The Annals of Probability
2010, Vol. 38, No. 6, 2224–2257
DOI: 10.1214/10-AOP544

RANDOM LIE GROUP ACTIONS ON COMPACT MANIFOLDS: A PERTURBATIVE ANALYSIS[1]

BY CHRISTIAN SADEL AND HERMANN SCHULZ-BALDES

Universität Erlangen–Nürnberg

A random Lie group action on a compact manifold generates a discrete time Markov process. The main object of this paper is the evaluation of associated Birkhoff sums in a regime of weak, but sufficiently effective coupling of the randomness. This effectiveness is expressed in terms of random Lie algebra elements and replaces the transience or Furstenberg's irreducibility hypothesis in related problems. The Birkhoff sum of any given smooth function then turns out to be equal to its integral w.r.t. a unique smooth measure on the manifold up to errors of the order of the coupling constant. Applications to the theory of products of random matrices and a model of a disordered quantum wire are presented.

1. Main results, discussion and applications. This work provides a perturbative calculation of invariant measures for a class of Markov chains on continuous state spaces and shows that these perturbative measures are unique and smooth. Let us state the main result right away in detail, and then place it into context with other work towards the end of this section and explain our motivation to study this problem.

Suppose given a Lie group $\mathcal{G} \subset \mathrm{GL}(L, \mathbb{C})$, a compact, connected, smooth Riemannian manifold $\mathcal{M}$ without boundary and a smooth, transitive group action $\cdot : \mathcal{G} \times \mathcal{M} \to \mathcal{M}$. Thus, $\mathcal{M}$ is a homogeneous space. Furthermore, let $\mathcal{T}_{\lambda,\sigma} \in \mathcal{G}$ be a family of group elements depending on a coupling constant $\lambda \geq 0$ and a parameter σ varying in some probability space $(\Sigma, \mathbf{p})$, which is of the following form:

$$\mathcal{T}_{\lambda,\sigma} = \mathcal{R} \exp\left(\sum_{n=1}^{\infty} \lambda^n \mathcal{P}_{n,\sigma}\right), \tag{1}$$

where $\mathcal{R} \in \mathcal{G}$ and $\mathcal{P}_{n,\sigma}$ are measurable maps on Σ with compact image in the Lie algebra $\mathfrak{g}$ of $\mathcal{G}$ such that

$$\limsup_{n\to\infty} \sup_{\sigma\in\Sigma} (\|\mathcal{P}_{n,\sigma}\|)^{1/n} < \infty \tag{2}$$

for some norm on $\mathfrak{g}$. This implies that $\mathcal{T}_{\lambda,\sigma}$ is well defined and analytic in λ for λ sufficiently small. The expectation value of the first-order term $\mathcal{P}_{1,\sigma}$ will be denoted by $\mathcal{P} = \int \mathbf{p}(d\sigma)\mathcal{P}_{1,\sigma}$.

Received February 2008; revised August 2009.

[1]Supported by the DFG.

AMS 2000 subject classifications. 60J05, 37H05, 37H15.

Key words and phrases. Group action, invariant measure, Birkhoff sum.

Let us consider the product probability space $(\Omega, \mathbf{P}) = (\Sigma^{\mathbb{N}}, \mathbf{p}^{\mathbb{N}})$. Associated to $\omega = (\sigma_n)_{n\in\mathbb{N}} \in \Omega$, there is a sequence $(\mathcal{T}_{\lambda,\sigma_n})_{n\in\mathbb{N}}$ of group elements. An $\mathcal{M}$-valued Markov process $x_n(\lambda, \omega)$ with starting point $x_0 \in \mathcal{M}$ is defined iteratively by

$$x_n(\lambda, \omega) = \mathcal{T}_{\lambda,\sigma_n} \cdot x_{n-1}(\lambda, \omega). \tag{3}$$

The averaged Birkhoff sum of a complex function f on $\mathcal{M}$ is

$$I_{\lambda,N}(f) = \mathbf{E}_\omega \frac{1}{N} \sum_{n=0}^{N-1} f(x_n(\lambda, \omega)) = \frac{1}{N} \sum_{n=0}^{N-1} (T_\lambda^n f)(x_0), \tag{4}$$

where in the second expression we used the Markov transition operator $(T_\lambda f)(x) = \mathbf{E}_\sigma(f(\mathcal{T}_{\lambda,\sigma} \cdot x))$. Here and below, expectation values w.r.t. $\mathbf{P}$ (or $\mathbf{p}$) will be denoted by $\mathbf{E}$ (or $\mathbf{E}_\omega$ and $\mathbf{E}_\sigma$). Next, recall that an invariant measure ν_λ on $\mathcal{M}$ is defined by the property $\int \nu_\lambda(dx) f(x) = \int \nu_\lambda(dx)(T_\lambda f)(x)$. The operator ergodic theorem [16], Theorem 19.2, then states that $I_{\lambda,N}(f)$ converges almost surely (in x_0) w.r.t. any invariant measure ν_λ and for any integrable function f. In the case that $\mathcal{M}$ is a projective space and the action is matrix multiplication, one is in the world of products of random matrices. If then the group generated by $\mathcal{T}_{\lambda,\sigma}$, with σ varying in the support of $\mathbf{p}$, is noncompact and strongly irreducible, Furstenberg, Guivarch and Raugi have proved [2, 9, 11] that there is a unique invariant measure ν_λ which is, moreover, Hölder continuous [2]. To our best knowledge, little seems to be known in more general situations and also concerning the absolute continuity of ν_λ (except if $\mathbf{p}$ is absolutely continuous [18], for some and under supplementary hypothesis [4, 28]).

