.11 implies that the process $W(t) *' t\mathcal{Y} *' (-t(B + \mathcal{Y}))$ is a diffusion on $(\mathfrak{g}, *')$ with infinitesimal generator \mathcal{L}_+ . This trick allows to suppress the drift parameter in the generator. It will be useful later for explicit computations of the support of σ (Section 3.3).

PROOF. We start by noticing that for every $s \geq 0$, if $F \in C_c^\infty(\mathbb{R}^+ \times \mathfrak{g})$ is a \mathcal{F}_s -measurable variable, then the stochastic process $(I(t))_{t \geq s}$ given by

$$(25) \quad I(t) := F(t, c(t)) - F(s, c(s)) - \int_s^t (\partial_{e_1} + \mathcal{L}_r) F(r, c(r)) dr$$

is a martingale on $(\Omega(\mathfrak{g}), (\mathcal{F}_t)_{t \geq s}, \sigma)$. Here ∂_{e_1} denotes the partial derivative with respect to the \mathbb{R}^+ variable. Indeed, for some martingales $(M_i(t))_{i \leq 3}$, we have

$$\begin{aligned} & F(t, c(t)) - F(t, c(s)) - \int_s^t (\mathcal{L}_r) F(r, c(r)) dr \\ &= \int_s^t \mathcal{L}_r [F(t, \cdot) - F(r, \cdot)](c(r)) dr + M_1(t) \\ &= \int_s^t \mathcal{L}_r \left[\int_r^t \partial_{e_1} F(z, \cdot) dz \right] (c(r)) dr + M_1(t) \\ &= \int_s^t \int_r^t \mathcal{L}_r [\partial_{e_1} F(z, \cdot)] dz (c(r)) dr + M_1(t) \\ &= \int_s^t \int_s^z \mathcal{L}_r [\partial_{e_1} F(z, \cdot)] (c(r)) dr dz + M_1(t) \\ &= \int_s^t \partial_{e_1} F(z, c(z)) - \partial_{e_1} F(z, c(s)) + M_2(z) dz + M_1(t) \\ &= \int_s^t \partial_{e_1} F(z, c(z)) dz - F(t, c(s)) + F(s, c(s)) + M_3(t) \end{aligned}$$

which rewrites as $I(t) = M_3(t)$.

Now set $F(s, x) = f(x *' s\mathcal{Y})$ where $f \in C_c^\infty(\mathfrak{g})$ is a \mathcal{F}_s -measurable variable. Observe that

$$(26) \quad \partial_{e_1} F(r, c(r)) = (\mathcal{Y}.f)(c(r)) *' r\mathcal{Y}$$

and for any vector $X \in \mathfrak{g}$,

$$(27) \quad (X.F)(r, c(r)) = (\text{Ad}(-r\mathcal{Y})X.f)(c(r) *' r\mathcal{Y})$$

The result now follows by combining (25), (26), (27). \square

We now combine Section 3.1, Lemmas 3.10, 3.11 to prove Proposition 3.8.

PROOF OF PROPOSITION 3.8. From Lemma 3.10 and the background material from Section 3.1, one may write $\sigma_t = u(t, x)dx$ where $u : \mathbb{R}_{>0} \times \mathfrak{g} \rightarrow \mathbb{R}^+$ is a smooth function. The relation $\sigma_t = D_{\sqrt{t}}\sigma_1$ obtained in Proposition 3.7 can be restated as

$$(28) \quad u(t, x) = t^{-\frac{d_{\overline{X}}}{2}} u(1, D_{\frac{1}{\sqrt{t}}}x)$$

As $D_{\frac{1}{\sqrt{t}}}$ acts linearly on \mathfrak{g} , and $D_{\frac{1}{\sqrt{t}}}e_j^{(i)} = t^{-i/2}e_j^{(i)}$, the upper bound (24) reduces to the case where $t = 1$.

Notice also that, in order to prove the upper bound (24), we may replace the additive derivative $\partial^\alpha u(t, x)$ by an iterated multiplicative derivative $(\underline{X}u)(1, x)$ where $\underline{X} = X_1 \dots X_k$ is a finite sequence of vectors in \mathfrak{g} . Indeed, every additive derivation with respect to some $e_j^{(i)}$ can be written as a combination with polynomial coefficients (in the x variable) of right derivations for $*'$, and this remains true under iteration. The polynomial coefficients appearing do not matter, as they are absorbed by the exponential term (here $t = 1$).

We introduce the bounded neighborhood of $(1, 0) \in \mathbb{R} \times \mathfrak{g}$ given by $\Omega :=]\frac{1}{2}, \frac{3}{2}[\times \{x \in \mathfrak{g}, |x|_{\mathfrak{g}}' < 1\}$. For $x \in \mathfrak{g}$, $(t, y) \mapsto u(t, x *' y)$ is a solution to 21. Therefore Equation (22) guarantees the existence of a constant $C_1 > 0$ that is independent of x and for which

$$(29) \quad |\underline{X}u(1, x)| \leq C_1 \int_{\Omega} u(t, x *' y) dt dy$$

On the other hand, the combination of Fact 3.5 applied to $H := (\tilde{\mathfrak{g}}, *')$ and Lemma 3.11 implies that

$$I := \sup_{t \in [0, 3/2]} \int_{\mathfrak{g}} \exp(\varepsilon |y *' t\chi|_{\mathfrak{g}}'^2) u(t, y) dy < \infty$$

for some $\varepsilon > 0$. For $(t, y) \in \Omega$, we have $|x *' y *' t\chi|_{\mathfrak{g}}' \geq |x|_{\mathfrak{g}}'/2$ whenever $|x|_{\mathfrak{g}}' \geq 2(1 + \sup_{t \in [0, 3/2]} |t\chi|_{\mathfrak{g}}')$. Hence

$$\int_{\Omega} u(t, x *' y) dt dy \leq I \exp(-\frac{\varepsilon}{4} |x|_{\mathfrak{g}}'^2)$$

Finally, from eq. (29) we get:

$$|\underline{X}u(1, x)| \leq C_1 I \exp(-\frac{\varepsilon}{4} |x|_{\mathfrak{g}}'^2)$$

which concludes the proof. \square

3.3. Support of the limiting diffusion. We keep the notations of Section 3.2 and describe the support of the distribution σ_1 . After a general and explicit description, we turn to two remarkable examples: for the upper triangular nilpotent Lie algebra, the support of σ_1 is always the full Lie algebra, whereas for non-centered walks (i.e. $\overline{X} \neq 0$) on a free nilpotent Lie algebra of step at least 3, the support is always a proper closed subset. The emergence of a limiting diffusion without full support is unexpected, and specific to the non-centered case.

Let us begin with a general description of the support \mathcal{S} of σ_1 . It is clearly a closed subset of \mathfrak{g} , with dense interior because σ_1 is smooth. We use the *support theorem* (Fact 3.2) to give an explicit description in our context.

Recall that if $I \subseteq \mathbb{R}$ is an open interval and $\gamma : I \rightarrow \mathfrak{g}$ is C^1 , we may define its (multiplicative) derivative at $t \in I$ by

$$\partial\gamma(t) = \lim_{\substack{\varepsilon \rightarrow 0 \\ \varepsilon \neq 0}} \frac{1}{\varepsilon} \gamma(t)^{-1} *' \gamma(t + \varepsilon)$$

When $*'$ is abelian, it is just the usual derivative $\gamma'(t)$, and we have in general the relation $\gamma'(t) = TL_{\gamma(t)}\partial\gamma(t)$ where $TL_{\gamma(t)}$ denotes the tangent map of the left $*'$ -multiplication by $\gamma(t)$.

A theorem of Strichartz [Str87] ensures that the derivation ∂ admits an inverse, generalizing the classical Riemann integral in the abelian case, expressed in terms of *iterated integrals* ([Che77]) as follows. Given a bounded and measurable path $c : [a, b] \rightarrow \mathfrak{g}$, set

$$\begin{aligned} \int_{[a,b]}^{*'} c &:= \sum_{r=1}^s \sum_{\tau \in \mathfrak{S}_r} \frac{(-1)^{e(\tau)}}{r^2 \binom{n-1}{e(\tau)}} \int_{\{a < t_1 < \dots < t_r < b\}} L^{[r]'}(c(t_{\tau(1)}), \dots, c(t_{\tau(r)})) dt_1 \dots dt_r \\ (30) \quad &= \int_{[a,b]} c(t) dt + \frac{1}{2} \int_{\{a < t_1 < t_2 < b\}} [c(t_1), c(t_2)]' dt_1 dt_2 \\ &+ \frac{1}{6} \int_{\{a < t_1 < t_2 < t_3 < b\}} [[c(t_1), c(t_2)]', c(t_3)]' + [[c(t_3), c(t_2)]', c(t_1)]' dt_1 dt_2 dt_3 + \dots \end{aligned}$$

where $L^{[r]}'$ refers to r iterations of the bracket $[\cdot, \cdot]'$ (see Section 2.5) and $e(\tau)$ represents the number of consecutive subscripts whose order is reversed under the permutation τ , i.e. $e(\tau) = \#\{i \in \{1, \dots, r-1\}, \tau(i) > \tau(i+1)\}$.

FACT 3.12 (Strichartz [Str87]). The map that sends c to $\gamma(t) = \int_{[a,t]}^{*'} c(s) ds$ is a bijection between the set of piecewise continuous paths c on $[a, b]$ and the set of continuous piecewise C^1 paths γ on $[a, b]$. The inverse is given by the multiplicative derivative $\gamma \mapsto \partial\gamma$.

Recall from Section 3.1 that a piecewise C^1 path $\gamma : [a, b] \rightarrow \mathfrak{g}$ is said to be \mathcal{L} -horizontal if

$$\partial\gamma(t) \in \text{Ad}(t\mathcal{Y})(\mathfrak{m}^{(1)} + B)$$

for any t outside of the finite set of discontinuity points of $\partial\gamma$. Combining Fact 3.2 and Fact 3.12, we get

LEMMA 3.13. *The support of σ_1 is the closure of the set*

$$\left\{ \int_{[0,1]}^{*'} \text{Ad}(t\mathcal{Y})(u(t) + B) dt : u : [0, 1] \rightarrow \mathfrak{m}^{(1)} \text{ piecewise continuous} \right\}$$

This formula has the advantage of reducing the computation of the support to the problem of solving a finite explicit family of polynomial equations in the variable v . We will use this point of view to show that walks on the upper triangular unipotent subgroup always have a fully supported limit distribution.

We also present a discrete characterization of the support of σ_1 that will be useful below to describe an example with non full support. Given vectors $u_1, \dots, u_k \in \tilde{\mathfrak{g}}$, we set $\prod_{1 \leq i \leq k}^{*'} u_i = u_1 *' \dots *' u_k$.

LEMMA 3.14. *The support of σ_1 is the closure of the set*

$$\left\{ \prod_{1 \leq i \leq k}^{*'} (u_i *' t_i (B + \mathcal{Y})) : k \geq 1, u_i \in \mathfrak{m}^{(1)}, t_i \geq 0, \sum_{i=1}^k t_i = 1 \right\} *' (-\mathcal{Y})$$

PROOF. Case $B = 0$. Call \mathcal{S} the support of σ_1 and \mathcal{S}' the closure of the set above.

We first check the inclusion $\mathcal{S} \subseteq \mathcal{S}'$. It is clear from the formula defining the multiplicative integral $\int^{*'}$ that for every smooth $u : [0, 1] \rightarrow \mathfrak{m}^{(1)}$,

$$\begin{aligned} \int_{[0,1]}^{*'} \text{Ad}(t\mathcal{Y})u(t) dt &= \prod_{0 \leq i \leq k-1}^{*'} \int_{[\frac{i}{k}, \frac{i+1}{k}]}^{*'} \text{Ad}(t\mathcal{Y})u(t) dt \\ &= \prod_{0 \leq i \leq k-1}^{*'} \left(\frac{1}{k} \text{Ad}\left(\frac{i}{k}\mathcal{Y}\right)u\left(\frac{i}{k}\right) + o_u(k^{-1}) \right) \\ &= \prod_{0 \leq i \leq k-1}^{*'} \left(\text{Ad}\left(\frac{i}{k}\mathcal{Y}\right)\left(\frac{1}{k}u\left(\frac{i}{k}\right)\right) \right) + o_u(1) \\ &= \prod_{0 \leq i \leq k-1}^{*'} \left(\frac{1}{k}u\left(\frac{i}{k}\right) *' \frac{\mathcal{Y}}{k} \right) *' (-\mathcal{Y}) + o_u(1) \end{aligned}$$

Letting k go to infinity, we get that $\int_{[0,1]}^{*'} \text{Ad}(t\mathcal{Y})u(t) \in \mathcal{S}'$. By continuous approximation and the formula defining $\int^{*'}$, we still have $\int_{[0,1]}^{*'} \text{Ad}(t\mathcal{Y})u(t) \in \mathcal{S}'$ if u is only piecewise-continuous. We then deduce from Lemma 3.13 that $\mathcal{S} \subseteq \mathcal{S}'$.

For the reverse inclusion, one may assume that $t_i > 0$ for all i . Let $0 < \varepsilon < \min_i t_i$. Set $\bar{t}_i = t_1 + \dots + t_{i-1}$ and define a piecewise continuous function u_ε by

$$u_\varepsilon(t) = \begin{cases} \varepsilon^{-1}u_i & \text{if } t \in [\bar{t}_i, \bar{t}_i + \varepsilon) \\ 0 & \text{otherwise} \end{cases}$$

Then $\int_{[0,1]}^{*'} \text{Ad}(t\mathcal{Y})u_\varepsilon(t) dt \rightarrow \prod_{1 \leq i \leq k}^{*'} (u_i *' t_i \mathcal{Y}) *' (-\mathcal{Y})$ as ε goes to zero, yielding $\mathcal{S}' \subseteq \mathcal{S}$.

General case. Call $W_+(t)$ a diffusion process with infinitesimal generator $\mathcal{L}_+ = \frac{1}{2} \sum_i \text{Ad}(t(B + \mathcal{Y}))E_i^2$. It follows from Lemma 3.11 that the stochastic processes $W_+(t) *' t(B + \mathcal{Y})$ and $W(t) *' t\mathcal{Y}$ have same distribution. As \mathcal{L}_+ has no drift, we may apply the previous discussion to characterize the support of $W_+(1) *' (B + \mathcal{Y})$, in other terms $\mathcal{S} *' \mathcal{Y}$, and the result follows. \square

3.3.1. *Example 1: the case of upper triangular matrices.* We assume here that $(\mathfrak{g}, [\cdot, \cdot])$ satisfies the following double cancellation assumption

(DC) *There exist vectors $v_1, \dots, v_q \in \mathfrak{g}$ spanning $\mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]$ such that for all $1 \leq i \leq q$, $k \geq 0$, $w \in \mathfrak{g}$, one has*

$$[v_i, \text{ad}(w)^k v_i] = 0$$

and we show that the support of σ_1 must be \mathfrak{g} . Here the notation ad refers to the non-graded bracket $[\cdot, \cdot]$, i.e. $\text{ad}(w)x = [w, x]$. Condition **(DC)** clearly holds in every nilpotent Lie algebra with step at most 2. As we show now it holds for the group of upper triangular unipotent matrices.

LEMMA 3.15. *Assume \mathfrak{g} is the Lie algebra of a maximal unipotent subgroup of a special linear group $SL_n(\mathbb{R})$. Then \mathfrak{g} satisfies condition **(DC)**.*

PROOF. In this context, \mathfrak{g} admits a vector space basis $(E_{i,j})_{1 \leq i < j \leq q_1+1}$ with the bracket relations

$$[E_{i,j}, E_{k,\ell}] = \mathbb{1}_{j=k} E_{i,\ell} - \mathbb{1}_{i=\ell} E_{k,j}$$

Setting $v_i = E_{i,i+1}$, the claim follows by direct computation. \square

The main result of the section is Proposition 3.16 below, which implies full support under condition **(DC)**, and will even yield a lower bound on σ_1 later in Section 3.4. Given a \mathcal{L} -horizontal path $\gamma : I \rightarrow \mathfrak{g}$, we introduce its *driving vector* $v_\gamma : I \rightarrow \mathfrak{m}^{(1)}$ defined (outside a finite subset) by the relation $\partial\gamma(t) = \text{Ad}(t\mathcal{Y})(v_\gamma(t) + B)$. We call *horizontal energy* of γ the quantity

$$E_{hor}(\gamma) = \int_I \|v_\gamma(t)\|^2 dt$$

It is related to the classical notion of energy, but better suited for computations and more meaningful in our context. Indeed it expresses how hard it is for the \mathcal{L} -diffusion to follow γ . We fix a left-invariant Riemannian metric on $(\mathfrak{g}, *')$ and denote by $|\cdot|'_\mathfrak{g}$ the distance to the origin.

PROPOSITION 3.16. *Assume **(DC)**. Then for every point $x \in \mathfrak{g}$, there exists a continuous, piecewise C^1 , \mathcal{L} -horizontal path $\gamma : [0, 1] \rightarrow \mathfrak{g}$ such that $\gamma(0) = 0$, $\gamma(1) = x$. Moreover, one may choose γ such that $E_{hor}(\gamma) \leq C|x|'_\mathfrak{g}^2 + C$ where $C > 0$ is a constant independent from x .*

Combining Proposition 3.16 and Fact 3.2 we get

COROLLARY 3.17 (Full support). *Under condition **(DC)**, the support of σ_1 is \mathfrak{g} .*

We now engage in the proof of Proposition 3.16. In accordance with condition **(DC)**, we fix $v_1, \dots, v_{q_1} \in \mathfrak{g}$ projecting to a basis of $\mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]$ and satisfying $[v_i, \text{ad}(w)^k v_i] = 0$ for all $w \in \mathfrak{g}$, $k \geq 0$.