Let $\mathbf{p}_1$ be the distribution of the random variable $\mathcal{P}_{1,\sigma}$ on the Lie algebra $\mathfrak{g}$, that is, for any measurable $\mathfrak{b} \subset \mathfrak{g}$ one has $\mathbf{p}_1(\mathfrak{b}) = \mathbf{p}(\{\mathcal{P}_{1,\sigma} \in \mathfrak{b}\})$. We are interested in a perturbative calculation of $I_{\lambda,N}(f)$ in λ for smooth functions f with rigorous control on the error terms. This can be achieved if the support of $\mathbf{p}_1$ is large enough in the following sense. First, let us focus on the special case $\mathcal{R} = \mathbf{1}$ and $\mathcal{P} = 0$.

THEOREM 1. *Let $\mathcal{T}_{\lambda,\sigma}$ be of the form* (1) *and assume $\mathcal{R} = \mathbf{1}$, $\mathcal{P} = \mathbf{E}(\mathcal{P}_{1,\sigma}) =$* 0. *Let x_n be the associated Markov process on $\mathcal{M}$ as given by* (3) *and let $\mathfrak{v} =$* $\mathrm{Lie}(\mathrm{supp}(\mathbf{p}_1))$ *be the smallest Lie subalgebra of $\mathfrak{g}$ that contains the support of $\mathbf{p}_1$. Recall that $\mu(dx)$ denotes the Riemannian volume measure on $\mathcal{M}$.*

Coupling hypothesis: *Suppose that the smallest subgroup $\mathcal{V}$ of $\mathcal{G}$ containing* $\{\exp(\lambda\mathcal{P}), \mathcal{P} \in \mathfrak{v}, \lambda \in [0, 1]\}$ *acts transitively on $\mathcal{M}$.* (*This is a Lie subgroup with Lie algebra $\mathfrak{v}$, but it may not be a submanifold.*)

Then there is a sequence of smooth functions ρ_m with $\int_{\mathcal{M}} d\mu \rho_m = \delta_{m,0}$ and $\rho_0 > 0$ μ-almost surely, such that for any $M \in \mathbb{N}$ and any function $f \in C^\infty(\mathcal{M})$, one obtains

$$I_{\lambda,N}(f) = \sum_{m=0}^{M} \lambda^m \int_{\mathcal{M}} \mu(dx) \rho_m(x) f(x) + \mathcal{O}\left(\frac{1}{N\lambda^2}, \lambda^{M+1}\right). \tag{5}$$

Here, the expression $\mathcal{O}(\frac{1}{N\lambda^2}, \lambda^{M+1})$ means that there are two error terms, one of which is bounded by $C_1\frac{1}{N\lambda^2}$ and the other by $C_2\lambda^{M+1}$ with C_1, C_2 depending on f and M. Especially, C_2 may grow in M so that we cannot deduce uniqueness of the invariant measure for small λ this way (cf. Remark 1 below).

When $\mathcal{R} \neq \mathbf{1}$ or $\mathcal{P} \neq 0$ further assumptions are needed in order to control the Birkhoff sums. We assume that $\mathcal{R}$ and $\mathcal{P}$ generate commuting compact groups, that is, $\mathcal{R}\mathcal{P}\mathcal{R}^{-1} = \mathrm{Ad}_{\mathcal{R}}(\mathcal{P}) = \mathcal{P}$ and the closed Abelian Lie groups $\langle\mathcal{R}\rangle = \overline{\{\mathcal{R}^k : k \in \mathbb{Z}\}}$ and $\langle\mathcal{P}\rangle = \overline{\{\exp(\lambda\mathcal{P}) : \lambda \in \mathbb{R})\}}$ are compact. While $\langle\mathcal{P}\rangle$ is always connected, $\langle\mathcal{R}\rangle$ can possibly be disconnected. However, there exists $K \in \mathbb{N}$ such that $\langle\mathcal{R}^K\rangle$ is connected. By considering the suspended Markov process $(y^n)_{n\in\mathbb{N}}$ with $y_n = x_{Kn}$ corresponding to the family

$$\mathcal{T}_{\lambda,\sigma_1,\dots,\sigma_K} = \mathcal{T}_{\lambda,\sigma_K} \cdots \mathcal{T}_{\lambda,\sigma_1}$$