LEMMA 3.18. *In order to prove Proposition 3.16, one may assume*

$$v_1, \dots, v_{q_1} \in \mathfrak{m}^{(1)} \quad B = 0$$

PROOF. Let $\mathfrak{g} = \bigoplus_b \mathfrak{m}^{(b)}$ be some other weight decomposition of \mathfrak{g} such that $\mathfrak{m}^{(1)} = \bigoplus \mathbb{R}v_i$, and $\phi : (\tilde{\mathfrak{g}}, [\cdot, \cdot]') \rightarrow (\tilde{\mathfrak{g}}, [\cdot, \cdot]')$ the isomorphism of graded structures from Lemma 3.6. The map ϕ sends a left invariant metric on $(\mathfrak{g}, *')$ to a left-invariant metric on $(\tilde{\mathfrak{g}}, \tilde{*}')$, and any \mathcal{L} -horizontal path γ on $(\mathfrak{g}, *')$ to the $\phi(\mathcal{L})$ -horizontal path $\phi \circ \gamma$ on $(\tilde{\mathfrak{g}}, \tilde{*}')$, which also satisfies

$E_{hor}(\gamma) \ll_{\phi} E_{hor}(\phi \circ \gamma) \ll_{\phi} E_{hor}(\gamma)$. Hence, it is enough to prove Proposition 3.16 for the process generated by $\phi(\mathcal{L})$, in other words one may assume $v_1, \dots, v_{q_1} \in \mathfrak{m}^{(1)}$.

To remove the drift, observe that γ is \mathcal{L} -horizontal if and only if the path $c : t \mapsto \gamma(t) *' t\mathcal{Y} *' -t(B + \mathcal{Y})$ is \mathcal{L}_+ -horizontal where $\mathcal{L}_+ = \frac{1}{2} \sum_i (\text{Ad}(t(\mathcal{Y} + B))E_i)^2$. A direct computation shows that γ and c have the same driving vector (for their respective generators \mathcal{L} , \mathcal{L}_+), in particular they have the same horizontal energy. We infer that it is sufficient to prove Proposition 3.16 for the process generated by \mathcal{L}_+ , in other words one may assume $B = 0$. \square

Say that a vector $w \in \mathfrak{g}$ satisfies the *property (P)* if for every non-trivial interval $[a, b] \subseteq \mathbb{R}$, there exists a smooth function $u : [a, b] \rightarrow \mathfrak{m}^{(1)}$ such that for all $r \in \mathbb{R}$,

$$(31) \quad \int_{[a,b]}^{*' } \text{Ad}(t\mathcal{Y})(ru(t)) dt = rw$$

LEMMA 3.19. *The elements v_1, \dots, v_{q_1} satisfy the property (P).*

We recall that throughout the paper, Ad refers implicitly to the graded structure $*'$ at study, i.e. $\text{Ad}(y)x = y *' x *' - y = \sum_{n=0}^s (n!)^{-1} L^{[n+1]'}(y^{\otimes n}, x)$ for all $x, y \in \tilde{\mathfrak{g}}$.

PROOF. Let $v \in \{v_1, \dots, v_{q_1}\}$. The crucial¹ input of the double cancellation property is that it implies that for any $t_0, t_1 \in \mathbb{R}$,

$$(32) \quad [\text{Ad}(t_0\mathcal{Y})v, \text{Ad}(t_1\mathcal{Y})v]' = 0$$

Indeed, one may assume $t_0 = 0$, then expanding $\text{Ad}(t_1\mathcal{Y})$ it is enough to check that $L^{[n+2]'}(v, \mathcal{Y}^{\otimes n}, v) = 0$ for any $n \geq 0$. As $v \in \mathfrak{m}^{(1)}$, $\mathcal{Y} \in \mathfrak{m}^{(2)}$, we have by definition that $L^{[n+2]'}(v, \mathcal{Y}^{\otimes n}, v)$ is the projection to $\mathfrak{m}^{(2+2n)}$ of the non-graded bracket $L^{[n+2]}(v, \mathcal{Y}^{\otimes n}, v)$ which is equal to zero by assumption **(DC)**. Hence (32).

Now fix a non-trivial interval $[a, b] \subseteq \mathbb{R}$. We look for u in the form $u(t) = \alpha(t)v$ where $\alpha : [a, b] \rightarrow \mathbb{R}$ is a smooth function. For any such u , it follows from Strichartz' formula (30) defining $\int^{*'}$ and (32) above that

$$\begin{aligned} \int_{[a,b]}^{*' } \text{Ad}(t\mathcal{Y})(u(t)) dt &= \int_{[a,b]} \text{Ad}(t\mathcal{Y})(u(t)) dt \\ &= \sum_{n=0}^s \left(\int_{[a,b]} \alpha(t)t^n dt \right) (n!)^{-1} L^{[n+1]'}(\mathcal{Y}^{\otimes n}, v) \end{aligned}$$

where the first line converts the multiplicative integral into an additive integral, in the classical sense. To conclude that v satisfies (P), we just need to choose α such that

$$\int_{[a,b]} \alpha(t)t^n dt = \begin{cases} 1 & \text{if } n = 0 \\ 0 & \text{if } n \in \{1, \dots, s\} \end{cases}$$

The existence of such α is clear by independence of the linear forms $\alpha \mapsto \int_{[a,b]} \alpha(t)t^n dt$ on $C^\infty([a, b], \mathbb{R})$. \square

We use Lemma 3.19 to generate more elements that satisfy (P).

¹Indeed, we only use **(DC)** to get (32) for every v in some basis of $\mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]$ lifted to \mathfrak{g} .

LEMMA 3.20 (horizontal pieces). *There exist elements $w_1, \dots, w_p \in \mathfrak{g}$ such that (w_i) project to a basis of $\mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]'$ and each w_i satisfies the property (P).*

PROOF. This follows from Lemma 3.19 and the two next observations:

- If w satisfies (P), then $\text{Ad}(t_0\mathcal{Y})w$ also satisfies (P) for any $t_0 \in \mathbb{R}$.
- The sub Lie algebra of $(\mathfrak{g}, [\cdot, \cdot]')$ generated by $\{\text{Ad}(t\mathcal{Y})\mathfrak{m}^{(1)}, t \in \mathbb{R}\}$ contains $\mathfrak{m}^{(1)}$ and is stable by $[\mathcal{Y}, \cdot]'$, so it must be \mathfrak{g} by Lemma 3.9. It follows that the abelianization $\mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]'$ is spanned by the projection of $\{\text{Ad}(t\mathcal{Y})v_i, t \in \mathbb{R}, i \leq q_1\}$.

□

The next lemma claims that we can combine the horizontal paths generating the $\mathbb{R}w_i$'s via the property (P) in order to reach any point of \mathfrak{g} , with a total path of length comparable to the distance to the origin for a \ast' -left invariant metric.

LEMMA 3.21 (combining pieces). *Given elements $w_1, \dots, w_p \in \mathfrak{g}$ such that (w_i) projects to a basis of $\mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]'$, there exists $R_0, n_0 \geq 1$ and a sequence of subscripts $i_1, \dots, i_{n_0} \in \{1, \dots, p\}$ such that every $x \in \mathfrak{g}$ can be written*

$$x = r_1 w_{i_1} \ast' \dots \ast' r_{n_0} w_{i_{n_0}}$$

for some parameters r_i satisfying $\max |r_i| \leq R_0 |x|_{\mathfrak{g}}' + R_0$.

PROOF. Straightforward induction on the step of the Lie algebra $(\mathfrak{g}, [\cdot, \cdot]')$.

□

END OF PROOF OF PROPOSITION 3.16. Combine Lemma 3.20 and Lemma 3.21.

□

3.3.2. *Example 2: the case of the free nilpotent Lie algebra.* We describe situations for which the limiting distribution σ_1 does not have full support in \mathfrak{g} . We start with an example of Lie algebra \mathfrak{g} for which *any biased limiting distribution* (i.e. $\overline{X} \neq 0 \pmod{[\mathfrak{g}, \mathfrak{g}]}$) is properly supported. Note that the restriction to biased processes is necessary because for $\overline{X} = 0 \pmod{[\mathfrak{g}, \mathfrak{g}]}$, full support is guaranteed by [VSCC92]. Combined with the central limit theorem established in Section 5, the following statement implies Theorem 1.2 from the Introduction.

PROPOSITION 3.22 (Never full support). *Assume that \mathfrak{g} is a free nilpotent Lie algebra of step at least 3, and $\overline{X} \in \mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]$ is non zero. Then the support of σ_1 does not contain 0 in its interior.*

Remark. For Lie algebras of step at most 2, full support is assured by Corollary 3.17.

The strategy to prove Proposition 3.22 is to find a concrete example of step-3 nilpotent Lie algebra with a limiting distribution that is not fully supported, then use the assumption that \mathfrak{g} is free to project \mathfrak{g} to this example in an equivariant way. To carry out this plan, we introduce $(\mathfrak{l}, [\cdot, \cdot])$ the so-called *filiform nilpotent Lie algebra* of step 3, defined by $\mathfrak{l} = \bigoplus_{i=1}^4 \mathbb{R}e_i$ where

$$\begin{aligned} [e_1, e_2] &= e_3, [e_1, e_3] = e_4, [e_1, e_4] = 0 \\ [e_2, e_3] &= [e_2, e_4] = [e_3, e_4] = 0 \end{aligned}$$

We shall denote $A = e_1, T = e_2$ (as one can realise \mathfrak{l} in the affine group of \mathbb{R}^3 , and A corresponds to a linear transformation, while T is a translation). We let $\bar{X}_\mathfrak{l} = T \pmod{[\mathfrak{l}, \mathfrak{l}]}$ and choose the adapted weight decomposition

$$\tilde{\mathfrak{l}} = \underbrace{\mathbb{R}A \oplus \mathbb{R}T}_{\mathfrak{m}_\mathfrak{l}^{(1)}} \oplus \underbrace{\mathbb{R}\chi_\mathfrak{l}}_{\tilde{\mathfrak{m}}_\mathfrak{l}^{(2)}} \oplus \underbrace{\mathbb{R}[A, T]}_{\mathfrak{m}_\mathfrak{l}^{(3)}} \oplus \underbrace{\mathbb{R}[A, [A, T]]}_{\mathfrak{m}_\mathfrak{l}^{(4)}}$$

where $[\chi_\mathfrak{l}, \cdot] = [T, \cdot]$ on \mathfrak{l} . Observe that the associated graded bracket $[\cdot, \cdot]'$ coincides with $[\cdot, \cdot]$ except that T becomes central.

LEMMA 3.23 (Explicit support). *Let $\sigma_\mathfrak{l}$ be the diffusion law on $(\mathfrak{l}, *')$ associated to an operator $\mathcal{L}_\mathfrak{l}$ as in (23). Then the support of the time-1 distribution is*

$$\text{supp } \sigma_{\mathfrak{l}1} = \{s_1 A + s_2 T + s_3 e_3 + s_4 e_4 : 2s_4 \geq (s_1 + s_3)s_3\}$$

Remark. Of course, we may replace the condition $\bar{X}_\mathfrak{l} = T$ by $\bar{X}_\mathfrak{l} \in \mathbb{R}^* T$, and multiply T appropriately in the support formula. However, if $\bar{X}_\mathfrak{l} \notin \mathbb{R}^* T \pmod{[\mathfrak{l}, \mathfrak{l}]}$, then the support is always full. We prove below that the limiting distribution is Gaussian (in the classical Euclidean sense) if and only if $\bar{X}_\mathfrak{l} \notin \mathbb{R} T \pmod{[\mathfrak{l}, \mathfrak{l}]}$, see Theorem 3.33.

PROOF. By Lemma 3.14 and the facts that T is central in $(\tilde{\mathfrak{l}}, *')$ and $\mathfrak{m}_\mathfrak{l}^{(2)} = 0$, the support of $\sigma_{\mathfrak{l}1}$ is the closure of the subset of \mathfrak{l} given by

$$\mathbb{R}T *' \left\{ \prod_{1 \leq i \leq N}^{*'} (\alpha_i A *' t_i \chi_\mathfrak{l}), t_i \geq 0, \sum t_i = 1, \alpha_i \in \mathbb{R} \right\} *' (-\chi_\mathfrak{l})$$

Setting $\bar{\alpha}_i = \sum_{j=1}^i \alpha_j$, one has

$$\prod_{1 \leq i \leq N}^{*'} (\alpha_i A *' t_i \chi_\mathfrak{l}) = \left(\prod_{1 \leq i \leq N}^{*'} \text{Ad}(\bar{\alpha}_i A)(t_i \chi_\mathfrak{l}) \right) *' (\bar{\alpha}_N A)$$

The terms in the product commute because $\tilde{\mathfrak{l}}^{(2)}$ is abelian. Hence,

$$\begin{aligned} \prod_{1 \leq i \leq N}^{*'} (\alpha_i A *' t_i \chi_\mathfrak{l}) &= \left(\sum_{i=1}^N \text{Ad}(\bar{\alpha}_i A)(t_i \chi_\mathfrak{l}) \right) *' (\bar{\alpha}_N A) \\ &= \left(\sum_{i=1}^N t_i \chi_\mathfrak{l} + \sum_{i=1}^N \bar{\alpha}_i t_i [A, \chi_\mathfrak{l}]' + \frac{1}{2} \sum_{i=1}^N \bar{\alpha}_i^2 t_i [A, [A, \chi_\mathfrak{l}]]' \right) *' (\bar{\alpha}_N A) \\ &= \left(\frac{1}{2} \sum_{i=1}^N \bar{\alpha}_i^2 t_i \right) [A, [A, \chi_\mathfrak{l}]]' *' \left(\sum_{i=1}^N \bar{\alpha}_i t_i \right) [A, \chi_\mathfrak{l}]' *' \left(\sum_{i=1}^N t_i \right) \chi_\mathfrak{l} *' (\bar{\alpha}_N A) \end{aligned}$$

The above computation justifies that the support of $\sigma_{\mathfrak{l}1}$ is exactly

$$\{r_4 e_4 *' r_3 e_3 *' \chi_\mathfrak{l} *' r_0 A *' (-\chi_\mathfrak{l}) : 2r_4 \geq r_3^2\} \oplus \mathbb{R}T$$

We can formulate this subset of \mathfrak{l} with linear combinations of the (e_i) 's:

$$\begin{aligned} r_4 e_4 *' r_3 e_3 *' \chi_\mathfrak{l} *' r_0 A *' (-\chi_\mathfrak{l}) &= r_4 e_4 *' r_3 e_3 *' (r_0 A - r_0 e_3) \\ &= \underbrace{\left(r_4 - \frac{r_0 r_3}{2} \right)}_{s_4} e_4 + \underbrace{(r_3 - r_0)}_{s_3} e_3 + \underbrace{r_0}_{s_1} A \end{aligned}$$

The condition $2r_4 \geq r_3^2$ amounts to $2s_4 \geq (s_1 + s_3)s_3$ so the support of $\sigma_{\mathfrak{l}_1}$ is

$$\{s_1A + s_2T + s_3e_3 + s_4e_4 : 2s_4 \geq (s_1 + s_3)s_3\}$$

□

Remark. Setting $t_0 = 0$, $2t_4 = t_3^2$, we see that $\text{supp } \sigma_{\mathfrak{l}_1}$ is not stable by the dilation D_r for $r > 1$. However, it is stable by D_r for $r \leq 1$. This second observation is true in general for any diffusion as in Section 3.2 with $B = 0$. It comes from the equality

$$\sigma_1 = D_{\sqrt{r}}\sigma_1 *' \text{Ad}(r\mathcal{Y})D_{\sqrt{1-r}}\sigma_1$$

justified by Proposition 3.7, and the fact that the support of σ_1 contains 0 (in its boundary at least) by Lemma 3.13.

Remark. In [Ben21], Benoist shows that for random walks on the discrete step-3 filiform nilpotent group, there may exist positive harmonic functions that are neither characters nor translates of harmonic functions induced from characters. This is in contrast with the Heisenberg case, and the argument also relies on the analysis of the central component.

We may now conclude the proof of Proposition 3.22.

PROOF OF PROPOSITION 3.22. As \mathfrak{g} is free of step at least 3 and \overline{X} is non zero, there must exist a surjective morphism of Lie algebras $\pi : (\mathfrak{g}, [\cdot, \cdot]) \rightarrow (\mathfrak{l}, [\cdot, \cdot])$ such that $\pi(\overline{X}) = T \bmod [\mathfrak{l}, \mathfrak{l}]$. Observe that π sends the weight filtration of $(\mathfrak{g}, \overline{X})$ to that of (\mathfrak{l}, T) , meaning $\pi(\mathfrak{g}^{(b)}) = \mathfrak{l}^{(b)}$ for every b . However, if $\mathfrak{l} = \bigoplus \mathfrak{m}_{\mathfrak{l}}^{(b)}$ is the weight decomposition from Lemma 3.23, it is not necessarily true that $\pi(\mathfrak{m}^{(b)}) = \mathfrak{m}_{\mathfrak{l}}^{(b)}$ or even that $(\pi(\mathfrak{m}^{(b)}))$ is a direct sum decomposition of \mathfrak{l} .

To reduce to this situation, we introduce some other weight decomposition $\mathfrak{g} = \bigoplus \mathring{\mathfrak{m}}^{(b)}$ such that $\pi(\mathring{\mathfrak{m}}^{(b)}) = \mathfrak{m}_{\mathfrak{l}}^{(b)}$ and set $\phi : \tilde{\mathfrak{g}} \rightarrow \tilde{\mathfrak{g}}$ the unique linear map such that $\phi : \tilde{\mathfrak{m}}^{(b)} \rightarrow \tilde{\mathfrak{m}}^{(b)}$ is the isomorphism that factorizes via the identity on the vector space $\mathfrak{g}^{(b)}/\mathfrak{g}^{(b+1)}$. Call $\overset{*}{\sigma}$ the induced graded product on \mathfrak{g} , set $\overset{\circ}{E}_i = \phi(E_i)$, $\overset{\circ}{B} = \phi(B)$, $\overset{\circ}{\mathcal{Y}} = \phi(\mathcal{Y})$, and $\overset{\circ}{\sigma}$ the induced diffusion law on $(\mathfrak{g}, \overset{*}{\sigma})$ as in (23). We know from Lemma 3.6 that

$$\sigma = \phi^{-1} \circ \overset{\circ}{\sigma}$$

In particular, $\text{supp } \sigma_1 = \phi^{-1}(\text{supp } \overset{\circ}{\sigma}_1)$, so it is sufficient to check the claim for $\overset{\circ}{\sigma}$.

Observe now that the condition $\pi(\mathring{\mathfrak{m}}^{(b)}) = \mathfrak{m}_{\mathfrak{l}}^{(b)}$ guarantees that p extends to a morphism from $(\tilde{\mathfrak{g}}, \overset{*}{\sigma})$ to $(\tilde{\mathfrak{l}}, *)$ by setting $p(\overset{\circ}{\chi}) = \chi_{\mathfrak{l}}$. As $\mathfrak{m}_{\mathfrak{l}}^{(2)} = \{0\}$, we have $p(\overset{\circ}{\mathcal{Y}}) = \chi_{\mathfrak{l}}$, and it follows that $p \circ \overset{\circ}{\sigma}$ is the diffusion law generated by

$$\frac{1}{2} \sum_{i=1}^q \left(\text{Ad}(t\chi_{\mathfrak{l}})p(\overset{\circ}{E}_i) \right)^2 + \text{Ad}(t\chi_{\mathfrak{l}})p(\overset{\circ}{B})$$

By Lemma 3.23, the support of $p \circ \overset{\circ}{\sigma}_1$ does not contain 0 in its interior, whence that of $\overset{\circ}{\sigma}_1$ does not either, and this concludes the proof. □

3.4. *Discussion on a lower bound.* We keep the notations of Section 3.2 and address the question of a lower bound for the distribution σ_1 , that is the distribution at time 1 of the diffusion law σ associated to \mathcal{L} from eq. (23). The fact that σ_1 may not have full support on \mathfrak{g} yields some difficulties. They lead us to formulate a lower bound in terms of K -regular chain, where K is a compact in the interior of the support of σ_1 , see Theorem 3.28. When σ_1 has full support, we interpret Theorem 3.28 in terms of the distance determined by horizontal paths, or even a left \ast' -invariant metric on \mathfrak{g} if \mathfrak{g} satisfies the condition (DC) from Section 3.3.1. Finally, we use Theorem 3.28 to characterize the situations where σ_1 is a Gaussian distribution, in the Euclidean sense.

To obtain a general lower bound, we follow the classical strategy of iterating a suitable Harnack inequality. Establishing such an inequality requires more information on the set of arrival points of (continuous piecewise C^1) horizontal paths. We start with a lemma telling us that this set has non empty interior, even if we only consider paths of bounded speed.

LEMMA 3.24. *Let $0 < t_0 < t_1$. The set of arrival points $\gamma(t_1)$, for \mathcal{L} -horizontal paths $\gamma : [t_0, t_1] \rightarrow \mathfrak{g}$ starting at $\gamma(t_0) = 0$ and whose derivative satisfies $\|\partial\gamma\|_\infty \leq 1$, contains an open set of \mathfrak{g} .*

PROOF. Recall from the remark following Lemma 3.11 that a path γ is \mathcal{L} -horizontal if and only if $t \mapsto \gamma(t) \ast' t\mathcal{Y} \ast' (-t(B + \mathcal{Y}))$ is \mathcal{L}_+ -horizontal, where $\mathcal{L}_+ := \frac{1}{2} \sum_{i=1}^q \text{Ad}(t(B + \mathcal{Y}))E_i^2$. We may thus assume that $B = 0$. Denote by \tilde{X} the vector field on $\mathbb{R} \times \mathfrak{g}$ given by $(\frac{d}{dt}, \text{Ad}(t\mathcal{Y})X)$. We saw in Lemma 3.10 that the family $\{\tilde{X}, X \in \mathfrak{m}^{(1)}\}$ is bracket-generating on $\mathbb{R} \times \mathfrak{g}$. Moreover, the associated flows project into horizontal paths. The result now follows from Chow's theorem [Bel96]. \square

Denote by \mathcal{S} the support of σ_1 , and $\mathring{\mathcal{S}}$ the interior of \mathcal{S} . We now use the previous lemma to show that any point in $\mathring{\mathcal{S}}$ can be reached by a horizontal path defined on the time interval $[0, 1]$ and starting at 0. Moreover, the length of these paths is uniformly bounded if the set of prescribed arrival points is compact in $\mathring{\mathcal{S}}$.

LEMMA 3.25 (Accessibility). *Let $K \subseteq \mathring{\mathcal{S}}$ be a compact subset. For $R > 0$ large enough, for any $x \in K$, there exists $\gamma : [0, 1] \rightarrow \mathfrak{g}$ \mathcal{L} -horizontal, such that $\|\partial\gamma\|_\infty \leq R$ and $\gamma(0) = 0$, $\gamma(1) = x$.*

The idea is to use the characterization of $\mathring{\mathcal{S}}$ as the closure of arrival points of horizontal paths on $[0, 1]$ starting at 0 (Fact 3.2) to generate a path from 0 to some y very close to x on the time interval $[0, 1 - r]$ with r small. Then we complete the path from y to x , or any x' in a neighborhood of x , on the remaining time $[1 - r, 1]$ using Lemma 3.24.

PROOF. For $r \in [0, 1]$, set $\mathcal{P}_{1-r} = \{\gamma(1-r) \mid \gamma : [0, 1-r] \rightarrow \mathfrak{g} \text{ } \mathcal{L}\text{-horizontal, } \gamma(0) = 0\}$. Recall from Fact 3.2 and Proposition 3.7 that \mathcal{P}_{1-r} is dense in $D_{\sqrt{1-r}}\mathring{\mathcal{S}}$. Fix $U \subseteq \mathfrak{g}$ a compact symmetric neighborhood of 0 such that $K \ast' U \subseteq \mathring{\mathcal{S}}$, then there exists $r_0 < 1$ such that $D_{\sqrt{1-r}}\mathring{\mathcal{S}} \supseteq K \ast' U$ for $0 \leq r \leq r_0$. It follows that for any open set $V \subseteq U$, one has $\mathcal{P}_{1-r} \ast' V \supseteq K$, and by compactness, K is covered by a finite number of translates $\gamma_i(1-r) \ast' V$ where $\gamma_i(1-r) \in \mathcal{P}_r$.

In order to conclude, we only need to show that for $r < 1$ large enough, any point in some open set $V \subseteq U$ can be reached by a \mathcal{L} -horizontal path on the interval $[1-r, 1]$ starting from 0 and of bounded speed. This is achieved by Lemma 3.24. \square

We now combine the result of Kogoj-Polidoro from Fact 3.3 and the previous analysis to deduce the following Harnack Inequality.

LEMMA 3.26 (Harnack Inequality). *Let $K \subseteq \mathcal{S}$ be a compact subset. For all $\varepsilon_0 > 0$ small enough, there exists $C > 0$ such that for every non-negative function $v : (\frac{\varepsilon_0}{2}, 2) \times \mathfrak{g} \rightarrow \mathbb{R}^+$ solution of $\partial_t - \frac{1}{2} \sum_{i=1}^q (\text{Ad}(t\mathcal{Y})E_i)^2 + \text{Ad}(t\mathcal{Y})B = 0$, we have*

$$v(\varepsilon_0, 0) \leq C \inf_{x \in K} v(1, x)$$

PROOF. Recall from Lemma 3.10 that the vector fields $\partial_t, (\text{Ad}(t\mathcal{Y})E_i)^2, \text{Ad}(t\mathcal{Y})B$ satisfy the Hörmander condition from Fact 3.3. Let $\varepsilon_0 > 0$, let $0 \in U \subseteq \mathfrak{g}$ an open ball, denote by \mathcal{P} the set of couples $(\tau, x) \in (\varepsilon_0, 2) \times U$ such that x is accessible by a U -valued \mathcal{L} -horizontal path on $[\varepsilon_0, \tau]$ with starting point 0. By Lemma 3.25 and using dilations, one may choose $\varepsilon_0 > 0$ small and U large so that $\{1\} \times K$ is included in the interior of the closure of \mathcal{P} . Now the result follows by application of the Harnack Inequality mentioned in Fact 3.4. \square

We combine the previous Harnack Inequality with invariance properties of \mathcal{L} to obtain the following.

LEMMA 3.27 (Harnack Inequality on a space-time cone). *Let $K \subseteq \mathcal{S}$ be a compact subset, let $\varepsilon > 0$. There exists $R > 1$ such that for every non-negative solution $v : \mathbb{R}_{>0} \times \mathfrak{g} \rightarrow \mathbb{R}^+$ of $\partial_t - \frac{1}{2} \sum_{i=1}^q (\text{Ad}(t\mathcal{Y})E_i)^2 + \text{Ad}(t\mathcal{Y})B = 0$, for all $t \in [\varepsilon, +\infty)$, $r \in [0, 1]$, $x \in \mathfrak{g}$, we have*

$$v(t, x) \leq R \inf_{y \in \text{Ad}(t\mathcal{Y})D_{\sqrt{r}}K} v(t+r, x *' y)$$

PROOF. Observe first that if v defined $I \times \mathfrak{g}$ is a solution $\partial_t - \frac{1}{2} \sum_{i=1}^q (\text{Ad}(t\mathcal{Y})E_i)^2 + \text{Ad}(t\mathcal{Y})B = 0$, then for any $x_0 \in \mathfrak{g}$, $\tau_0 \in \mathbb{R}$, $r_0 > 0$, it is also the case of the functions

- $v(t, x_0 *' x)$ defined on $(t, x) \in I \times \mathfrak{g}$
- $v(\tau_0 + t, \text{Ad}(\tau_0\mathcal{Y})x)$ defined on $(t, x) \in (I - \tau) \times \mathfrak{g}$
- $v(r_0 t, D_{\sqrt{r_0}}x)$ defined on $(t, x) \in r_0^{-1}I \times \mathfrak{g}$

Combining these invariance properties, we get the solution

$$v(\tau_0 + r_0 t, x_0 *' \text{Ad}(\tau_0\mathcal{Y})D_{\sqrt{r_0}}x) \text{ defined on } (t, x) \in (r_0^{-1}I - \tau) \times \mathfrak{g}$$

Let $K' \subseteq \mathcal{S}$ be a compact neighborhood of K . Choose $\varepsilon_0 \in (0, \frac{\varepsilon}{1+\varepsilon})$ as in Lemma 3.26 for the compact K' , and denote by $C > 1$ the associated constant. It follows that for any non-negative solution $v : \mathbb{R}_{>0} \times \mathfrak{g} \rightarrow \mathbb{R}^+$ of $\partial_t - \frac{1}{2} \sum_{i=1}^q (\text{Ad}(t\mathcal{Y})E_i)^2 + \text{Ad}(t\mathcal{Y})B = 0$, any $x_0 \in \mathfrak{g}$, $\tau_0, r_0 > 0$, one has

$$(33) \quad v(\tau_0 + r_0 \varepsilon_0, x_0) \leq C \inf_{y \in \text{Ad}(\tau_0\mathcal{Y})D_{\sqrt{r_0}}K'} v(\tau_0 + r_0, x_0 *' y)$$

Let $t \in [\varepsilon, +\infty)$, $r \in [0, 1]$. We now choose (τ_0, r_0) so that $\tau_0 + r_0 \varepsilon_0 = t$ and $\tau_0 + r_0 = t + r$, i.e. we set $\tau_0 = t - r \frac{\varepsilon_0}{1-\varepsilon_0}$, $r_0 = \frac{r}{1-\varepsilon_0}$. The domain restrictions on (t, r) guarantee that $\tau_0 > 0$. Plugging into (33) yields

$$v(t, x_0) \leq C \inf_{y \in \text{Ad}(\tau_0\mathcal{Y})D_{\sqrt{r_0}}K'} v(t+r, x_0 *' y)$$

The result now follows by choosing ε_0 small enough from the start so that $\text{Ad}(\tau_0\mathcal{Y})D_{\sqrt{r_0}}K'$ contains $\text{Ad}(t\mathcal{Y})D_{\sqrt{r}}K'$. Indeed, this condition can be satisfied independently of (t, r) because it amounts to the inclusion $K \subseteq D_{(1-\varepsilon_0)^{-1/2}} \text{Ad}(-\varepsilon_0\mathcal{Y})K'$. \square

We now deduce a general lower bound for the density σ_1 . Given a compact subset $K \subseteq \mathring{\mathcal{S}}$, say that a finite sequence $(t_i, x_i)_{i=0, \dots, n}$ in $\mathbb{R}_{>0} \times \mathfrak{g}$ is a K -regular chain if for all $i \leq n-1$, one has $t_i < t_{i+1}$ and

$$x_i^{-1} *' x_{i+1} \in \text{Ad}(t_i \mathcal{Y}) D_{\sqrt{t_{i+1} - t_i}} K$$

Given $a < b \in \mathbb{R}_{>0}$ and $x, y \in \mathfrak{g}$, set $E_K(a, x, b, y)$ the smallest integer $n \geq 1$ such that there exists a K -regular chain $(t_i, x_i)_{i=0, \dots, n}$ from (a, x) to (b, y) .

An iterated application of Lemma 3.27 yields the following lower bound on σ_1 . Recall for $t > 0$ the notation $\sigma_t = u(t, x) dx$.

THEOREM 3.28 (General lower bound). *Let $K \subseteq \mathring{\mathcal{S}}$ be a compact subset, let $1 > \varepsilon > 0$. There exists $R > 1$ such that for all $x, y \in \mathfrak{g}$,*

$$u(\varepsilon, x) \leq R^{E_K(\varepsilon, x, 1, y)} u(1, y)$$

Theorem 3.28 has several concrete consequences, the first of which is the following:

COROLLARY 3.29. *The density σ_1 is positive on the interior of its support:*

$$\forall x \in \mathring{\mathcal{S}}, \quad u(1, x) > 0$$

PROOF. Choose the compact set $K \subseteq \mathring{\mathcal{S}}$ such that $x \in \mathring{K}$. There is a fixed small neighborhood U of x that is included in $\text{Ad}(\varepsilon \mathcal{Y}) D_{\sqrt{1-\varepsilon}} \mathring{K}$ for any $\varepsilon > 0$ small enough. By Proposition 3.7 taken with $r = \varepsilon$, we see that for small ε the support of $u(\varepsilon, \cdot)$ will intersect any given neighbourhood of 0. In particular, one may choose such ε so that for some $x_0 \in \{u(\varepsilon, \cdot) > 0\}$, one has $x_0^{-1} *' x \in U$. Now $(\varepsilon, x_0), (1, x)$ is a K -regular chain, and the result follows from Theorem 3.28. \square

In case σ_1 has full support, one may compare E_K with the energy functional determined by \mathcal{L} -horizontal paths. More precisely, given $a < b \in \mathbb{R}_{>0}$ and $x, y \in \mathfrak{g}$, define the *horizontal cost* $E_{hor}(a, x, b, y)$ to be the infimum of the energy $E_{hor}(\gamma) = \int_{[a, b]} \|v_\gamma\|^2$ for $\gamma : [a, b] \rightarrow \mathfrak{g}$ continuous, piecewise C^1 , \mathcal{L} -horizontal, such that $\gamma(a) = x, \gamma(b) = y$.

COROLLARY 3.30 (Lower bound when full support). *Assume $\text{supp } \sigma_1 = \mathfrak{g}$. For all $\varepsilon > 0$, there exists $\alpha, A > 0$ such that for all $x \in \mathfrak{g}$,*

$$u(1, x) \geq \alpha \exp(-A E_{hor}(\varepsilon, 0, 1, x))$$

Corollary 3.30 follows from the combination of Theorem 3.28, Corollary 3.29 and the next lemma.

LEMMA 3.31. *Let K be a large enough compact subset of \mathfrak{g} . For $a < b \in \mathbb{R}, x, y \in \mathfrak{g}$,*

$$E_K(a, x, b, y) \leq E_{hor}(a, x, b, y) + 1$$

PROOF. It is sufficient to check that for some K compact, every horizontal path γ on an interval $[a, b]$ whose energy is bounded by 1 is K -regular, in the sense that

$$\gamma(b) \in \gamma(a) *' \text{Ad}(a \mathcal{Y}) D_{\sqrt{b-a}} K$$

Case $B = 0$. We may replace γ by $\text{Ad}(-a \chi)(\gamma(a)^{-1} *' \gamma)$ which is horizontal on $[0, b-a]$ and has same energy as γ . This allows to assume $a = 0$. Then we need check that $D_{\frac{1}{\sqrt{b}}} \gamma(b)$

is bounded uniformly in γ . We set $c(t) = D_{\frac{1}{\sqrt{b}}} \gamma(bt)$. It is a continuous, piecewise- C^1 path defined on $[0, 1]$. Moreover, using that $D_{\frac{1}{\sqrt{b}}}$ is an automorphism for $*'$, that $B = 0$ and $v_\gamma \in \mathfrak{m}^{(1)}$, we get

$$\begin{aligned} \partial c(t) &= D_{\frac{1}{\sqrt{b}}} b \lim_{\varepsilon \rightarrow 0} \frac{\gamma(bt)^{-1} *' \gamma(bt + b\varepsilon)}{b\varepsilon} \\ &= D_{\frac{1}{\sqrt{b}}} b \text{Ad}(bt\mathcal{Y})v_\gamma(bt) \\ &= \text{Ad}(t\mathcal{Y})\sqrt{b}v_\gamma(bt) \end{aligned}$$

Hence, the path c is horizontal on $[0, 1]$ with driving vector $v_c(t) = \sqrt{b}v_\gamma(bt)$. A change of variable then yields that c and γ have same energy, in particular $E_{hor}(c) \leq 1$. It follows that $c(1)$ is bounded, independently of γ , thus concluding the proof in the case $B = 0$.

General case. Define the path $\gamma_0(t) = \gamma(t) *' t\mathcal{Y} *' (-tB - t\mathcal{Y})$ and observe γ_0 is \mathcal{L}_0 -horizontal for the (driftless) generator $\mathcal{L}_0 = \frac{1}{2} \sum_{i=1}^q (\text{Ad}(t(\mathcal{Y} + B))E_i)^2$, with same driving vector as γ . By the first case examined above, we know that

$$\gamma_0(b) \in \gamma_0(a) *' \text{Ad}(a(\mathcal{Y} + B))D_{\sqrt{b-a}}K_0$$

for some compact K_0 independent of γ . Direct computation shows that this implies

$$\gamma(b) \in \gamma(a) *' \text{Ad}(a\mathcal{Y})D_{\sqrt{b-a}}K$$

where $K = K_0 *' (-Y + B) *' Y$, thus concluding the proof. \square

Applying Corollary 3.30 to Lie algebras satisfying the double cancelling assumption **(DC)**, we obtain

COROLLARY 3.32 (Lower bound when **(DC)).** *Assume \mathfrak{g} satisfies the assumption **(DC)** from Section 3.3. There exist constants $\alpha, A > 0$ such that for every $x \in \mathfrak{g}$,*

$$u(1, x) \geq \alpha \exp\left(-A|x|_{\mathfrak{g}}'^2\right)$$

PROOF. It follows from Proposition 3.16 that for some constant $C > 0$, for every $x \in \mathfrak{g}$, we have

$$E_{hor}(1/2, 0, 1, x) \leq C|x|_{\mathfrak{g}}'^2 + C$$

Now the result follows from Corollary 3.30 and Corollary 3.29. \square

PROOF OF THEOREM 1.3. This is a combination of Corollary 3.17, Corollary 3.32 and our central limit theorem established in Section 5. \square

To conclude, we characterize the case when the limit distribution σ_1 is an ordinary Euclidean Gaussian distribution. It is pleasant that the criterion below is expressed only in terms of the weight filtration of \mathfrak{g} , and ultimately only in terms of the adjoint action of the mean \bar{X} on successive quotients of the descending central series. The forward direction (4. \implies 6. below) is due to Crépel and Raugi [CR78, §4]. We are able to show that the converse holds thanks to our lower bound, Corollary 3.30. The following statement encompasses Theorem 1.4 from the introduction.