for $(\sigma_1, \dots, \sigma_K) \in (\Sigma^K, \mathbf{p}^K)$, one can always assume that $\langle\mathcal{R}\rangle$ is connected and we shall do so from now on. Note that the product $\langle\mathcal{R}\rangle\langle\mathcal{P}\rangle$ is also a compact, connected, Abelian subgroup of $\mathcal{G}$ which will be denoted by $\langle\mathcal{R}, \mathcal{P}\rangle$. All these groups are tori in $\mathcal{G}$ and their dimensions are $L_{\mathcal{R}}$, $L_{\mathcal{P}}$ and $L_{\mathcal{R},\mathcal{P}}$. Hence, $\langle\mathcal{R}\rangle \cong \mathbb{T}^{L_{\mathcal{R}}}$, $\langle\mathcal{P}\rangle \cong \mathbb{T}^{L_{\mathcal{P}}}$ and $\langle\mathcal{R},\mathcal{P}\rangle \cong \mathbb{T}^{L_{\mathcal{R},\mathcal{P}}}$, where $\mathbb{T}^L = \mathbb{R}^L/(2\pi\mathbb{Z})^L$ is the L-dimensional torus. The (chosen) isomorphisms shall be denoted by $R_{\mathcal{R}}$, $R_{\mathcal{P}}$ and $R_{\mathcal{R},\mathcal{P}}$, respectively, for example, $R_{\mathcal{R}}(\theta) \in \langle\mathcal{R}\rangle \subset \mathrm{GL}(L, \mathbb{C})$ for $\theta = (\theta_1, \dots, \theta_{L_{\mathcal{R}}}) \in \mathbb{T}^{L_{\mathcal{R}}}$.

The isomorphism $R_{\mathcal{R}}$ directly leads to the Fourier decomposition of the function $\theta \in \mathbb{T}^{L_{\mathcal{R}}} \mapsto f(R_{\mathcal{R}}(\theta) \cdot x)$, notably

$$f(R_{\mathcal{R}}(\theta) \cdot x) = \sum_{j \in \mathbb{Z}^{L_{\mathcal{R}}}} f_j(x) e^{\imath j \cdot \theta}, \tag{6}$$

where

$$f_j(x) = \int_{\mathbb{T}^{L_{\mathcal{R}}}} \frac{d\theta}{(2\pi)^{L_{\mathcal{R}}}} e^{-\imath j\cdot\theta} f\big(R_{\mathcal{R}}(\theta) \cdot x\big), \qquad j \cdot \theta = \sum_{l=1}^{L_{\mathcal{R}}} j_l \theta_l.$$

Similarly, the maps $\theta \in \mathbb{T}^{L_{\mathcal{P}}} \mapsto f(R_{\mathcal{P}}(\theta) \cdot x)$ and $\theta \in \mathbb{T}^{L_{\mathcal{R},\mathcal{P}}} \mapsto f(R_{\mathcal{R},\mathcal{P}}(\theta) \cdot x)$ lead to Fourier series.

DEFINITION 1. A function $f \in C^\infty(\mathcal{M})$ is said to consist of only low frequencies w.r.t. $\langle\mathcal{R}\rangle$ if the Fourier coefficients $f_j \in C^\infty(\mathcal{M})$ vanish for j with norm $\|j\| = \sum_{l=1}^{L_{\mathcal{R}}} |j_l|$ larger than some fixed integer $J > 0$. Similarly, f is defined to consist of only low frequencies w.r.t. $\langle\mathcal{P}\rangle$ or $\langle\mathcal{R},\mathcal{P}\rangle$.

The following definitions are standard (see [17] for references).

DEFINITION 2. Let us define $\hat{\theta}_{\mathcal{R}} \in \mathbb{T}^{L_{\mathcal{R}}}$ by $R_{\mathcal{R}}(\hat{\theta}_{\mathcal{R}}) = \mathcal{R}$ and $\hat{\theta}_{\mathcal{P}} \in \mathbb{R}^{L_{\mathcal{P}}}$ by $R_{\mathcal{P}}(\lambda\hat{\theta}_{\mathcal{P}}) = \exp(\lambda\mathcal{P})$. Then $\mathcal{R}$ is said to be a Diophantine rotation or simply Diophantine if there is some $s > 1$ and some constant C such that for any nonzero multi-index $j \in \mathbb{Z}^{L_{\mathcal{R}}} \setminus \{0\}$ one has

$$|e^{\imath j\cdot\hat{\theta}_{\mathcal{R}}} - 1| \geq C\|j\|^{-s}.$$

Similar, $\mathcal{P}$ is said to be Diophantine, or a Diophantine generator of a rotation, if there is some $s > 1$ and some constant C, such that for any nonzero multi-index $j \in \mathbb{Z}^{L_{\mathcal{P}}} \setminus \{0\}$ one has

$$|j \cdot \hat{\theta}_{\mathcal{P}}| \geq C\|j\|^{-s}.$$

As final preparation before stating the result, let us introduce the measure $\overline{\mathbf{p}}$ on the Lie algebra $\mathfrak{g}$ obtained from averaging the distribution $\mathbf{p}_1$ of the lowest-order terms $\mathcal{P}_{1,\sigma}$ w.r.t. the Haar measure dR on the compact group $\langle\mathcal{R},\mathcal{P}\rangle$, namely for any measurable set $\mathfrak{b} \subset \mathfrak{g}$,

$$\overline{\mathbf{p}}(\mathfrak{b}) = \int_{\langle\mathcal{R},\mathcal{P}\rangle} dR\,\mathbf{p}(\{\sigma \in \Sigma : R\mathcal{P}_{1,\sigma}R^{-1} \in \mathfrak{b}\}).$$

THEOREM 2. *Let $\mathcal{T}_{\lambda,\sigma}$ be of the form* (1) *and x_n the associated Markov process on $\mathcal{M}$ as given in* (3). *Denote the Lie algebra of $\langle\mathcal{R},\mathcal{P}\rangle$ by $\mathfrak{r}$ and let $\mathfrak{v} = \mathrm{Lie}(\mathrm{supp}(\overline{\mathbf{p}}), \mathfrak{r})$ be the Lie subalgebra of $\mathfrak{g}$ generated by the support of $\overline{\mathbf{p}}$ and $\mathfrak{r}$. Suppose that the smallest subgroup $\mathcal{V}$ of $\mathcal{G}$ containing $\{\exp(\lambda\mathcal{P}) : \mathcal{P} \in \mathfrak{v}, \lambda \in [0,1]\}$ acts transitively on $\mathcal{M}$. Further, suppose that $f \in C^\infty(\mathcal{M})$ and one of the following conditions hold*:

(i) *$\mathcal{R}$ and $\mathcal{P}$ are Diophantine and $\mathcal{M} = \mathfrak{K}/\mathfrak{H}$ where $\mathfrak{K}$ and $\mathfrak{H} \subset \mathfrak{K}$ are compact Lie groups.*

(ii) *f consist of only low frequencies w.r.t. $\langle\mathcal{R},\mathcal{P}\rangle$.*

Then there is a μ-almost surely positive function $\rho_0 \in C^\infty(\mathcal{M})$ normalized w.r.t. the Riemannian volume measure μ on $\mathcal{M}$, such that

$$I_{\lambda,N}(f) = \int_{\mathcal{M}} \mu(dx)\rho_0(x)f(x) + \mathcal{O}\left(\frac{1}{N\lambda^2}, \lambda\right), \tag{7}$$

where μ is the Riemannian volume measure on $\mathcal{M}$. Moreover, the probability measure $\rho_0\mu$ is invariant under the action of $\langle\mathcal{R},\mathcal{P}\rangle$.

The probability measures $\sum_{m=0}^{M}\lambda^m\rho_m\mu$ in Theorem 1 and $\rho_0\mu$ in Theorem 2 can be interpreted as perturbative approximations of invariant measures ν_λ. In fact, integrating (5) over the initial condition x_0 w.r.t. any invariant measure ν_λ and then taking the limit $N \to \infty$, shows that for any smooth function

$$\int_{\mathcal{M}} \nu_\lambda(dx)f(x) = \sum_{m=0}^{M}\lambda^m\int_{\mathcal{M}} \mu(dx)\rho_m(x)f(x) + \mathcal{O}(\lambda^{M+1}). \tag{8}$$

This means that the invariant measure is unique in a perturbative sense and, moreover, its unique approximations are absolutely continuous with smooth density. In fact, one obtains the following.

COROLLARY 1. *Let the assumptions of Theorems 1 or 2 be fulfilled and* $(\nu_\lambda)_{\lambda>0}$ *be a family of invariant probability measures for the Markov processes* $x_n(\lambda)$. *Then*

$$\text{w}^*\text{-}\lim_{\lambda\to 0} \nu_\lambda = \rho_0\mu,$$

where $\text{w}^*\text{-}\lim$ *denotes convergence in the weak-$*$ topology on the set of Borel measures.*

PROOF. Approximating a continuous function by its Fourier series shows that the set of smooth functions consisting of only low frequencies w.r.t. $\langle \mathcal{R}, \mathcal{P}\rangle$ is dense in the set of continuous functions w.r.t. the $\|\cdot\|_\infty$-norm. The set of probability measures is norm bounded by 1 w.r.t. the dual norm. Now, let $g \in C(\mathcal{M})$. For any $\varepsilon > 0$, there is a smooth function g consisting of only low frequencies such that $\|f - g\|_\infty < \varepsilon$. Then one has

$$\begin{aligned}|\nu_\lambda(f) - \rho_0\mu(f)| &\leq |\nu_\lambda(f-g)| + |\nu_\lambda(g) - \rho_0\mu(g)| + |\rho_0\mu(g-f)|\\ &\leq 2\varepsilon + |\nu_\lambda(g) - \rho_0\mu(g)|.\end{aligned}$$

One obtains $\limsup_{\lambda\to 0}|\nu_\lambda(f) - \rho_0\mu(f)| \leq 2\varepsilon$ for any $\varepsilon > 0$, so that by (8)

$$\limsup_{\lambda\to 0}|\nu_\lambda(f) - \rho_0\mu(f)| = 0,$$

for any continuous function $f \in C(\mathcal{M})$, which gives the desired result. □

REMARK 1. According to the unique weak-$*$-limit for a family of invariant measures ν_λ, one might expect uniqueness for the invariant measure at least in a small interval around 0. However, we will briefly describe a simple example satisfying all conditions of Theorem 1 such that for any rational λ the invariant measure is not unique. Let $\mathcal{G} = \mathcal{M} = \mathbb{S}^1 = \{z \in \mathbb{C} : |z| = 1\}$ and let the Lie group action be the ordinary multiplication. Furthermore, let $\mathcal{R} = 1$ and $\mathcal{P}_{1,\sigma}$ be Bernoulli distributed with probability $\frac{1}{2}$ at $\imath\pi$ and $-\imath\pi$ and let $\mathcal{P}_{n,\sigma} = 0$ for $n \geq 2$. Any measure on $\mathbb{S}^1$ which is invariant under a rotation by $\lambda\pi$ is an invariant measure and for rational λ there are many of them. Therefore, we expect the following to hold: given the conditions of Theorem 2 one finds $\lambda_0 > 0$ such that for Lebesgue a.e. $\lambda \in [0, \lambda_0]$ there is a unique invariant measure.