THEOREM 3.33 (Characterization of the Gaussian case). *Let σ be the diffusion law associated to \mathcal{L} as in eq. (23). The following are equivalent.*

1. The Lie algebra $(\mathfrak{g}, [\cdot, \cdot]')$ is abelian.
2. For all $a \in \{1, \dots, s\}$

$$\mathfrak{g}^{[a]} = \mathfrak{g}^{(2a-1)}.$$

3. For all $i \in \{1, \dots, 2s-1\}$,

$$\mathfrak{m}^{(i)} \neq \{0\} \iff i \text{ is odd.}$$

4. For all $a \in \{1, \dots, s\}$, we have

$$\mathfrak{g}^{[a]} = L^{[a]}(X^{\otimes(a-1)} \otimes \mathfrak{g}).$$

5. For all $a \in \{1, \dots, s\}$, we have

$$[X, \mathfrak{g}^{[a]} / \mathfrak{g}^{[a+1]}] = \mathfrak{g}^{[a+1]} / \mathfrak{g}^{[a+2]}.$$

6. The distribution σ_1 is Gaussian (in the classical Euclidean sense) on \mathfrak{g} .

PROOF. 1. \implies 2. By definition of $[\cdot, \cdot]'$, the assumption $[\cdot, \cdot]' = 0$ means that for all $1 \leq i, j \leq 2s-1$,

$$[\mathfrak{g}^{(i)}, \mathfrak{g}^{(j)}] \subseteq \mathfrak{g}^{(i+j+1)}$$

The result follows by induction on a . Indeed, the case $a = 1$ is tautological. For the induction step, we observe that

$$\mathfrak{g}^{[a+1]} = [\mathfrak{g}^{[a]}, \mathfrak{g}] = [\mathfrak{g}^{(2a-1)}, \mathfrak{g}^{(1)}] \subseteq \mathfrak{g}^{(2a+1)}$$

and the reverse inclusion is clear from the definition of $\mathfrak{g}^{(2a+1)}$.

2. \implies 3. An element in $\mathfrak{g}^{(2a)}$ must be in $\mathfrak{g}^{[a+1]}$, i.e. in $\mathfrak{g}^{(2a+1)}$ by assumption. Hence $\mathfrak{g}^{(2a)} = \mathfrak{g}^{(2a+1)}$. The claim on the $\mathfrak{m}^{(i)}$'s now follows from their definition.

3. \implies 1. Since $[\mathfrak{m}^{(i)}, \mathfrak{m}^{(j)}]' \subseteq \mathfrak{m}^{(i+j)}$, by assumption the left hand side is zero unless both i and j are odd, in which case the right hand side is zero as $i+j$ is even.

2. \iff 4. Only 2. \implies 4. needs a proof. We argue by downwards induction on a . The result is clear for a $a = s$. For the induction step, it is enough to notice that

$$\mathfrak{g}^{[a]} = \mathfrak{g}^{(2a-1)} \subseteq L^{[a]}(X^{\otimes(a-1)} \otimes \mathfrak{g}) + \mathfrak{g}^{[a+1]}$$

with $\mathfrak{g}^{[a+1]} = L^{[a+1]}(X^{\otimes a} \otimes \mathfrak{g}) \subseteq L^{[a]}(X^{\otimes(a-1)} \otimes \mathfrak{g})$ by the induction hypothesis.

4. \iff 5. Similar as 2. \iff 4.

1. \implies 6. One may replace \mathcal{Y} by χ in the definition of the generator \mathcal{L} which because if $(\mathfrak{g}, [\cdot, \cdot]')$ is abelian, then $\text{Ad}(t\mathcal{Y})|_{\mathfrak{g}} = \text{Ad}(t\chi)|_{\mathfrak{g}}$. One may also assume that $q = q_1$, i.e. the vectors E_i in the second order term for \mathcal{L} define a basis of $\mathfrak{m}^{(1)}$. Now, $\nu = \sigma_1$ is the limit measure in the central limit theorem for μ . In this specific context where 4. holds, the result is due to Crepel-Raugi [CR78, Section 4], and their theorem also states that $\nu = \sigma_1$ is Gaussian with expectation B .

6. \implies 1. If $(\mathfrak{g}, [\cdot, \cdot]')$ is not abelian, then there exists a horizontal path $\gamma : [0, 1] \rightarrow \mathfrak{g}$ such that the ratio $\|\gamma(1)\|/\ell(\gamma)$ is arbitrary large. It follows that the ratio $d_{hor}(1/2, 0, 1, x)/\|x\|$ converges to 0 along a sequence of x going to infinity. The lower bound from Corollary 3.30 then contradicts the assumption that σ_1 is Gaussian (it does not decrease fast enough). \square

PROOF OF THEOREM 1.4. Consequence of Theorem 3.33 and our central limit theorem established in Section 5. \square

EXAMPLE (Full upper-triangular group). If $\mathfrak{g} = \bigoplus_{1 \leq i < j \leq q_1+1} \mathbb{R}E_{i,j}$ is the Lie algebra of the group of upper triangular matrices (see Section 3.3.1), then the distribution σ_1 is Gaussian if and only if $\bar{X} = \sum_{1 \leq i \leq q_1} c_i E_{i,i+1} \pmod{\mathfrak{g}^{[2]}}$ with at most one the c_i 's equal to zero. Indeed, if c_i and c_{i+k-1} both vanish, then the application $[\bar{X}, \cdot]$ from $\mathfrak{g}^{[k-1]}/\mathfrak{g}^{[k]}$ to $\mathfrak{g}^{[k]}/\mathfrak{g}^{[k+1]}$ is not surjective: one checks easily that $E_{i,i+k}$ is not in the image. This prevents σ_1 to be Gaussian by Theorem 3.33. The converse implication is straightforward.

EXAMPLE (3-step filiform case). If $\mathfrak{g} = \mathbb{R}A \oplus \mathbb{R}T \oplus \mathbb{R}[A, T] \oplus \mathbb{R}[A, [A, T]]$ is the step-3 filiform Lie algebra (see Section 3.3.2), then the distribution σ_1 is Gaussian if and only if $\bar{X} \in \mathbb{R}^*A \oplus \mathbb{R}T \pmod{\mathfrak{g}^{[2]}}$.

4. The three reductions. To prove the limit theorems announced earlier, we need to perform a series of reductions. This section is devoted to them. We shall show that the random product can be altered slightly (allowing a small error) in three different ways. First, we shall truncate the random variables in order to deal with the moment assumptions. Then (graded replacement) we shall replace the Lie product by the graded Lie product (for which dilations are automorphisms) and third, we shall show that we can replace the driving measure by any measure with the same moments of order at most 2 (Lindeberg replacement). While performing those reductions, we will need to allow moderate deviations of the random product. This flexibility will be used (in a weak form) to take into account the recentering in the proof of the central limit theorem in Section 5. The full extent of moderate deviations is to be used in our follow-up paper [BB23] where we establish the local central limit theorem.

The notations $(\mathfrak{g}, *, \tilde{\mathfrak{g}}, \bar{X}, (\mathfrak{m}^{(b)})_{b \leq 2s-1}, X, \chi, *')$ as well as $\mathcal{D}_N(\delta_0)$ and $\Pi^{[a,b]}$ are those of Section 2. We assume μ is a probability measure on \mathfrak{g} with finite second moment for the weight filtration induced by \bar{X} , and such that μ_{ab} has expectation \bar{X} . In particular $X = \mathbb{E}(\mu^{(1)})$. We will mostly work with $\tilde{\mu} = \mu - X + \chi$ the associated probability measure on $\tilde{\mathfrak{g}}$.

4.1. *Truncation and moment estimates.* We first establish moment estimates. To fix ideas, we start with the case where μ has a finite moment of every order. Consider a monomial function $M : \tilde{\mathfrak{g}}^t \rightarrow \mathbb{R}$ of degree $b \geq 1$, see Section 2.7. We define its statistics M^∞ on $\tilde{\mathfrak{g}}^{\mathbb{N}^*}$ as the formal series

$$M^\infty(\underline{x}) = \sum_{1 \leq n_1 < \dots < n_t < \infty} M(x_{n_1}, \dots, x_{n_t}).$$

For every $N \geq 1$, we identify $\tilde{\mathfrak{g}}^N$ as the subset of $\tilde{\mathfrak{g}}^{\mathbb{N}^*}$ made of sequences that vanish after the N first terms. In particular, M^∞ is a well-defined function on $\tilde{\mathfrak{g}}^N$.

LEMMA 4.1. *Let $m \in \mathbb{N}$ and assume μ has a finite moment of order $2mb$. Then for $N \geq 1$,*

$$\mathbb{E}_{\tilde{\mu}^{\otimes N}} [M^\infty(\underline{x})^{2m}] \ll_m N^{bm},$$

and in particular for $\delta_0 > 0$, $g, h \in \mathcal{D}_N(\delta_0)$, $\underline{y} = (g, \underline{x}, h)$, $a, b \geq 1$,

$$\mathbb{E}_{\underline{x} \sim \tilde{\mu}^{\otimes N}} [\|\Pi^{[a,b]}(\underline{y})\|^{2m}] \ll_m N^{(b+2\delta_0)m}.$$

PROOF. Given $\alpha \in \mathbb{N}^{\dim \tilde{\mathfrak{g}}}$, denote by $x \mapsto x^\alpha$ the corresponding monomial on $\tilde{\mathfrak{g}}$ (say for some ordering of the basis $e_j^{(i)}$ fixed in Section 2) and write $d(\alpha)$ its degree for the weight

filtration. Expanding the power $2m$, we can write for $N \geq 1$, $\underline{x} \in \tilde{\mathfrak{g}}^N$,

$$M^\infty(\underline{x})^{2m} = \sum_{\ell=1}^{2mb} \sum_{\underline{\alpha} \in \mathcal{A}_\ell^1} \sum_{\underline{n} \in \llbracket 1, N \rrbracket^\ell} C_{\underline{\alpha}} x_{n_1}^{\alpha_1} \dots x_{n_\ell}^{\alpha_\ell}$$

where $\sup_{\underline{\alpha}} |C_{\underline{\alpha}}| = O_m(1)$ and

$$\mathcal{A}_\ell^1 = \left\{ \underline{\alpha} \in (\mathbb{N}^{\dim \tilde{\mathfrak{g}}})^\ell : \sum_{i=1}^{\ell} d(\alpha_i) = 2mb \text{ and } \inf_i d(\alpha_i) \geq 1 \right\}$$

Then by linearity and independence,

$$\mathbb{E}_{\tilde{\mu}^{\otimes N}} [M^\infty(\underline{x})^{2m}] = \sum_{\ell=1}^{2mb} \sum_{\underline{\alpha} \in \mathcal{A}_\ell^1} \sum_{\underline{n} \in \llbracket 1, N \rrbracket^\ell} C_{\underline{\alpha}} \mathbb{E}_{\tilde{\mu}}(x_{n_1}^{\alpha_1}) \dots \mathbb{E}_{\tilde{\mu}}(x_{n_\ell}^{\alpha_\ell})$$

The terms for which some $x_i^{\alpha_i}$ has degree 1 satisfy $\mathbb{E}_{\tilde{\mu}}(x_i^{\alpha_i}) = 0$, hence the $\underline{\alpha}$ that contribute belong to

$$\mathcal{A}_\ell^2 = \left\{ \underline{\alpha} \in (\mathbb{N}^{\dim \tilde{\mathfrak{g}}})^\ell : \sum_{i=1}^{\ell} d(\alpha_i) = 2mb \text{ and } \inf_i d(\alpha_i) \geq 2 \right\}$$

thus forcing ℓ to belong to $\{1, \dots, mb\}$.

It follows that

$$\mathbb{E}_{\tilde{\mu}^{\otimes N}} [M^\infty(\underline{x})^{2m}] = \sum_{\ell=1}^{mb} \sum_{\underline{\alpha} \in \mathcal{A}_\ell^2} \sum_{\underline{n} \in \llbracket 1, N \rrbracket^\ell} C_{\underline{\alpha}} \mathbb{E}_{\tilde{\mu}}(x_{n_1}^{\alpha_1}) \dots \mathbb{E}_{\tilde{\mu}}(x_{n_\ell}^{\alpha_\ell})$$

and the bound for $\mathbb{E}_{\tilde{\mu}^{\otimes N}} [M^\infty(\underline{x})^{2m}]$ follows because the parameters $\ell, \underline{\alpha}$ vary in a bounded set independent of N , the parameter \underline{n} varies in a set of cardinality $N^\ell \leq N^{mb}$, and the product of expectations is $O_m(1)$ by the moment assumption on μ .

To prove the second assertion, we note that $\Pi^{[a,b]}(g, \underline{x}, h)$ is a finite formal linear combination (depending on a, b but not N) of statistics in \underline{x} associated to monomials of degree at most b , and such that each statistics of degree $b' \leq b$ has its coefficients bounded above by

$$O(1) \sum_{i_1 + \dots + j_\ell = b - b'} \|g^{(i_1)}\| \dots \|g^{(i_k)}\| \|h^{(j_1)}\| \dots \|h^{(j_\ell)}\|$$

if $b' < b$, and $O(1)$ otherwise. Given the condition on $g, h \in \mathcal{D}_N(\delta_0)$, this bound is in particular $O(N^{\frac{b-b'}{2} + \delta_0})$. The result follows by Hölder Inequality and the previous bound established for $\mathbb{E}_{\tilde{\mu}^{\otimes N}} [M^\infty(\underline{x})^{2m}]$. \square

In order to obtain similar moment estimates but with weaker moment assumptions on μ , we introduce the *truncation* $T_N \tilde{\mu}$ as the image of $\tilde{\mu}$ by $T_N : \tilde{\mathfrak{g}} \rightarrow \tilde{\mathfrak{g}}, x = \oplus x^{(b)} \mapsto \oplus T_N x^{(b)}$ where

$$T_N x^{(b)} = \begin{cases} x^{(b)} & \text{if } \|x^{(b)}\| \leq N^{b/2} \\ 0 & \text{if } b \geq 2 \text{ and } \|x^{(b)}\| > N^{b/2} \\ c_N & \text{if } b = 1 \text{ and } \|x^{(1)}\| > N^{1/2} \end{cases}$$

and $c_N = -\tilde{\mu}(\|x^{(1)}\| > N^{1/2})^{-1} \mathbb{E}_{\tilde{\mu}}(x^{(1)} \mathbb{1}_{\|x^{(1)}\| \leq N^{1/2}})$ is chosen so that $T_N \tilde{\mu}$ is centered in the $\mathfrak{m}^{(1)}$ -coordinate.

It is good to notice first that one may replace $\tilde{\mu}$ by $T_N \tilde{\mu}$ at a very low cost.

LEMMA 4.2 (Cost of truncation). *Let $q \in [2, +\infty)$. If μ has finite q -th moment then for all $N \geq 1$,*

$$\|\tilde{\mu}^{\otimes N} - (T_N \tilde{\mu})^{\otimes N}\| \leq \epsilon_N N^{-\frac{q}{2}+1}$$

where $\|\cdot\|$ is the norm in total variation, and the sequence $\epsilon_N \in \mathbb{R}_{>0}^{N^*}$ goes to zero while being bounded by $\epsilon_N \leq C(1 + m_q(\mu))$ where $C > 0$ is a constant that may only depend on q .

PROOF. Note that

$$\tilde{\mu}\{T_N x \neq x\} \leq \tilde{\mu}\{\max_b \|x^{(b)}\|^{1/b} > N^{1/2}\} \leq \epsilon_N N^{-\frac{q}{2}}$$

with ϵ_N as in the statement. Hence $T_N^{\otimes N}$ is the identity on a subset of $\tilde{\mathfrak{g}}^N$ of $\tilde{\mu}^{\otimes N}$ -measure at least $(1 - \epsilon_N N^{-\frac{q}{2}+1})$, up to multiplying ϵ_N by a bounded constant. \square

We now show that $(T_N \tilde{\mu})^{\otimes N}$ satisfies the moment estimates from Lemma 4.1.

PROPOSITION 4.3 (Moment estimate). *Assume that μ has a finite second moment. For $N \geq 1$, $(N_i)_{i \leq N} \in [0, N]^N$, $m \geq 1$ an integer,*

$$\mathbb{E}_{\otimes_{i=1}^N T_{N_i} \tilde{\mu}} [M^\infty(\underline{x})^{2m}] \ll_m N^{bm},$$

and in particular for $\delta_0 \geq 0$, $g, h \in \mathcal{D}_N(\delta_0)$, $\underline{y} = (g, \underline{x}, h)$, $a, b \geq 1$,

$$\mathbb{E}_{\underline{x} \sim \otimes_{i=1}^N T_{N_i} \tilde{\mu}} [\|\Pi^{[a,b]}(\underline{y})\|^{2m}] \ll_m N^{(b+2\delta_0)m}.$$

Moreover, the implicit constant in \ll_m can be taken of the form $C(1 + m_2(\mu)^{mb})$ where $C = C(\mathfrak{g}, [\cdot, \cdot], \mathbb{R}\bar{X}, (\mathbf{m}^{(b)})_b, (e_j^{(i)})_{i,j}, m) > 0$ is a constant and $m_2(\mu)$ is the second moment of μ .

It is quite remarkable that μ is only assumed to have finite second moment.