REMARK 2. The main hypothesis of Theorems 1 and 2 is that the Lie group associated to the Lie algebra $\mathfrak{v}$ acts transitively on $\mathcal{M}$. This can roughly be thought of as a Lie algebra equivalent of Furstenberg's irreducibility condition or the

Goldsheid–Margulis criterion [10]. Let us note that nontrivial $\mathcal{R}$, $\mathcal{P}$ lead to a larger support for $\overline{\mathbf{p}}$ and hence weaken this hypothesis. A second hypothesis is that the group $\langle\mathcal{R}\rangle$ is compact. This excludes many situations appearing in physical models where hyperbolic or parabolic channels appear. In some particular situations, this could be dealt with [24, 25].

REMARK 3. As by the main hypothesis the action of $\mathcal{G}$ on $\mathcal{M}$ is transitive, $\mathcal{M}$ is always a homogeneous space and given as a quotient of $\mathcal{G}$ w.r.t. some isotropy group, but hypothesis (i) requires that $\mathcal{M}$ is, moreover, a quotient of a compact group (which in the examples of Section 5 is a subgroup of $\mathcal{G}$). The assumption that $\mathcal{G} \subset \mathrm{GL}(L, \mathbb{C})$ (or, equivalently $\mathcal{G}$ has a faithful representation) is only needed for the proof of Theorem 2 under hypothesis (ii).

REMARK 4. Suppose $\mathfrak{K}$ is a compact subgroup of $\mathcal{G}$ acting transitively on $\mathcal{M}$ [which is a special case of the condition in Theorem 2(i)]. Then the Haar measure dk on $\mathfrak{K}$ induces a unique natural $\mathfrak{K}$-invariant measure on $\mathcal{M}$ which one may choose to be μ (which is also the volume measure of the metric $\int dK\, K_* g$). It is interesting to examine whether $\rho_0 = 1_{\mathcal{M}}$, that is, the lowest-order approximation of the invariant measure is given by the natural measure. The proof below provides a technique to check this. More precisely, in the notation developed below, $\hat{\mathcal{L}}^* 1_{\mathcal{M}} = 0$ implies that ρ_0 is constant. An example, where this can indeed be checked is developed in Section 5. Note that, if $\mathfrak{K}$ is as above, then any conjugation $\mathcal{N}\mathfrak{K}\mathcal{N}^{-1}$ with an element $\mathcal{N} \in \mathcal{G}$ has another natural measure, given by $J_{\mathcal{N}}\mu$ where $J_{\mathcal{N}}$ is the Jacobian of the map $x \mapsto \mathcal{N} \cdot x$. Unless μ is invariant under all of $\mathcal{G}$, the equality $\rho_0 = 1_{\mathcal{M}}$ is hence linked to a good choice of $\mathfrak{K}$. If μ is invariant under $\mathcal{G}$, then it is also an invariant measure for the Markov process and under the hypothesis of Theorem 2 one therefore has $\rho_0 = 1_{\mathcal{M}}$.

REMARK 5. If $\langle\mathcal{R}, \mathcal{P}\rangle$ acts transitively on $\mathcal{M}$, then the measure $\rho_0\mu$ is uniquely determined by the fact that it is invariant under the action of $\langle\mathcal{R}, \mathcal{P}\rangle$ and normalized. Moreover, $\mathcal{M}$ is isomorphic to the quotient of $\langle\mathcal{R}, \mathcal{P}\rangle$ and the stabilizer $\mathcal{S}_x$ of any point $x \in \mathcal{M}$ (which is a compact Abelian subgroup of $\langle\mathcal{R}, \mathcal{P}\rangle$). Hence, in this case, $\mathcal{M}$ is a torus and the action is simply the translation on the torus. Consequently, the measure $\rho_0\mu$ is the Haar measure. Note that, if $\mathcal{P} = 0$, this holds independently of the perturbation and is imposed by the deterministic process for $\lambda = 0$.

REMARK 6. If the action of $\langle\mathcal{R}\rangle$ on $\mathcal{M}$ is not transitive, there are many invariant measures ν_0 for the deterministic dynamics (in particular, if $\mathcal{R} = \mathbf{1}$ any measure is invariant under $\langle\mathcal{R}\rangle$). Under the hypothesis of Theorems 1 and 2, the random perturbations $\mathcal{P}_{1,\sigma}$ and $\mathcal{P}_{2,\sigma}$ single out a unique perturbative invariant measure $\rho_0\mu$.

REMARK 7. We believe that condition that $\mathcal{R}$ and $\mathcal{P}$ commute is unnecessary. In fact, we expect that conditions on $\mathcal{P}$ can be replaced by conditions on $\hat{\mathcal{P}} = \int_{\langle\mathcal{R}\rangle} dR\, R\mathcal{P}R^{-1}$.

REMARK 8. Let us cite prior work on the rigorous perturbative evaluation of the averaged Birkhoff sums (4). In the case of $\mathcal{G}=\mathrm{SL}(2,\mathbb{R})$, $\mathcal{M}=\mathbb{R}P(1)$ and a rotation matrix $\mathcal{R}$ in (1), Pastur and Figotin [20] showed (7) for the lowest two harmonics whenever $\mathcal{R},\mathcal{R}^2 \neq \pm\mathbf{1}$. The above result combined with the calculations in Section 5 shows that (7) holds also for other functions with $\rho_0 = 1_{\mathcal{M}}$. Without the conditions $\mathcal{R},\mathcal{R}^2 \neq \pm\mathbf{1}$, Theorem 2 was proved in [24, 26]. Moreover, when $\mathcal{R}^K = \mathbf{1}$ (at so-called anomalies) and for an absolutely continuous distribution on $\mathcal{G}$, Theorem 1 was proved by Campanino and Klein [4]. Quasi-one-dimensional generalizations of [20] in the case where $\mathcal{G}$ is a symplectic group were obtained in [25, 27]. The work [7] is an attempt to treat higher-dimensional anomalies. To further generalize, the above results to quasi-one-dimensional systems was our main motivation for this work.

REMARK 9. Our main application presented in Section 5 is the perturbative calculation of Lyapunov exponents associated to products of random matrices of the form (1). Moreover, we show how to choose $\mathcal{N}$ (cf. Remark 5) such that $\rho_0 = 1_{\mathcal{M}}$. This property is called the *random phase property* in [22] which is related to the maximal entropy Ansatz in the physics literature. Section 5 can be read directly at this point if Theorem 2 is accepted without proof.

REMARK 10. The recent work by Dolgopyat and Krikorian [6] on random diffeomorphisms on $\mathbb{S}^d$ contains results on the associated invariant measure and Lyapunov spectrum which are related to the results of the present paper. The main difference is that [6] assume the random diffeomorphisms to be close to a set of rotations which generate $\mathrm{SO}(d+1)$ while in the present work the diffeomorphisms are to lowest order given by the identity (Theorem 1) or close to one fixed rotation (Theorem 2). As a result, the invariant measure in Proposition 2 of [6] is close to the Haar measure while it is determined by the random perturbations in the present paper. In the particular situation of the example studied in Section 5, the randomness is such that the invariant measure is the Haar measure and as a consequence the Lyapunov spectrum is equidistant, just as in [6].

In order to clearly exhibit the strategy of the proof of the theorems, we first focus on the case $\mathcal{R} = \mathbf{1}$ and $\mathcal{P} = 0$ in Sections 2 and 3, which corresponds to a higher-dimensional anomaly in the terminology of our prior work [24, 26]. The main idea is then to expand $T_\lambda f$ into a Taylor expansion in λ. This directly leads to a second-order differential operator $\mathcal{L}$ on $\mathcal{M}$ of the Fokker–Planck type, for which the Birkhoff sums $I_{\lambda,N}(\mathcal{L}f)$ vanish up to order λ. Under the hypothesis of Theorem 1, it can be shown to be a sub-elliptic Hörmander operator on the smooth functions on $\mathcal{M}$ with a one-dimensional cokernel. Then one can deduce that $\mathbb{C} + \mathcal{L}(C^\infty(\mathcal{M})) = C^\infty(\mathcal{M})$ and that the kernel of $\mathcal{L}^*$ is spanned by a smooth positive function ρ_0. These are the main elements of the proof of Theorem 1 for $M = 1$. Then using the properties of the operators $\mathcal{L}$ and $\mathcal{L}^*$ and a further Taylor expansion of $T_\lambda f$ one can prove Theorem 1 by induction. The additional difficulties for other $\mathcal{R}$, $\mathcal{P}$ in Theorem 2 are dealt with in the more technical Section 4. The applications to Lyapunov exponents are presented in Section 5.

2. Fokker–Planck operator and its properties. In this section, we suppose $\mathcal{R} = \mathbf{1}$ and $\mathcal{P} = \mathbf{E}(\mathcal{P}_{1,\sigma}) = 0$ in (1) and introduce in this case the backward Kolmogorov operator $\mathcal{L}$ and its adjoint $\mathcal{L}^*$, called forward Kolmogorov or also Fokker–Planck operator [21]. Their use for the calculation of the averaged Birkhoff sum is exhibited and several properties of these operators are studied. One way to define the operator $\mathcal{L}: C^\infty(\mathcal{M}) \to C^\infty(\mathcal{M})$ is

$$(\mathcal{L} f)(x) = \frac{d^2}{d\lambda^2}\bigg|_{\lambda=0} (T_\lambda f)(x). \tag{9}$$

Let us rewrite this using the smooth vector fields ∂_P associated to any element $P \in \mathfrak{g}$ by

$$\partial_P f(x) = \frac{d}{d\lambda}\bigg|_{\lambda=0} f(e^{\lambda P} \cdot x). \tag{10}$$

Then $\mathcal{L}$ is given by

$$\mathcal{L} = \mathbf{E}_\sigma(\partial^2_{\mathcal{P}_{1,\sigma}} + 2\partial_{\mathcal{P}_{2,\sigma}}). \tag{11}$$