PROOF. Arguing as for Lemma 4.1 (with $\otimes_i T_{N_i} \tilde{\mu}$ in place of $\tilde{\mu}^{\otimes N}$) gives

$$\mathbb{E}_{\otimes_{i=1}^N T_{N_i} \tilde{\mu}} [M^\infty(\underline{x})^{2m}] = \sum_{\ell=1}^{mb} \sum_{\underline{\alpha} \in \mathcal{A}_\ell^2} \sum_{\underline{n} \in [1, N]^\ell} C_{\underline{\alpha}} \mathbb{E}_{T_{N_{n_1}} \tilde{\mu}}(x_{n_1}^{\alpha_1}) \cdots \mathbb{E}_{T_{N_{n_\ell}} \tilde{\mu}}(x_{n_\ell}^{\alpha_\ell})$$

but it is no longer true that the product of expectation is $O_m(1)$. Nonetheless, note that for $d(\alpha) \geq 2$, $N' \geq 1$,

$$\mathbb{E}_{T_{N'} \tilde{\mu}}(|x^\alpha|) = \mathbb{E}_{\tilde{\mu}}\left(|(T_{N'} x)^{\frac{2\alpha}{d(\alpha)}}|^{\frac{d(\alpha)}{2}-1} |(T_{N'} x)^{\frac{2\alpha}{d(\alpha)}}|\right) = O(N'^{\frac{d(\alpha)}{2}-1})$$

because $|T_{N'} x^{\frac{2\alpha}{d(\alpha)}}| = O(N')$ and $\mathbb{E}_{\tilde{\mu}}(|(T_{N'} x)^{\frac{2\alpha}{d(\alpha)}}|)$ is bounded independently of N' (by a constant of the form $C(1 + m_2(\mu))$ with C as in the statement).

This bound implies that

$$\begin{aligned} \mathbb{E}_{\otimes_{i=1}^N T_{N_i} \tilde{\mu}} [M^\infty(\underline{x})^{2m}] &= \sum_{\ell=1}^{mb} \sum_{\underline{\alpha} \in \mathcal{A}_\ell^2} \sum_{\underline{n} \in [1, N]^\ell} O_m(N^{mb-\ell}) \\ &= O_m(N^{mb}) \end{aligned}$$

as the parameters $\ell, \underline{\alpha}$ vary in a bounded set independent of N , while the parameter \underline{n} varies in a set of cardinality N^ℓ .

The estimate for $\mathbb{E}_{\underline{x} \sim \otimes_{i=1}^N T_{N_i} \tilde{\mu}} [\|\Pi^{[a,b]}(\underline{y})\|^{2m}] \ll_m N^{(b+2\delta_0)m}$ follows by the same argument as for Lemma 4.1.

The comment on the constant in \ll_m is just an inspection of the proof. \square

4.2. *Graded replacement.* In the remainder of Section 4 we keep the assumption that μ has a finite second moment for the weight filtration induced by \overline{X} . We now show how to replace the product $*$ by the its graded counterpart $*'$, for which the dilations $(D_r)_{r>0}$ act as automorphisms.

PROPOSITION 4.4 (Graded replacement). *For $N \geq 1$, $(N_i)_{i \leq N} \in [0, N]^N$, $g, h \in \mathcal{D}_N(\delta_0)$, $\underline{y} = (g, \underline{x}, h)$, $m \geq 1$,*

$$\mathbb{E}_{\underline{x} \sim \otimes_{i=1}^N T_{N_i} \tilde{\mu}} \left(\|D_{\frac{1}{\sqrt{N}}} \Pi(\underline{y}) - D_{\frac{1}{\sqrt{N}}} \Pi'(\underline{y})\|^{2m} \right) = O_m(N^{(-1+2\delta_0)m}).$$

Moreover, the implicit constant in \ll_m can be taken of the form $C(1 + m_2(\mu)^{mb})$ where $C = C(\mathfrak{g}, [\cdot, \cdot], \mathbb{R}\overline{X}, (\mathbf{m}^{(b)})_b, (e_j^{(i)})_{i,j}, m) > 0$ and $m_2(\mu)$ is the second moment of μ .

PROOF. Using the triangle inequality, it is sufficient to show that the bound for

$$\mathbb{E}_{\underline{x} \sim \otimes_{i=1}^N T_{N_i} \tilde{\mu}} \left(\|D_{\frac{1}{\sqrt{N}}} \Pi^{(b)}(\underline{y}) - D_{\frac{1}{\sqrt{N}}} \Pi'^{(b)}(\underline{y})\|^{2m} \right)$$

But $\Pi^{(b)}(\underline{y}) - \Pi'^{(b)}(\underline{y})$ is the projection of $\Pi^{(b)}(\underline{y})$ to $\tilde{\mathfrak{g}}^{(b+1)}$ parallel to $\tilde{\mathfrak{m}}^{(b)}$. Hence,

$$\|D_{\frac{1}{\sqrt{N}}} (\Pi^{(b)}(\underline{y}) - \Pi'^{(b)}(\underline{y}))\| \leq N^{-(b+1)/2} \|\Pi^{(b)}(\underline{y})\|$$

and the result follows because we proved in Proposition 4.3 that

$$\mathbb{E}_{\underline{x} \sim \otimes_{i=1}^N T_{N_i} \tilde{\mu}} \left[\|\Pi^{(b)}(\underline{y})\|^{2m} \right] \ll_m N^{(b+2\delta_0)m}$$

where the implicit constant in \ll_m has the desired form. \square

Remark. The same holds for $\tilde{\mu}$ in place of $T_N \tilde{\mu}$ under the assumption of a finite moment of order $2mb_{\max}$, where $b_{\max} \leq 2s - 1$ is the length of the weight filtration. Just use Lemma 4.1 in place of Proposition 4.3.

We infer a graded approximation at the level of Fourier transforms. It will be used to prove the central limit theorem with Berry-Esseen estimate (and the local limit theorem in [BB23]).

COROLLARY 4.5. *For ξ a linear form on $\tilde{\mathfrak{g}}$, $N \geq 1$, $(N_i)_{i \leq N} \in [0, N]^N$, $g, h \in \mathcal{D}_N(\delta_0)$, $\underline{y} = (g, \underline{x}, h)$,*

$$\left| \mathbb{E}_{\underline{x} \sim \otimes_{i=1}^N T_{N_i} \tilde{\mu}} [e_\xi(\Pi(\underline{y})) - e_\xi(\Pi'(\underline{y}))] \right| \leq C(1 + m_2(\mu)^s) \|D_{\sqrt{N}} \xi\| N^{-1/2+\delta_0}$$

where $C = C(\mathfrak{g}, [\cdot, \cdot], \mathbb{R}\overline{X}, (\mathbf{m}^{(b)})_b, (e_j^{(i)})_{i,j}) > 0$.

PROOF.

$$\begin{aligned} \left| \mathbb{E}_{\underline{x} \sim \otimes_{i=1}^N T_{N_i} \tilde{\mu}} [e_\xi(\Pi(\underline{y})) - e_\xi(\Pi'(\underline{y}))] \right| &\ll \mathbb{E}_{\underline{x} \sim \otimes_{i=1}^N T_{N_i} \tilde{\mu}} (|\xi \Pi(\underline{y}) - \xi \Pi'(\underline{y})|) \\ &\leq \|D_{\sqrt{N}} \xi\| \mathbb{E}_{\underline{x} \sim \otimes_{i=1}^N T_{N_i} \tilde{\mu}} (\|D_{\frac{1}{\sqrt{N}}} (\Pi(\underline{y}) - \Pi'(\underline{y}))\|) \\ &\leq C(1 + m_2(\mu)^s) \|D_{\sqrt{N}} \xi\| N^{-1/2+\delta_0} \end{aligned}$$

where the last line is given by the inequality $\|\cdot\|_{L^1} \leq \|\cdot\|_{L^2}$ followed by Proposition 4.4 (case $m = 1$). \square

We also deduce a result of graded approximation that is uniform on the time interval $\{0, \dots, N\}$. It will be useful to prove our theorem à la Donsker in Section 5.3. Here we do not use truncation.

COROLLARY 4.6. *Let $\alpha \in [0, \frac{1}{2})$. For $N \geq 1$, set*

$$F_N = \left\{ \underline{x} \in \tilde{\mathfrak{g}}^N : \sup_{k \leq N} \|D_{\frac{1}{\sqrt{N}}} \Pi \left((x_i)_{i=1}^k, -k\chi \right) - D_{\frac{1}{\sqrt{N}}} \Pi' \left((x_i)_{i=1}^k, -k\chi \right)\| > N^{-\alpha} \right\}$$

If μ has finite moment of order $2mb_{\max}$ where b_{\max} is the length of the weight filtration, then

$$\tilde{\mu}^{\otimes N}(F_N) = O_{\alpha, m}(N^{(-1+2\alpha+\frac{1}{m})m})$$

In particular $\tilde{\mu}^{\otimes N}(F_N)$ goes to 0 provided $m > 1/(1 - 2\alpha)$.

PROOF. The result follows by replacing the supremum by a sum, then using the Markov inequality and (the remark after the proof of) Proposition 4.4 with $\delta_0 = 0$. \square

4.3. *Lindeberg replacement.* We show how to replace the driving measure by any other measure which coincides up to order 2 for the weight filtration. The substitution takes place at the level of Fourier transforms.

PROPOSITION 4.7 (Lindeberg replacement). *Let η be a probability measure on \mathfrak{g} with finite second moment for the weight filtration induced by \overline{X} and which coincides with μ up to order 2. For ξ a linear form on $\tilde{\mathfrak{g}}$, $N \geq 1$, $(N_i)_{i \leq N} \in [0, N]^N$, $g, h \in \mathcal{D}_N(\delta_0)$, $\underline{y} = (g, \underline{x}, h)$,*

$$|\mathbb{E}_{\underline{x} \sim \otimes_{i=1}^N T_{N_i} \tilde{\mu}}[e_{\xi}(\Pi'(\underline{y}))] - \mathbb{E}_{\underline{x} \sim \otimes_{i=1}^N T_{N_i} \tilde{\eta}}[e_{\xi}(\Pi'(\underline{y}))]| \leq (\|D_{\sqrt{N}}\xi\| + \|D_{\sqrt{N}}\xi\|^3) N^{3s\delta_0} \sum_{i=1}^N N_i^{-1} \varepsilon_N$$

where $\varepsilon_N = o_{\eta}(1)$. More generally, assuming μ, η both have finite m -th moment for some $m \in [2, 3]$, we may take $\varepsilon_N = o_{\eta}(N^{-\frac{m}{2}+1})$ if $m < 3$ and $\varepsilon_N = O_{\eta}(N^{-1/2})$ if $m = 3$.

Moreover, in the $m = 3$ case, the implicit constant in O_{η} can be taken of the form $C(1 + m_2(\mu) + m_2(\eta))^{6s^2} (1 + m_3(\mu) + m_3(\eta))$ where $C = C(\mathfrak{g}, [\cdot, \cdot], \mathbb{R}\overline{X}, (\mathbf{m}^{(b)})_b, (e_j^{(i)})_{i,j}) > 0$ and $m_q(\mu), m_q(\eta)$ denote the q -th moments of μ and η .

In this paper, we will apply this proposition with (N_i) constant equal to N . In [BB23], we will need to consider gradually increasing (N_i) . In both cases, $\sum_{i=1}^N N_i^{-1} = O(1)$, so the right-hand side is $O_{\eta}(1)(\|D_{\sqrt{N}}\xi\| + \|D_{\sqrt{N}}\xi\|^3) N^{3s\delta_0} \varepsilon_N$.

PROOF. We replace $\tilde{\mu}$ -variables by $\tilde{\eta}$ -variables successively. At the step i , we set

$$\kappa_i = g \otimes T_{N_1} \tilde{\mu} \otimes \dots \otimes T_{N_{i-1}} \tilde{\mu} \quad \theta_i = T_{N_{i+1}} \tilde{\eta} \otimes \dots \otimes T_{N_N} \tilde{\eta} \otimes h$$

and we need show that

$$\Delta_i := |\mathbb{E}_{\kappa_i \otimes (T_{N_i} \tilde{\mu} - T_{N_i} \tilde{\eta}) \otimes \theta_i}(e_{\xi}(\Pi'(\underline{u}, v, \underline{w})))| \leq (\|D_{\sqrt{N}}\xi\| + \|D_{\sqrt{N}}\xi\|^3) N_i^{-1+3s\delta_0} \varepsilon_N$$

Setting $\xi_N = D_{\sqrt{N}}\xi$, $\underline{u}_N = D_{\frac{1}{\sqrt{N}}}\underline{u}$, $v_N = D_{\frac{1}{\sqrt{N}}}v$, $\underline{w}_N = D_{\frac{1}{\sqrt{N}}}\underline{w}$, we can use that dilations commute with $*$ ' to rewrite

$$\Delta_i = |\mathbb{E}_{\kappa_i \otimes (T_{N_i} \tilde{\mu} - T_{N_i} \tilde{\eta}) \otimes \theta_i}(e_{\xi_N}(\Pi'(\underline{u}_N, v_N, \underline{w}_N)))|$$

We now isolate the part of $\Pi'(\underline{u}_N, v_N, \underline{w}_N)$ depending on v . To do so, we observe that for $g_1, g_2, g_3 \in \tilde{\mathfrak{g}}$,

$$g_1 *' g_2 *' g_3 = g_1 *' g_3 + P_{g_1, g_3}(g_2)$$

where P_{g_1, g_3} is a polynomial of bounded degree without constant term and whose coefficients depend polynomially on g_1, g_3 . Setting $x_N = \Pi' \underline{u}_N$, $z_N = \Pi' \underline{w}_N$, we then rewrite

$$\begin{aligned} \Delta_i &= |\mathbb{E}_{\kappa_i \otimes (T_{N_i} \tilde{\mu} - T_{N_i} \tilde{\eta}) \otimes \theta_i} (e_{\xi_N}(x_N *' z_N + P_{x_N, z_N}(v_N)))| \\ &= |\mathbb{E}_{\kappa_i \otimes \theta_i} [e_{\xi_N}(x_N *' z_N) \mathbb{E}_{(T_{N_i} \tilde{\mu} - T_{N_i} \tilde{\eta})} (e_{\xi_N}(P_{x_N, z_N}(v_N)))]| \\ &\leq \mathbb{E}_{\kappa_i \otimes \theta_i} |\mathbb{E}_{(T_{N_i} \tilde{\mu} - T_{N_i} \tilde{\eta})} (e_{\xi_N}(P_{x_N, z_N}(v_N)))| \end{aligned}$$

Using Taylor expansion up to order 3, we have

$$\begin{aligned} \mathbb{E}_{(T_{N_i} \tilde{\mu})} (e_{\xi_N}(P_{x_N, z_N}(v_N))) &= 1 + \mathbb{E}_{(T_{N_i} \tilde{\mu})} (-2i\pi \xi_N P_{x_N, z_N}(v_N)) + \mathbb{E}_{(T_{N_i} \tilde{\mu})} (-4\pi^2 (\xi_N P_{x_N, z_N}(v_N))^2) \\ &\quad + \mathbb{E}_{(T_{N_i} \tilde{\mu})} (O(\xi_N P_{x_N, z_N}(v_N))^3) \end{aligned}$$

and the analogous formula for $T_{N_i} \tilde{\eta}$. Our purpose is to compare those two polynomial expansions, and we do this one monomial at a time. Recall that for $\alpha \in \mathbb{N}^{\dim \tilde{\mathfrak{g}}}$, we write $x \mapsto x^\alpha$ the induced monomial on $\tilde{\mathfrak{g}}$ (for some ordering of the basis $e_j^{(i)}$ fixed in Section 2) and $d(\alpha)$ its degree. We compare the expectations of $v_N^\alpha := (v_N)^\alpha$ when v varies with respect to the measures $T_{N_i} \tilde{\mu}$ and $T_{N_i} \tilde{\eta}$.

- If $d(\alpha) = 1$, then

$$\mathbb{E}_{T_{N_i} \tilde{\mu}}(v_N^\alpha) = \mathbb{E}_{T_{N_i} \tilde{\eta}}(v_N^\alpha) = 0$$

because our truncations are centered by definition.

- If $d(\alpha) = 2$, then we use that μ and η coincide up to order 2 to write

$$\begin{aligned} |\mathbb{E}_{T_{N_i} \tilde{\mu} - T_{N_i} \tilde{\eta}}(v_N^\alpha)| &= N^{-1} |\mathbb{E}_{T_{N_i} \tilde{\mu} - \tilde{\mu}}(v^\alpha) + \mathbb{E}_{\tilde{\eta} - T_{N_i} \tilde{\eta}}(v^\alpha)| \\ &\leq N^{-1} \mathbb{E}_{\tilde{\mu} + \tilde{\eta}} (|v^\alpha| + |(T_{N_i} v)^\alpha|) 1_{E_{N_i}}(v) \end{aligned}$$

where $E_{N_i} = \{v : T_{N_i} v \neq v\}$ has measure $\tilde{\mu}(E_{N_i}), \tilde{\eta}(E_{N_i}) \leq o_\eta(N_i^{-m/2})$ if μ, η have finite m -th moment. Using the Hölder inequality $\|f_1 f_2\|_{L^1} \leq \|f_1\|_{L^{m/2}} \|f_2\|_{L^{\frac{m}{m-2}}}$, we obtain

$$\mathbb{E}_{T_{N_i} \tilde{\mu} - T_{N_i} \tilde{\eta}}(v_N^\alpha) = o_\eta(N_i^{-m/2}).$$

- If $d(\alpha) \geq 3$ and μ, η both have finite m -th moment, then we may use Lemma 4.8 below to get

$$\mathbb{E}_{T_{N_i} \tilde{\mu} + T_{N_i} \tilde{\eta}}(|v_N^\alpha|) = \begin{cases} o_\eta(N_i^{-m/2}) & \text{if } m \in [2, 3) \\ O_\eta(N_i^{-3/2}) & \text{if } m \in [3, +\infty) \end{cases}$$

Combining the above Taylor expansions and the monomial estimates, we deduce that

$$|\mathbb{E}_{(T_{N_i} \tilde{\mu} - T_{N_i} \tilde{\eta})} (P_{x_N, z_N}(v_N))| \leq (\|\xi_N\| + \|\xi_N\|^3)(1 + \|x_N\| + \|z_N\|)^{3s} N_i^{-1} \varepsilon_N$$

and the domination of Δ_i follows by integrating in $\underline{u}_N, \underline{w}_N$ and using the moment estimate from Proposition 4.3.