PROPOSITION 1. *For $F \in C^\infty(\mathcal{M})$, one has*

$$I_{\lambda,N}(\mathcal{L}F) = \mathcal{O}\left(\frac{1}{N\lambda^2}, \lambda\right).$$

PROOF. For $P \in \mathfrak{g}$, a Taylor expansion with Lagrange remainder gives

$$F(e^P \cdot x) = F(x) + (\partial_P F)(x) + \tfrac{1}{2}(\partial_P^2 F)(x) + \tfrac{1}{6}(\partial_P^3 F)(e^{\chi P} \cdot x),$$

for some $\chi \in [0, 1]$. Choose $P = \lambda\mathcal{P}_{1,\sigma} + \lambda^2\mathcal{P}_{2,\sigma} + \lambda^3\mathcal{S}_\sigma(\lambda)$, where $\mathcal{S}_\sigma(\lambda) = \sum_{n=3}^\infty \lambda^{n-3}\mathcal{P}_{n,\sigma}$ and use that $\mathcal{P}_{1,\sigma}$ is centered to obtain

$$\begin{aligned}\mathbf{E}_\sigma F(\mathcal{T}_{\lambda,\sigma} \cdot x) &= F(x) + \mathbf{E}_\sigma\big(\lambda^2\big(\tfrac{1}{2}\partial^2_{\mathcal{P}_{1,\sigma}} F(x) + \partial_{\mathcal{P}_{2,\sigma}} F(x)\big)\big) + \mathcal{O}(\lambda^3)\\ &= F(x) + \tfrac{1}{2}\lambda^2\mathcal{L}F(x) + \mathcal{O}(\lambda^3).\end{aligned}$$

The error terms depend on derivatives of F up to order 3 and are uniform in x because $\mathcal{M}$ is compact and $\mathcal{P}_{1,\sigma}, \mathcal{P}_{2,\sigma}$ and $\mathcal{S}_\sigma(\lambda)$ are compactly supported by (2). Due to definition (3), this implies

$$\mathbf{E}_\omega \frac{1}{N}\sum_{n=1}^{N} F(x_n(\lambda, \omega)) = \mathbf{E}_\omega \frac{1}{N}\sum_{n=0}^{N-1} F(x_n(\lambda, \omega)) + \frac{\lambda^2}{2} I_{\lambda,N}(\mathcal{L}F) + \mathcal{O}(\lambda^3).$$

As the appearing sums only differ by a boundary term, resolving for $I_{\lambda,N}(\mathcal{L}F)$ finishes the proof. $\square$

Next, let us bring the operator $\mathcal{L}$ into a normal form. According to Appendix A, one can decompose $\mathcal{P}_{1,\sigma}$ into a finite linear combination of fixed Lie algebra vectors $\mathcal{P}_i \in \mathfrak{g}$, $i \in I$, with uncorrelated real random coefficients, namely

$$\mathcal{P}_{1,\sigma} = \sum_{i=1}^{I} v_{i,\sigma}\mathcal{P}_i, \qquad v_{i,\sigma} \in \mathbb{R}, \qquad \mathbf{E}_\sigma(v_{i,\sigma}) = 0, \qquad \mathbf{E}_\sigma(v_{i,\sigma}v_{i',\sigma}) = \delta_{i,i'}.$$

Then (11) implies that $\mathcal{L}$ is in the so-called Hörmander form

$$\mathcal{L} = \sum_{i=1}^{I} \partial_{\mathcal{P}_i}^2 + 2\partial_{\mathcal{Q}},$$

where $\mathcal{Q} = \mathbf{E}_\sigma(\mathcal{P}_{2,\sigma})$. Using the main assumption of Theorem 1 (i.e., $\mathfrak{v} \subset \mathfrak{u}$), one can show that $\mathcal{L}$ satisfies the strong Hörmander property of rank $r \in \mathbb{N}$ [14, 15, 23].

PROPOSITION 2. *Under the assumptions of Theorem 1, there exists $r \in \mathbb{N}$ such that $\mathcal{L}$ satisfies a strong Hörmander property of rank r, that is, the vector fields $\partial_{\mathcal{P}_i}$ and their r-fold commutators span the whole tangent space at every point of $\mathcal{M}$.*

In order to check this, one needs to calculate the commutators of vector fields ∂_P, ∂_Q for $P, Q \in \mathfrak{g}$. Let X_P, X_Q denote the left-invariant vector fields on $\mathcal{G}$ and furthermore introduce for each $x \in \mathcal{M}$ a function on $\mathcal{G}$ by $f_x(\mathcal{T}) = f(\mathcal{T} \cdot x)$, $\mathcal{T} \in \mathcal{G}$. Then one obtains

$$\begin{aligned}\partial_P \partial_Q f(x) &= \frac{d}{d\lambda}\bigg|_{\lambda=0} (\partial_Q f)(e^{\lambda P} \cdot x) = \frac{d^2}{d\lambda\, d\mu}\bigg|_{\lambda,\mu=0} f(e^{\mu Q} e^{\lambda P} \cdot x) \\ &= X_Q X_P f_x(\mathbf{1}),\end{aligned}$$

which implies

$$\begin{aligned}(\partial_P \partial_Q - \partial_Q \partial_P) f(x) &= (X_Q X_P - X_P X_Q) f_x(\mathbf{1}) = X_{[Q,P]} f_x(\mathbf{1}) \\ &= \partial_{[Q,P]} f(x),\end{aligned} \tag{12}$$

where $[Q, P]$ denotes the Lie bracket (this is well known, see Theorem II.3.4 in [12]). We also need the following lemma for the proof of Proposition 2.