The final claim on ε_n in the $m = 3$ case follows from inspecting the proof. More precisely, the dependence in $(1 + m_3(\mu) + m_3(\eta))$ originates from the estimates of $\mathbb{E}_{(T_{N_i} \tilde{\mu} - T_{N_i} \tilde{\eta})}(v_N^\alpha)$ performed above, and the dependence in $(1 + m_2(\mu) + m_2(\eta))^{6s^2}$ appears in the very last step, when we integrate with respect to $\underline{u}_N, \underline{w}_N$ and use Proposition 4.3. \square

To complete the proof of Proposition 4.7, we record the following.

LEMMA 4.8. *Let $\varphi : (\Omega, \mathbb{P}) \rightarrow \mathbb{R}^+$ be an integrable non-negative random variable. For $\beta > 1$,*

$$\int_{\Omega} \varphi^{\beta} \mathbb{1}_{\varphi \leq N} d\mathbb{P} = o_{\varphi}(N^{\beta-1})$$

PROOF. Just observe that

$$\int_{\Omega} \left(\frac{\varphi}{N}\right)^{(\beta-1)} \mathbb{1}_{\varphi \leq N} \varphi d\mathbb{P} \rightarrow 0$$

by dominated convergence. \square

5. Global limit theorems. This section is dedicated to the proof of the global limit theorems announced in the introduction, namely Theorems 1.1, 1.6, 1.7 and Corollary 1.5. We will combine Section 3 and Section 4. The notations $(\mathfrak{g}, *, \tilde{\mathfrak{g}}, \tilde{X}, (\mathfrak{m}^{(b)})_{b \leq 2s-1}, X, \chi, *')$ are those of Section 2. We assume that μ is a probability measure on \mathfrak{g} with finite second moment for the weight filtration induced by \tilde{X} , and such that μ_{ab} has expectation \tilde{X} and non-degenerate covariance matrix. In particular, we have $X = \mathbb{E}(\mu^{(1)})$.

Let $(E_i)_{1 \leq i \leq q_1}$ be a basis of $\mathfrak{m}^{(1)}$ in which the covariance matrix of $\mu^{(1)}$ is the identity. Let $B \in \mathfrak{m}^{(2)}$ be the expectation of $\mu^{(2)}$. Set $\mathcal{L} = (\mathcal{L}_t)_{t \in \mathbb{R}^+}$ the time-dependent left-invariant differential operator on $(\mathfrak{g}, *')$ given by

$$(34) \quad \mathcal{L}_t = \frac{1}{2} \sum_{i=1}^{q_1} (\text{Ad}(t\chi)E_i)^2 + \text{Ad}(t\chi)B$$

We denote by $(W(t))_{t \in \mathbb{R}^+}$ a diffusion process on $(\mathfrak{g}, *)$ of infinitesimal generator \mathcal{L} , and set $Z(t) = W(t) *' t\chi$ as in Lemma 3.11. The notations σ, λ correspond to the respective distributions of W and Z in the space of continuous paths. Finally we write $\nu = \sigma_1$ the time-1 distribution of σ , i.e. the law of $W(1)$.

5.1. *Central limit theorem and convergence rate.* We prove the central limit theorem and the Berry-Esseen estimate announced in the introduction. We start with a lemma that further connects μ to the diffusion σ , via their respective bias extensions.

LEMMA 5.1. *The time-1 distribution λ_1 is the bias extension $\tilde{\eta}$ to $\tilde{\mathfrak{g}}$ of a measure η on \mathfrak{g} with finite p -th moment for all p and which coincides with μ up to order 2. In particular, one has for all $N \geq 1$,*

$$\tilde{\eta}^{*N} = D_{\sqrt{N}}\nu *' N\chi$$

Later, we will bound all the moments of η using the second moment of μ , see Proposition 5.4.

PROOF. The only possible candidate is $\eta = \lambda_1 - \chi + X$. It does have finite polynomial moments by Proposition 3.8 and satisfies $\tilde{\eta} = \lambda_1$. By Lemma 3.11, we have

$$\tilde{\eta}^{*N} = \lambda_1^{*N} = \lambda_N = D_{\sqrt{N}}\nu *' N\chi$$

It remains to compute the covariance and expectations. First recall from Lemma 3.11 that for any $f \in C_c^\infty(\tilde{\mathfrak{g}})$,

$$(35) \quad \lambda_1(f) - f(0) = \int_0^1 \lambda_s(\tilde{\mathcal{L}}.f) ds$$

Rewriting $\lambda_s(\widetilde{\mathcal{L}}.f) = \lambda_1[(\widetilde{\mathcal{L}}.f) \circ D_{\sqrt{s}}]$ thanks to Proposition 3.7, and using that λ_1 has finite polynomial moments, we see that the relation (35) holds more generally for all $f \in C^\infty(\widehat{\mathfrak{g}})$ whose derivatives have polynomial growth (indeed, just approximate $f(x)$ by $f(x)\rho(\varepsilon x)$ where ρ is a bump function and $\varepsilon \rightarrow 0^+$). Now, if f is a linear form on $\mathfrak{m}^{(1)}$, then $\widetilde{\mathcal{L}}.f = 0$, so the projection $\lambda_1^{(1)}$ is centered, i.e. $\mathbb{E}(\eta^{(1)}) = X$. If $(E'_i : \mathfrak{m}^{(1)} \rightarrow \mathbb{R})$ refers to the dual basis of (E_i) , then $\widetilde{\mathcal{L}}.f(E'_i E'_j) = \delta_{ij}$, whence $\lambda_1^{(1)}$ and η have covariance matrix the identity in the basis (E_i) . Finally, if f is a linear form on $\mathfrak{m}^{(2)}$, then $\widetilde{\mathcal{L}}.f = f(B)$, so $\eta(f) = \lambda_1(f) = f(B)$. This concludes the proof. \square

We now prove Theorems 1.1, 1.6. With our current notations, they together correspond to Theorem 5.2 below. Recall that $X = \mathbb{E}(\mu^{(1)})$.

THEOREM 5.2 (CLT + Berry-Esseen). *Keep the above notations. Let $f \in C_c^\infty(\mathfrak{g})$ be a smooth function with compact support. Then for $N \geq 1$,*

$$|D_{\frac{1}{\sqrt{N}}}(\mu^{*N} * -NX)(f) - \nu(f)| \leq \varepsilon_N \max_{|\alpha| \leq \dim \mathfrak{g} + 4} \|\partial^\alpha f\|_{L^1}$$

where the notation ε_N stands for $\varepsilon_N = o(1)$. *Moreover, if μ has finite third moment for the weight filtration, then we can take $\varepsilon_N \leq C(1 + m_2(\mu))^{6s^2}(1 + m_3(\mu))N^{-1/2}$ where $C = C(\mathfrak{g}, [\cdot, \cdot], \mathbb{R}\overline{X}, (\mathfrak{m}^{(b)})_b, (e_j^{(i)})_{i,j}) > 0$.*

PROOF OF THEOREM 5.2. First we prove the statement without additional moment assumption, in particular ε_N stands for $o(1)$. By Fourier Inversion formula, we have

$$D_{\frac{1}{\sqrt{N}}}(\mu^{*N} * -NX)(f) = \int_{\widehat{\mathfrak{g}}} \widehat{f}(\xi) \widehat{\mu_{-NX}^{*N}}(D_{\frac{1}{\sqrt{N}}}\xi) d\xi$$

where a linear form $\xi \in \widehat{\mathfrak{g}}$ is seen as a linear form on $\widetilde{\mathfrak{g}}$ by setting $\xi(\chi) = 0$, and where $\widehat{\mu_{-NX}^{*N}} = \widehat{\mu^{*N}} * -N\chi$.

Using the truncation estimate from Lemma 4.2, followed by the graded replacement and the Lindeberg replacement from Corollary 4.5, Proposition 4.7 applied with $\delta_0 = 0$ and to the measure η from Lemma 5.1, we deduce

$$\begin{aligned} D_{\frac{1}{\sqrt{N}}}(\mu^{*N} * -NX)(f) &= \int_{\widehat{\mathfrak{g}}} \widehat{f}(\xi) (T_N \widehat{\mu}_{-NX}^{*N})'(D_{\frac{1}{\sqrt{N}}}\xi) d\xi + \varepsilon_N \int_{\mathfrak{g}} |\widehat{f}(\xi)| d\xi \\ &= \int_{\widehat{\mathfrak{g}}} \widehat{f}(\xi) (T_N \widehat{\eta}_{-NX}^{*N})'(D_{\frac{1}{\sqrt{N}}}\xi) d\xi + \varepsilon_N \int_{\mathfrak{g}} |\widehat{f}(\xi)|(1 + \|\xi\| + \|\xi\|^3) d\xi \\ &= \int_{\widehat{\mathfrak{g}}} \widehat{f}(\xi) \widehat{\eta_{-NX}^{*N}}(D_{\frac{1}{\sqrt{N}}}\xi) d\xi + \varepsilon_N \int_{\mathfrak{g}} |\widehat{f}(\xi)|(1 + \|\xi\| + \|\xi\|^3) d\xi \\ &= \nu(f) + \varepsilon_N \int_{\mathfrak{g}} |\widehat{f}(\xi)|(1 + \|\xi\| + \|\xi\|^3) d\xi \end{aligned}$$

where the last line uses that $\widehat{\eta_{-NX}^{*N}} = (D_{\sqrt{N}}\nu)' * N\chi$.

We can finally bound the dependence in f in terms of the L^1 -norm of derivatives of f . Indeed, using integration by parts, one sees that for all $k \geq 0$, $\xi \in \widehat{\mathfrak{g}} \setminus \{0\}$,

$$|\widehat{f}(\xi)| \ll_k \frac{1}{\|\xi\|^k} \sup_{|\alpha|=k} \|\partial^\alpha f\|_{L^1}$$

where ∂^α varies among the order k additive derivatives of f with respect to a fixed arbitrary basis. It follows that

$$\int_{\widehat{\mathfrak{g}}} \widehat{f}(\xi)(1 + \|\xi\| + \|\xi\|^3) d\xi = O(1) \sup_{|\alpha| \leq \dim \mathfrak{g} + 4} \|\partial^\alpha f\|_{L^1},$$

which concludes the first part of the proof.

We now turn to the case where μ has finite third moment. Inspecting the above argument, and using the description of constants involved in Lemma 4.2, Corollary 4.5, Proposition 4.7, it appears we may take

$$(36) \quad \varepsilon_N \leq C(1 + m_2(\mu)^{6s^2} + m_2(\eta)^{6s^2})(1 + m_3(\mu) + m_3(\eta)).$$

It remains to bound the moments of η using those of μ . We will do it in Lemma 5.3 and Proposition 5.4 below. Then Theorem 5.2 follows by combining (36) and Proposition 5.4. \square

In order to study the moments of η , we introduce

$$a(\widetilde{\eta}) := \int_{\widehat{\mathfrak{g}}} \|x^{(1)}\|^2 + \|x^{(2)}\| d\widetilde{\eta}(x) = m_2(\widetilde{\eta}^{(1)}) + m_1(\widetilde{\eta}^{(2)})$$

where we recall that $\widetilde{\eta}^{(i)} = \pi^{(i)}(\widetilde{\eta})$ is the projection of $\widetilde{\eta}$ to $\widetilde{\mathfrak{m}}^{(i)}$ parallel to $\bigoplus_{j \neq i} \widetilde{\mathfrak{m}}^{(j)}$. These notations also make sense for μ in place of $\widetilde{\eta}$. We start with a preliminary bound, relating $a(\widetilde{\eta})$ to $a(\mu)$.

LEMMA 5.3. *We have*

$$\|X\| \leq a(\mu)^{1/2} \quad \text{and} \quad a(\widetilde{\eta}) \leq C(1 + a(\mu))$$

where $C = C(\mathfrak{g}, [\cdot, \cdot], \mathbb{R}\overline{X}, (\mathfrak{m}^{(b)})_b, (e_j^{(i)})_{i,j}) > 0$.

PROOF. The estimate $\|X\| \leq m_1(\mu^{(1)}) \leq a(\mu)^{1/2}$ follows from the Cauchy-Schwarz inequality. Let us check the second one. Note that $\widetilde{\eta}^{(1)}$ is a centered probability measure on $\mathfrak{m}^{(1)}$ with same covariance as $\mu^{(1)}$. Hence $m_2(\widetilde{\eta}^{(1)}) = m_2(\mu^{(1)} - X) \leq m_2(\mu^{(1)})$. To conclude, we check that $m_1(\widetilde{\eta}^{(2)}) \leq C(1 + a(\mu))$. In the first part of the proof of Theorem 5.2, we established (without rate) the graded CLT

$$D_{\frac{1}{\sqrt{n}}}(\widetilde{\mu}^{*n}) \rightarrow \widetilde{\eta}.$$

It yields

$$(37) \quad \begin{aligned} \int_{\mathfrak{g}} \|x^{(2)}\| d\widetilde{\eta}(x) &\leq \liminf_n \frac{1}{n} \int_{\mathfrak{g}^n} \|(g_1 *' \dots *' g_n)^{(2)}\| d\widetilde{\mu}^{\otimes n}(\underline{g}) \\ &= \liminf_n \frac{1}{n} \int_{\mathfrak{g}^n} \|g_1^{(2)} + \dots + g_n^{(2)} + \sum_{i < j} [g_i^{(1)}, g_j^{(1)}]'\| d\widetilde{\mu}^{\otimes n}(\underline{g}). \end{aligned}$$

On the one hand, by the triangle inequality and recalling $\|X\| \leq m_2(\mu^{(1)})^{1/2} \leq 1 + a(\mu)$, we have

$$\int_{\mathfrak{g}^n} \|g_1^{(2)} + \dots + g_n^{(2)}\| d\widetilde{\mu}^{\otimes n}(\underline{g}) \leq 2n(1 + a(\mu)).$$

On the other hand, by the Cauchy-Schwarz inequality,

$$(38) \quad \int_{\mathfrak{g}^n} \left\| \sum_{i < j} [g_i^{(1)}, g_j^{(1)}]'\right\| d\widetilde{\mu}^{\otimes n}(\underline{g}) \leq \left(\int_{\mathfrak{g}^n} \left\| \sum_{i < j} [g_i^{(1)}, g_j^{(1)}]'\right\|^2 d\widetilde{\mu}^{\otimes n}(\underline{g}) \right)^{1/2}$$

then recalling that the norm $\|\cdot\|$ is Euclidean, and using that $\tilde{\mu}^{(1)}$ is centered, we get

$$\begin{aligned}
\int_{\mathfrak{g}^n} \left\| \sum_{i < j} [g_i^{(1)}, g_j^{(1)}]' \right\|^2 d\tilde{\mu}^{\otimes n}(\underline{g}) &= \int_{\mathfrak{g}^n} \sum_{i < j} \sum_{k < \ell} \langle [g_i^{(1)}, g_j^{(1)}]', [g_k^{(1)}, g_\ell^{(1)}]' \rangle d\tilde{\mu}^{\otimes n}(\underline{g}) \\
&= \int_{\mathfrak{g}^n} \sum_{i < j} \|[g_i^{(1)}, g_j^{(1)}]'\|^2 d\tilde{\mu}^{\otimes n}(\underline{g}) \\
(39) \qquad \qquad \qquad &\leq n^2 C(1 + m_2(\mu^{(1)}))^2
\end{aligned}$$

Combining (37), (38), (39) concludes the proof. \square

Using the previous lemma, we can bound all moments of η using only the quantity $a(\mu)$ (and in particular the second moment of μ).

PROPOSITION 5.4 (Bound for all moments). *Given $r > 0$, we have*

$$m_r(\eta) \leq C(1 + a(\mu)^{r/2}) \leq C(1 + m_r(\mu))$$

where $C > 0$ only depends on $\mathfrak{g}, [\cdot, \cdot], \mathbb{R}\bar{X}, (\mathfrak{m}^{(b)})_b, (e_j^{(i)})_{i,j}$ and an upper bound on r .

The proof is inspired by [BP96].

PROOF. We first establish that for any $\alpha \in \mathbb{N}^{\dim \mathfrak{g}}$, the $L^1(\tilde{\eta})$ -norm the monomial x^α is related to that of monomial of lower orders. Assume the degree of α satisfies $d(\alpha) \geq 3$. Using the invariance relation $\tilde{\eta} = D_{\frac{1}{\sqrt{2}}}(\tilde{\eta} *' \tilde{\eta})$, we have

$$\begin{aligned}
\int_{\mathfrak{g}} |x^\alpha| d\tilde{\eta}(x) &= \frac{1}{2^{d(\alpha)/2}} \int_{\mathfrak{g}} |(x *' y)^\alpha| d\tilde{\eta}(x) d\tilde{\eta}(y) \\
&\leq \frac{1}{2^{d(\alpha)/2}} \left(2 \int_{\mathfrak{g}} |x^\alpha| d\tilde{\eta}(x) + C' \sum_{\substack{d(\beta)+d(\gamma)=d(\alpha) \\ d(\beta), d(\gamma) < d(\alpha)}} \int_{\mathfrak{g}} |x^\beta| d\tilde{\eta}(x) \int_{\mathfrak{g}} |x^\gamma| d\tilde{\eta}(x) \right).
\end{aligned}$$

for some $C' = C'(\mathbf{D}, \alpha) \geq 0$, where \mathbf{D} refers to the data $\mathbf{D} = (\mathfrak{g}, [\cdot, \cdot], \mathbb{R}\bar{X}, (\mathfrak{m}^{(b)})_b, (e_j^{(i)})_{i,j})$. Using $d(\alpha) \geq 3$ to guarantee $1 - \frac{1}{2^{d(\alpha)/2-1}} \geq 1/10$, we get

$$\int_{\mathfrak{g}} |x^\alpha| d\tilde{\eta}(x) \leq 10C' \sum_{\substack{d(\beta)+d(\gamma)=d(\alpha) \\ d(\beta), d(\gamma) < d(\alpha)}} \int_{\mathfrak{g}} |x^\beta| d\tilde{\eta}(x) \int_{\mathfrak{g}} |x^\gamma| d\tilde{\eta}(x) \leq C'' a(\tilde{\eta})^{d(\alpha)/2} \leq C'''(1 + a(\mu))^{d(\alpha)/2}.$$

where the second inequality is obtained by iterating the first yielding $C'' = C''(\mathbf{D}, \alpha) > 0$, while the third inequality follows from Lemma 5.3. The same holds for η instead of $\tilde{\eta}$ due to the relation $\eta = \tilde{\eta} - \chi + X$ and the estimate $\|\chi\| = \|X\| \leq a(\mu)^{1/2}$.

This does not justify yet the formula that was announced in the statement because we have only looked at monomials of integer degree so far. So let $r > 0$. Write $2k_i$ the smallest even integer greater than r/i . Using the Hölder inequality, then the previous paragraph, we get

$$\begin{aligned}
\int_{\mathfrak{g}} \|x^{(i)}\|^{r/i} d\eta(x) &\leq \left(\int_{\mathfrak{g}} \|x^{(i)}\|^{2k_i} d\eta(x) \right)^{r/(i2k_i)} \\
&\leq C(1 + a(\mu)^{ik_i \cdot r/(i2k_i)}) \\
&= C(1 + a(\mu)^{r/2})
\end{aligned}$$

where $C = C(\mathbf{D}, r) > 0$. The second estimate in the statement follows by Hölder inequality. This concludes the proof of Proposition 5.4 (whence that of Theorem 5.2). \square

We conclude this section with a discussion of several papers that preceded us in attempting to prove the central limit theorem for non-centered random walks on a general nilpotent Lie group. The case of the Heisenberg group was handled successfully by Tutubalin [Tut64] but a substantial difficulty remained to formulate the correct renormalization for an arbitrary group.

In [Vir74], Virtser proposes a renormalization for walks on the full upper-triangular unipotent matrix group, i.e. $\mathfrak{g} = \sum_{1 \leq i < j \leq n} \mathbb{R}E_{i,j}$ with $[E_{i,j}, E_{k,\ell}] = \mathbb{1}_{j=k}E_{i,\ell} - \mathbb{1}_{i=\ell}E_{k,j}$. While Virtser's normalization is appropriate (and coincides with ours) for certain values of \bar{X} , as in [Vir74, Theorem 2 & 3], for certain other values his renormalization is too strong and may lead to non absolutely continuous limiting measures. This occurs for instance if $n = 4$, and $\bar{X} = sE_{1,2} + tE_{3,4} \pmod{[\mathfrak{g}, \mathfrak{g}]}$ with $s, t \neq 0$. Indeed, in this case direct computation shows that our subspace of weight at least 3 is $\mathfrak{g}^{(3)} = \mathbb{R}E_{1,4} + \mathbb{R}(E_{1,3} - t/sE_{2,4}) \subsetneq \mathfrak{g}^{[2]}$. However, Virtser renormalizes the whole subspace $\mathfrak{g}^{[2]}$ with coefficient $N^{-3/2}$, hence shrinking to a point the coordinate $\mathfrak{g}^{[2]}/\mathfrak{g}^{(3)}$ which lives at scale N .

In [Rau78], Raugi proposes a renormalization for walks on a general nilpotent Lie group. We explain below why his renormalization is too weak to prevent escape of mass in non-centered cases, despite the CLT formulated in his paper.

Raugi's renormalization is defined with respect to the sequence of ideals $\{\mathfrak{J}^{k,\ell}, 0 \leq \ell < k \leq s\}$, where $\mathfrak{J}^{k,\ell}$ is the linear span of $\mathfrak{g}^{[k+1]}$ and brackets containing k elements of \mathfrak{g} , at least ℓ of which are in $\mathbb{R}\bar{X}$. One also authorizes the case $(k, \ell) = (1, 1)$ and set $\mathfrak{J}^{(1,1)} = \mathfrak{g}^{[2]} + \mathbb{R}\bar{X}$. The ideals $\mathfrak{J}^{k,\ell}$ are then ordered by lexicographical order, i.e.

$$\mathfrak{J}^{1,0} \supseteq \mathfrak{J}^{1,1} \supseteq \mathfrak{J}^{2,0} \supseteq \mathfrak{J}^{2,1} \supseteq \dots \supseteq \mathfrak{J}^{k,0} \supseteq \mathfrak{J}^{k,1} \supseteq \dots \supseteq \mathfrak{J}^{k,k-1} \supseteq \dots \supseteq \mathfrak{J}^{s+1,0} = \{0\}$$

and for each each (k, ℓ) above, of successor $(k, \ell)_+$, one may choose a supplementary subspace $\mathfrak{r}^{k,\ell}$ such that

$$\mathfrak{J}^{k,\ell} = \mathfrak{r}^{k,\ell} \oplus \mathfrak{J}^{(k,\ell)_+}$$

in particular $\mathfrak{g} = \bigoplus \mathfrak{r}^{k,\ell}$. Raugi's system of dilation is the family R_t of linear maps defined on $\mathfrak{r}^{k,\ell}$ by $R_t x = t^{k+\ell}x$ except on $\mathfrak{r}^{1,1}$ where $R_t x = tx$. The *main problem* with Raugi's definition is that the weight of the dilation may be strictly subadditive when taking brackets: $[\mathfrak{r}^{k_1,\ell_1}, \mathfrak{r}^{k_2,\ell_2}]$ is not necessarily included in $\bigoplus \{\mathfrak{r}^{k,\ell} : k + \ell \geq k_1 + \ell_1 + k_2 + \ell_2\}$. This leaves room for escape of mass, contradicting [Rau78]. We now formulate a precise counterexample where this phenomenon occurs.

Counterexample to [Rau78]. Let \mathfrak{l} be the step-3 filiform algebra defined by $\mathfrak{l} = \bigoplus_{i=1}^4 \mathbb{R}e_i$ where $[e_1, e_2] = e_3$, $[e_1, e_3] = e_4$, $[e_2, e_3] = 0$, $e_4 \in \mathfrak{z}(\mathfrak{l})$. In particular,

$$\mathfrak{l} = \mathbb{R}e_1 \oplus \mathbb{R}e_2 \oplus \mathbb{R}[e_1, e_2] \oplus \mathbb{R}[e_1, [e_1, e_2]]$$

Let \mathfrak{l}' be a nilpotent Lie algebra of step 4, and $(e'_j)_{1 \leq j \leq d'}$ a basis adapted to its central descending series, in particular $e_{d'} \in \mathfrak{l}'^{[4]}$. Endow the vector space $\mathfrak{g} := \mathfrak{l} \oplus \mathfrak{l}'$ with the bracket $[\cdot, \cdot]$ defined by $[\mathfrak{l}, \mathfrak{l}'] = 0$, extending the bracket on \mathfrak{l}' , and whose restriction to \mathfrak{l} corresponds to the original bracket on \mathfrak{l} modified by imposing $[e_2, e_3] = e'_d$. Direct computation shows that $(\mathfrak{g}, [\cdot, \cdot])$ is indeed a Lie algebra (of step 4). Now let μ be a compactly supported non-lattice probability measure on \mathfrak{g} with expectation $\mathbb{E}(\mu_{ab}) = e_2 \pmod{[\mathfrak{g}, \mathfrak{g}]}$ in the abelianization. Notice $\mathfrak{J}^{3,2} = \mathfrak{J}^{4,0} = \mathfrak{l}'^{[4]}$ and $\mathfrak{J}^{4,1} = \{0\}$. Given an arbitrary choice of $(\mathfrak{r}^{k,\ell})_{k,\ell}$, Raugi renormalises the projection of $\mu^{*n} * -NX$ to $\mathfrak{r}^{4,0} = \mathfrak{l}'^{[4]}$ by $N^{-4/2}$, but Theorem 5.2 (and even Crepel-Raugi [CR78] in this particular case) predicts that this projection is spread at scale $N^{5/2}$, whence the escape of mass.

5.2. *Asymptotically close measures.* We deduce from our central limit theorem a necessary and sufficient condition for two measures on \mathfrak{g} to be asymptotically close, namely Corollary 1.5. We keep the notations from the beginning of Section 5 and start with a variant of the CLT whose recentering is not multiplicative, but additive.

LEMMA 5.5 (CLT with additive recentering). *As N goes to $+\infty$, we have*

$$D_{\frac{1}{\sqrt{N}}}(\mu^{*N} - NX) \longrightarrow (\nu *' \chi) - \chi$$

We emphasize that $D_{\frac{1}{\sqrt{N}}}(\mu^{*N} - NX)$ differs from the measure $D_{\frac{1}{\sqrt{N}}}(\mu^{*N} * -NX)$ that we studied until now. Indeed, $D_{\frac{1}{\sqrt{N}}}(\mu^{*N} - NX)$ refers to the image of μ^{*N} by the map $\mathfrak{g} \rightarrow \mathfrak{g}, x \mapsto D_{\frac{1}{\sqrt{N}}}(x - NX)$ (see conventions in Section 2.9), so here the recentering refers to the addition on \mathfrak{g} seen as a vector space, and not the Lie product $*$. In a similar way, the measure in the limit is the image of ν by the map $\tilde{\mathfrak{g}} \rightarrow \tilde{\mathfrak{g}}, x \mapsto (x *' \chi) - \chi$ which induces a diffeomorphism of \mathfrak{g} . In particular, it is also a smooth probability measure on \mathfrak{g} .

PROOF. We observe that $\mu^{*N} - NX = \tilde{\mu}^{*N} - N\chi$, so we just need to check that $D_{\frac{1}{\sqrt{N}}}\tilde{\mu}^{*N} \rightarrow \nu *' \chi$. By the graded replacement lemma (Corollary 4.5), we may replace $D_{\frac{1}{\sqrt{N}}}\tilde{\mu}^{*N}$ by $D_{\frac{1}{\sqrt{N}}}\tilde{\mu}^{*'N}$, so it boils down to showing the convergence $D_{\frac{1}{\sqrt{N}}}(\tilde{\mu}^{*'N} *' - N\chi) \rightarrow \nu$. Applying Corollary 4.5 again, we are left to check that $D_{\frac{1}{\sqrt{N}}}(\tilde{\mu}^{*'N} * - N\chi) \rightarrow \nu$. As $\tilde{\mu}^{*'N} * - N\chi = \mu^{*N} * - NX$, this follows from the central limit theorem 5.2. \square

PROOF OF COROLLARY 1.5. If two measures are asymptotically close, then so are their projections to the abelianization. Combining the classical CLT and the fact that the class of convex subsets is invariant by dilations and translations, one deduces that those projections must have same mean and covariance matrix. In particular, the weight filtrations for both measures coincide and we may thus choose a common family of dilations and recentering for both measures. Using again the invariance properties of \mathcal{C} , the limiting distributions in Lemma 5.5 must coincide, hence the corresponding measures ν as well, or more generally the underlying diffusion laws σ . Examining infinitesimal generators, we conclude that the commutator means, i.e. the B term in \mathcal{L} , also coincide.

The converse statement is a direct consequence of the fact that \mathcal{C} is closed under dilations and translations, Lemma 5.5, and the fact that weak convergence of measures toward an absolutely continuous distribution implies uniform convergence on the class \mathcal{C} [RR62, Theorem 4.2]. \square

5.3. *Convergence of processes.* In this subsection, we prove Theorem 1.7, namely the analog of Donsker's invariance principle [RY99, Theorem XIII.1.9] for the μ -walk on $(\mathfrak{g}, *)$. For $N \geq 1$, $t \geq 0$, we recall the interpolated random process $W^{(N)} : \mathbb{R}^+ \rightarrow \mathfrak{g}$ defined by

$$W^{(N)}(t) = D_{\frac{1}{\sqrt{N}}}(X_1 * \cdots * X_{[tN]} * (tN - [tN])X_{[tN]+1} * -tNX)$$

where the X_i 's are i.i.d. with law μ , and $X = \mathbb{E}(\mu^{(1)})$ is the lift to $\mathfrak{m}^{(1)}$ of $\bar{X} = \mathbb{E}(\mu_{ab})$. We also recall that we have associated to μ a $*$ -left-invariant diffusion process on $(\mathfrak{g}, *)$, namely the time-dependent diffusion $(W(t))_{t \geq 0}$ with infinitesimal generator $\mathcal{L}_t(\mu, (\mathfrak{m}^{(b)})_b)$ and law σ defined in (34).

Our goal is to show that $W^{(N)}$ converges in law to W in the topology of α -Hölder convergence on bounded intervals. We begin by rewriting the process as:

$$(40) \quad W^{(N)}(t) = D_{\frac{1}{\sqrt{N}}}(x_1 * \cdots * x_{[tN]} * (tN - [tN])x_{[tN]+1} * -tNX)$$

where $x_i = X_i - X + \chi$ is now a family of i.i.d. variables with law $\tilde{\mu}$. Then we set

$$S^{(N)}(t) = D_{\frac{1}{\sqrt{N}}} (x_1 *' \cdots *' x_{[tN]} *' (tN - [tN])x_{[tN]+1})$$

From Corollary 4.6 and the remark following it, Theorem 1.7 will reduce to the following. We recall that $Z(t) = W(t) *' t\chi$.

PROPOSITION 5.6. *Assume that μ has finite m -th moment for any $m > 0$. Given any $\alpha \in [0, \frac{1}{2})$, the process $S^{(N)}$ converges in law towards the process Z in $C^{0,\alpha}(\mathbb{R}_+, \tilde{\mathfrak{g}})$ endowed with the topology of α -Hölder convergence on bounded intervals.*

We shall use a classical approach, via Kolmogorov's tightness criterion [RY99, Chapter VIII]. This method reduces the proof of Proposition 5.6 to the combination of the next two lemmas.

LEMMA 5.7 (Convergence of finite-dimensional distributions). *For every $k \geq 1$, $0 \leq t_1 < \cdots < t_k < \infty$, one has the convergence in law, as $N \rightarrow +\infty$:*

$$(S^{(N)}(t_i))_{i \leq k} \longrightarrow (Z(t_i))_{i \leq k}.$$

PROOF. For $N \geq 1$, denote by $t_i^{(N)} = \inf\{k/N : k/N \geq t_i\}$. Then, for all $\varepsilon > 0$,

$$\mathbb{P}(\|S^{(N)}(t_i) - S^{(N)}(t_i^{(N)})\| > \varepsilon) = o(1) \quad \mathbb{P}(\|Z(t_i) - Z(t_i^{(N)})\| > \varepsilon) = o(1)$$

so it is sufficient to prove that for any $F : \tilde{\mathfrak{g}}^k \rightarrow \mathbb{R}$ continuous and bounded,

$$(S^{(N)}(t_i^{(N)}))_{i \leq k}(F) - (Z(t_i^{(N)}))_{i \leq k}(F) = o_F(1)$$

But, setting $s_i^{(N)} = t_i^{(N)} - t_{i-1}^{(N)}$ (where $t_0^{(N)} = 0$), one may rewrite

$$(S^{(N)}(t_i^{(N)}))_{i \leq k} = \Psi(\mathcal{S}_1^{(N)}, \dots, \mathcal{S}_k^{(N)}) \quad (Z(t_i^{(N)}))_{i \leq k} = \Psi(\mathcal{Z}_1^{(N)}, \dots, \mathcal{Z}_k^{(N)})$$

where $\Psi : \tilde{\mathfrak{g}}^k \rightarrow \tilde{\mathfrak{g}}^k, (x_1, \dots, x_k) \mapsto (x_1 *' \cdots *' x_i)_{i \leq k}$ is a continuous map, the variables $\mathcal{S}_1^{(N)}, \mathcal{Z}_1^{(N)}, \dots, \mathcal{S}_k^{(N)}, \mathcal{Z}_k^{(N)}$ are mutually independent, with each $\mathcal{S}_i^{(N)}$ (resp. $\mathcal{Z}_i^{(N)}$) distributed as $S^{(N)}(s_i^{(N)})$ (resp. $Z(s_i^{(N)})$). To conclude, it is now sufficient to check that $\mathcal{S}_i^{(N)} \rightarrow \mathcal{Z}_i^{(N)}$ for each i , but this follows from the central limit theorem 5.2. \square

LEMMA 5.8 (Tightness). *Let d' be a left-invariant Riemannian metric on $(\mathfrak{g}, *')$. For every segment $[0, R]$, every integer $m \geq 1$ there is $C_m > 0$ such that for all $N \geq 1$,*

$$(41) \quad \sup_{0 \leq s < t \leq R} \mathbb{E} \left(d'(S_s^{(N)}, S_t^{(N)})^{2m} \right) \leq C_m |s - t|^m.$$

PROOF. Suppose $s = \frac{i}{N}, t = \frac{i+j}{N}$ for integers $i, j \geq 0$. Then

$$\begin{aligned} d'(S_s^{(N)}, S_t^{(N)}) &= d'(0, -S_s^{(N)} *' S_t^{(N)}) = d'(0, D_{\frac{1}{\sqrt{N}}}(x_{i+1} *' \cdots *' x_{i+j})) \\ &\ll_{d'} \|D_{\frac{1}{\sqrt{N}}}(x_{i+1} *' \cdots *' x_{i+j})\| \end{aligned}$$

For any $m \geq 1$, Proposition 4.3 and the remark following it tells us that

$$\mathbb{E} \left(\|D_{\frac{1}{\sqrt{N}}}(x_{i+1} *' \cdots *' x_{i+j})\|^{2m} \right) \ll_m \left(\frac{j}{N} \right)^m = |t - s|^m,$$

as desired. In general with $i := \lfloor Ns \rfloor$, $i + j = \lfloor Nt \rfloor$, we may write

$$d'(S_s^{(N)}, S_t^{(N)})^{2m} \ll_m d'(S_s^{(N)}, S_{i/N}^{(N)})^{2m} + d'(S_{i/N}^{(N)}, S_{(i+j)/N}^{(N)})^{2m} + d'(S_{(i+j)/N}^{(N)}, S_t^{(N)})^{2m}$$

Taking expectations, this becomes $O_{d',m}(1)|t - s|^m$, as we may see after distinguishing the cases $|t - s| < 1/N$ and $|t - s| \geq 1/N$. \square

PROOF OF PROPOSITION 5.6. According to Kolmogorov's tightness criterion and the proof of Theorem 2.1 in [RY99] (using the Riemannian metric d' in place of the Euclidean metric), Lemma 5.8 implies that the sequence of processes $S^{(N)}$ is relatively compact in the topology of $C^{0,\alpha}(\mathbb{R}_+, \tilde{\mathfrak{g}})$ for any $\alpha < \frac{m-1}{2m}$. Together with Lemma 5.7 this ensures the desired convergence in law to λ . \square

PROOF OF THEOREM 1.7. Let $W'^{(N)}$ be the process obtained from the formula (40) for $W^{(N)}$ by replacing the product $*$ by the product $*'$. From Corollary 4.6 we see that the sequence of processes $W^{(N)}$ and $W'^{(N)}$ have the same limit in the α -Hölder topology for each $\alpha \in [0, \frac{1}{2})$.

On the other hand, $W'^{(N)}(t) = S^{(N)}(t) *' -t\chi$, and the operator on $C^{0,\alpha}(\mathbb{R}, \tilde{\mathfrak{g}})$ defined by $c(t) \mapsto c(t) *' t\chi$ is continuous. Hence it follows from Proposition 5.6 that $W'^{(N)}$ converges in law to the random process $(Z(t) *' -t\chi)_{t \geq 0}$, which is just $(W(t))_{t \geq 0}$ by definition. \square

Remark. When $\alpha = 0$, Proposition 5.6 can also be derived from the main result of [SV73].

Remark. (Rough paths) We note in passing that, in fact, a stronger version of Proposition 5.6 holds. Namely, for a suitable choice of dilations $D_{\frac{1}{\sqrt{N}}}$ (i.e. of subspaces $(\tilde{\mathfrak{m}}^{(b)})_b$), the convergence also holds in the finer *rough paths* topology, see [LQ02, FV10, FH14], namely in the Polish space of α -Hölder paths with respect to the $*'$ -invariant Carnot-Carathéodory metric $d'_{cc}(x, y)$ defined on $(\tilde{\mathfrak{g}}, *')$ by any Euclidean norm on $\tilde{\mathfrak{m}}^{(1)} \oplus \mathbb{R}\chi$. This is defined exactly as the Riemannian metric, except that we only allow paths that are tangent to the distribution of planes formed by the left translates of $\tilde{\mathfrak{m}}^{(1)} \oplus \mathbb{R}\chi$. If $\tilde{\mathfrak{m}}^{(b)}$ is chosen so that it lies in the span V_b of all brackets of length at most b in $\tilde{\mathfrak{m}}^{(1)} \oplus \mathbb{R}\chi$ (which is possible thanks to Lemma 3.9), then for small x (see “ball-box theorem”, [Bel96, Thm 7.3.4], [Gro96, Thm 0.5.A, p97])

$$d'_{cc}(0, x) \simeq \max_{b \geq 0} d_{eucl}(x, V_b)^{1/(b+1)} \ll \max_{b \geq 1} \|x^{(b)}\|^{1/b},$$

where $V_0 := 0$, while for large x , $d'_{cc}(0, x)$ is controlled by $d_{eucl}(0, x)$, where d_{eucl} is the Euclidean distance on $\tilde{\mathfrak{g}}$. The same proof then shows Lemma 5.8 with d'_{cc} in place of d' . As a consequence the sample paths of the diffusion λ are α -Hölder for the Carnot-Carathéodory metric d'_{cc} for each $\alpha \in [0, \frac{1}{2})$. In the centered case on the free nilpotent Lie group, Donsker's theorem in the rough paths topology was first obtained in [BFH09] by a similar method.

Acknowledgments. We warmly thank Yves Benoist, Peter Friz and Laurent Saloff-Coste for very helpful discussions, as well as the anonymous referee for their useful comments.

Funding. The first author has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 803711).

REFERENCES

- [AS04] ANDREI A. AGRACHEV AND YURI L. SACHKOV (2004). *Control theory from the geometric viewpoint*, volume **87** of *Encyclopaedia of Mathematical Sciences*. Springer-Verlag, Berlin, Control Theory and Optimization, II.
- [AKS93] SHIGEKI AIDA, SHIGEO KUSUOKA, AND DANIEL STROOCK (1993). On the support of Wiener functionals. In *Asymptotic Problems in Probability Theory: Wiener Functionals and Asymptotics*, K. D. Elworthy and N. Ikeda (eds.), Pitman Research Notes in Math. Series **284**, 3–34. Longman Scient. and Tech., New York.
- [Ale02a] GEORGIOS ALEXOPOULOS (2002). Centered densities on Lie groups of polynomial volume growth. *Probab. Theory Related Fields*, **124**(1):112–150.
- [Ale02b] GEORGIOS ALEXOPOULOS (2002). Random walks on discrete groups of polynomial volume growth. *Ann. Probab.*, **30**(2):723–801.
- [Ale02c] GEORGIOS ALEXOPOULOS (2002). Sub-Laplacians with drift on Lie groups of polynomial volume growth. *Mem. Amer. Math. Soc.*, **155**(739):x+101, 2002.
- [AG16] ASAAD, MALVA AND GORDINA, MARIA (2016). Hypoelliptic heat kernels on nilpotent Lie groups. In *Potential Analysis*, **45**, no. 2, 355–386.
- [Bel96] ANDRÉ BELLAÏCHE (1996). The tangent space in sub-Riemannian geometry. In *Sub-Riemannian geometry*, volume **144** of *Progr. Math.*, pages 1–78. Birkhäuser, Basel.
- [Ben89] GÉRARD BEN AROUS (1989). Flots et séries de Taylor stochastiques. In *Probab. Theory Related Fields*, **81**, no. 1, 29–77.
- [BeLe91] GÉRARD BEN AROUS AND RÉMI LÉANDRE (1991). Décroissance exponentielle du noyau de la chaleur sur la diagonale, I, II. *Probab. Theory Related Fields*, **90**, no. 2, 175–202. and no. 3, 377–402.
- [BB23] TIMOTHÉE BÉNARD AND EMMANUEL BREUILLARD. Local limit theorems for random walks on nilpotent Lie groups, arXiv:2304.14551.
- [Ben21] YVES BENOIST (2021). Positive harmonic functions on the Heisenberg group II. *J. Éc. Polytech., Math.*, **8**: 973–1003.
- [BQ16] YVES BENOIST AND JEAN-FRANÇOIS QUINT (2016). *Random walks on reductive groups*, volume 62 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics*. Springer, Cham.
- [BP96] VIDMANTAS BENTKUS AND GYULA PAP (1996). The accuracy of Gaussian approximations in nilpotent Lie groups. *J. Theoret. Probab.*, **9**(4):995–1017.
- [Bou81] PHILIPPE BOUGEROL (1982). Théorème central limite local sur certains groupes de Lie. *Ann. Sci. École Norm. Sup.*, **14**(4):403–432.
- [Bre04] EMMANUEL BREUILLARD (2004). *Equidistribution of random walks on nilpotent Lie groups and homogeneous spaces*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)–Yale University.
- [Bre05] EMMANUEL BREUILLARD (2005). Local limit theorems and equidistribution of random walks on the Heisenberg group. *Geom. Funct. Anal.*, **15**(1):35–82.
- [Bre10] EMMANUEL BREUILLARD (2010). Equidistribution of dense subgroups on nilpotent Lie groups. *Ergodic Theory Dynam. Systems*, **30**, no. 1, 131–150.
- [Bre14] EMMANUEL BREUILLARD (2014). Geometry of locally compact groups of polynomial growth and shape of large balls. *Groups Geom. Dyn.*, **8**(3):669–732.
- [BFH09] EMMANUEL BREUILLARD, PETER FRIZ, AND MARTIN HUESMANN (2009). From random walks to rough paths. *Proc. Amer. Math. Soc.*, **137**(10):3487–3496.
- [CCLP99] LUCIA CAMELLINO, ADRIANA CLIMESCU-HAULICA AND BARBARA PACCHIAROTTI (1999). Diffusion approximations for random walks on nilpotent Lie groups. *Statist. Probab. Lett.* **41**, no. 4, 363–377.
- [Cha06] SOURAV CHATTERJEE (2006). A generalization of the Lindeberg principle. *Ann. Probab.*, **34**(6):2061–2076.
- [Che77] KUO TSAI CHEN (1977). Iterated path integrals. *Bull. Amer. Math. Soc.*, **83**(5):831–879.
- [CHT21] YVES COLIN DE VERDIÈRE, LUC HILLAIRET, AND EMMANUEL TRÉLAT (2021). Small-time asymptotics of hypoelliptic heat kernels near the diagonal, nilpotentization and related results. *Ann. H. Lebesgue* **4**, 897–971.
- [CG90] LAWRENCE CORWIN AND FREDERICK GREENLEAF (1990). *Representations of nilpotent Lie groups and their applications. Part I*, volume **18** of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge. Basic theory and examples.
- [CR78] PIERRE CRÉPEL AND ALBERT RAUGI (1978). Théorème central limite sur les groupes nilpotents. *Ann. Inst. H. Poincaré Sect. B (N.S.)*, **14**(2):145–164.
- [DH21] PERSI DIACONIS AND ROBERT HOUGH (2021). Random walk on unipotent matrix groups. *Ann. Sci. Éc. Norm. Supér. (4)*, **54**(3):587–625.

- [DL57] JACQUES DIXMIER AND W. G. LISTER (1957). Derivations of nilpotent Lie algebras. *Proc. Amer. Math. Soc.*, **8**:155–158.
- [Dye70] JOAN DYER (1970). A nilpotent Lie algebra with nilpotent automorphism group. *Bull. Amer. Math. Soc.*, **76**:52–56.
- [Dyn47] EUGENE DYNKIN (1947). Calculation of the coefficients in the Campbell-Hausdorff formula. *Doklady Akad. Nauk SSSR (N.S.)*, **57**:323–326.
- [Fel68] WILLIAM FELLER (1968, 1971). *An introduction to probability theory and its applications*. Vol. I and II. John Wiley & Sons, Inc., New York-London-Sydney.
- [FH14] PETER FRIZ AND MARTIN HAIRER (2014). *A course on rough paths*. Universitext. Springer, Cham. With an introduction to regularity structures.
- [FV10] PETER FRIZ AND NICOLAS VICTOIR (2010). *Multidimensional stochastic processes as rough paths*, volume 120 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge. Theory and applications.
- [Gav77] BERNARD GAVEAU (1977). Principe de moindre action, propagation de la chaleur et estimées sous elliptiques sur certains groupes nilpotents. *Acta Math.*, **139**(1-2):95–153.
- [Gre63] ULF GRENANDER (1963). *Probabilities on algebraic structures*. John Wiley & Sons, Inc., New York-London; Almqvist & Wiksell, Stockholm-Göteborg-Uppsala.
- [Gro96] MIKHAELE GROMOV (1996). Carnot-Carathéodory spaces seen from within. In *Sub-Riemannian geometry*, volume **144** of *Progr. Math.*, pages 79–323. Birkhäuser, Basel.
- [Gui73] YVES GUIVARCH (1973). Croissance polynomiale et périodes des fonctions harmoniques. *Bull. Soc. Math. France*, **101**:333–379.
- [HS01] WILFRIED HAZOD AND EBERHARD SIEBERT (2001). *Stable probability measures on Euclidean spaces and on locally compact groups*, volume **531** of *Mathematics and its Applications*. Kluwer Academic Publishers, Dordrecht. Structural properties and limit theorems.
- [Heb89] WALDEMAR HEBISCH (1989). Sharp pointwise estimate for the kernels of the semigroup generated by sums of even powers of vector fields on homogeneous groups. *Studia Math.*, **95**(1):93–106.
- [Hör67] LARS HÖRMANDER (1967). Hypoelliptic second order differential equations. *Acta Math.*, **119**:147–171.
- [Hou19] ROBERT HOUGH (2019). The local limit theorem on nilpotent Lie groups. *Probab. Theory Relat. Fields*, **174**(3-4):761–786.
- [Hun56] G. A. HUNT (1956). Semi-groups of measures on Lie groups. *Trans. Amer. Math. Soc.*, **81**:264–293.
- [Ish03] SATOSHI ISHIWATA (2003). A central limit theorem on a covering graph with a transformation group of polynomial growth. *J. Math. Soc. Japan*, **55**(3):837–853.
- [IKN20] SATOSHI ISHIWATA, HIROSHI KAWABI, AND RYUYA NAMBA (2020). Central limit theorems for non-symmetric random walks on nilpotent covering graphs: Part I. *Electron. J. Probab.*, **25**:Paper No. 86, 46.
- [IKN21] SATOSHI ISHIWATA, HIROSHI KAWABI, AND RYUYA NAMBA (2021). Central limit theorems for non-symmetric random walks on nilpotent covering graphs: Part II. *Potential Anal.*, **55**(1):127–166.
- [Jør75] PALLE JØRGENSEN (1975). Representations of differential operators on a Lie group. *J. Functional Analysis*, **20**(2):105–135.
- [Kis79] JAN KISYŃSKI (1979). On semigroups generated by differential operators on Lie groups. *J. Functional Analysis*, **31**(2):234–244.
- [KP16] ALESSIA E. KOGOJ AND SERGIO POLIDORO (2016). Harnack inequality for hypoelliptic second order partial differential operators. *Potential Anal.*, **45**(3):545–555.
- [Kun80] HIROSHI KUNITA (1980). On the decomposition of solutions of stochastic differential equations. In *Stochastic integrals*. (Proc. Sympos., Univ. Durham), pp. 213–255, Lecture Notes in Math., **851**, Springer, Berlin.
- [Kup97] IVAN KUPKA (1997). Géométrie sous-riemannienne. In *Séminaire Bourbaki*, Vol. 1995/96. Astérisque No. **241**, Exp. No. 817, 5, 351–380.
- [Lév50] PAUL LÉVY (1950). Wiener’s random function, and other Laplacian random functions. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 171–187.
- [Lia04] MING LIAO (2004). *Lévy processes in Lie groups*, volume **162** of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge.
- [LQ02] TERRY LYONS AND ZHONGMIN QIAN (2002). *System control and rough paths*. Oxford Mathematical Monographs. Oxford University Press, Oxford. Oxford Science Publications.
- [MOO10] ELCHANAN MOSSEL, RYAN O’DONNELL, AND KRZYSZTOF OLESZKIEWICZ (2010). Noise stability of functions with low influences: invariance and optimality. *Ann. of Math. (2)*, **171**(1):295–341, 2010.
- [Nel61] EDWARD NELSON (1961). A proof of Liouville’s theorem. *Proc. Amer. Math. Soc.*, **12**:995.
- [Pap91] GYULA PAP (1991). Rate of convergence in CLT on stratified groups. *J. Multivariate Anal.*, **38**(2):333–365.

- [RR62] R. RANGA RAO (1962). Relations between weak and uniform convergence of measures with applications. *Ann. Math. Statist.*, **33**:659–680.
- [Rau78] ALBERT RAUGI (1978). Théorème de la limite centrale sur les groupes nilpotents. *Z. Wahrsch. Verw. Gebiete*, **43**(2):149–172.
- [Reu93] CHRISTOPHE REUTENAUER (1993). *Free Lie algebras*, volume 7 of *London Mathematical Society Monographs. New Series*. The Clarendon Press, Oxford University Press, New York. Oxford Science Publications.
- [RY99] DANIEL REVUZ AND MARC YOR (1999). *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition.
- [Rot75] V. I. ROTAR (1975). Limit theorems for multilinear forms and quasipolynomial functions. *Teor. Veroyatnost. i Primenen.*, **20**(3):527–546.
- [Sac07] YURI L. SACHKOV (2007). Control theory on Lie groups. *Sovrem. Mat. Fundam. Napravl.*, **27**:5–59.
- [SC01] LAURENT SALOFF-COSTE (2001). Probability on groups: random walks and invariant diffusions. *Notices Amer. Math. Soc.*, **48**(9):968–977.
- [ST66] VYACHESLAV SAZONOV AND VALERI TUTUBALIN (1966). Probability distributions on topological groups. *Teor. Veroyatnost. i Primenen.*, **11**:3–55.
- [Ser06] JEAN-PIERRE SERRE (2006). *Lie algebras and Lie groups*, volume **1500** of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin. 1964 lectures given at Harvard University, Corrected fifth printing of the second (1992) edition.
- [Sie82] EBERHARD SIEBERT (1982). Absolute continuity, singularity, and supports of Gauss semigroups on a Lie group. *Monatsh. Math.*, **93**(3):239–253.
- [Sie84] EBERHARD SIEBERT (1984). Densities and differentiability properties of Gauss semigroups on a Lie group. *Proc. Amer. Math. Soc.*, **91**(2):298–305.
- [Str87] ROBERT S. STRICHARTZ (1987). The Campbell-Baker-Hausdorff-Dynkin formula and solutions of differential equations. *J. Funct. Anal.*, **72**(2):320–345.
- [SV72] DANIEL W. STROOCK AND S. R. S. VARADHAN (1972). On the support of diffusion processes with applications to the strong maximum principle. *Proc. 6th Berkeley Sympos. math. Statist. Probab.*, Univ. Calif. 1970, **3**, 333-359.
- [SV73] DANIEL W. STROOCK AND S. R. S. VARADHAN (1973). Limit theorems for random walks on Lie groups. *Sankhyā Ser. A*, **35**(3):277–294.
- [Sus87] HECTOR SUSSMANN (1987). A general theorem on local controllability. *SIAM J. Control Optim.*, **25**(1):158–194.
- [Tut64] VALERI N. TUTUBALIN (1964). Compositions of measures on the simplest nilpotent group. *Teor. Veroyatnost. i Primenen.*, **9**:531–539.
- [Var88] NICHOLAS VAROPOULOS (1988). Analysis on Lie groups. *J. Funct. Anal.*, **76**(2):346–410, .
- [VSCC92] NICHOLAS VAROPOULOS, LAURENT. SALOFF-COSTE, AND THIERRY COULHON (1992). *Analysis and geometry on groups*, volume **100** of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge.
- [Var21] NICHOLAS VAROPOULOS (2021). *Potential theory and geometry on Lie groups*, volume **38** of *New Mathematical Monographs*. Cambridge University Press, Cambridge.
- [Vir74] ALEXANDER D. VIRTSER (1974). Limit theorems for compositions of distributions on certain nilpotent Lie groups. *Teor. Veroyatnost. i Primenen.*, **19**:84–103.
- [Weh62] DONALD WEHN (1962). Probabilities on Lie groups. *Proc. Nat. Acad. Sci. U.S.A.*, **48**:791–795.