LEMMA 1. *Let $\mathcal{U} \subset \mathcal{G}$ be a Lie subgroup of $\mathcal{G}$ that acts transitively on $\mathcal{M}$ and denote the Lie algebra of $\mathcal{U}$ by $\mathfrak{u}$. Then the vector fields ∂_P, $P \in \mathfrak{u}$, span the whole tangent space at each point of $\mathcal{M}$.*

PROOF. First, let us show that there is a dense set of points in $\mathcal{M}$ for which the vector fields ∂_P, $P \in \mathfrak{u}$, span the whole tangent space. Indeed, for a fixed $x \in \mathcal{M}$ consider the surjective, smooth map $\varphi_x : \mathcal{U} \to \mathcal{M}$, $\varphi_x(U) = U \cdot x$. A point $x' \in \mathcal{M}$ is called regular for φ_x if and only if for any point in the preimage of x' the differential $D\varphi_x$ is surjective. For each point x', the hypothesis implies that there is a $U \in \mathcal{U}$ such that $x' = \varphi_x(U) = U \cdot x$ and the regularity of x' then shows that the paths $\lambda \mapsto \varphi_x(e^{\lambda P} U) = e^{\lambda P} \cdot x'$, $P \in \mathfrak{u}$, span the whole tangent space at x'. By Sard's theorem [13], the set of regular points is dense in $\mathcal{M}$.

Actually, the existence of only 1 regular point x implies that all points are regular. In fact, again any other point is of the form $x' = U \cdot x$. As the map $x \mapsto x' = U \cdot x$ is a diffeomorphism, the push-forward of the paths $\lambda \mapsto \exp(\lambda P) \cdot x$, $P \in \mathfrak{u}$, given by the paths $\lambda \mapsto U \exp(\lambda P) \cdot x = e^{\lambda U P U^{-1}} \cdot x'$, $P \in \mathfrak{u}$, span the tangent space also at x'. $\square$

PROOF OF PROPOSITION 2. Define iteratively the subspaces $\mathfrak{v}_r \subset \mathfrak{g}$ by

$$\mathfrak{v}_1 = \operatorname{span}\{\mathcal{P}_i : 1 \le i \le I\}, \qquad \mathfrak{v}_r = \operatorname{span}(\mathfrak{v}_{r-1} \cup [\mathfrak{v}_{r-1}, \mathfrak{v}_1]). \tag{13}$$

By definition, one has $\mathfrak{v}_1 = \operatorname{span}(\operatorname{supp}(\mathcal{P}_\sigma))$. The space $\mathfrak{v} \subset \mathfrak{g}$ defined in Theorem 1 is equal to $\mathfrak{v} = \operatorname{Lie}(\mathfrak{v}_1)$. Due to (12), the strong Hörmander property of rank r is equivalent to the property that ∂_P, $P \in \mathfrak{v}_r$, spans the whole tangent space at every point $x \in \mathcal{M}$.

By the Lemma 1 and the assumption of Theorem 1, this is fulfilled if $\mathfrak{v}_r = \mathfrak{v}$ for some r. As the vector spaces $\mathfrak{v}_r$ are nested and $\mathfrak{g}$ is finite dimensional, the sequence has to become stationary. This means, there is some r such that $\mathfrak{v}_r = \mathfrak{v}_{r+1}$. Using the Jacobi identity, one then checks that $\mathfrak{v}_r$ is closed under the Lie bracket and therefore $\mathfrak{v}_r = \mathfrak{v}$. $\square$

Next, we want to recollect the consequences of the strong Hörmander property of rank r as proved in [14, 15, 23]. The first basic fact is the subelliptic estimate within any chart

$$\|f\|_{(1/r)} \le C\big(\|\mathcal{L} f\|_{(0)} + \|f\|_{(0)}\big), \tag{14}$$

where $\|\cdot\|_{(s)}$ denotes the Sobolev norms. Using a finite atlas of $\mathcal{M}$, one can define a global Sobolev space $H_s(\mathcal{M})$ with norm also denoted by $\|\cdot\|_{(s)}$. Then the estimate (14) holds also w.r.t. these global norms. Moreover, the norm $\|\cdot\|_{(0)}$ can be seen to be equivalent to the norm in $L^2(\mathcal{M}, \mu)$ where μ is the Riemannian volume measure. As usual, the embedding of $H_{s+\varepsilon}(\mathcal{M})$ in $H_s(\mathcal{M})$ is compact for any $\varepsilon > 0$.

The second basic fact is the hypoellipticity of $\mathcal{L}$. In order to state this property, let us first extend $\mathcal{L}$ in the usual dual way to an operator $\mathcal{L}_{\text{dis}}$ on the space $\mathcal{D}' = (C^\infty(\mathcal{M}))'$ of distributions on $\mathcal{M}$. Then hypoellipticity states that, for any smooth function g, the solution f of $\mathcal{L}_{\text{dis}} f = g$ is itself smooth.

The Fokker–Planck operator $\mathcal{L}^*$ is the adjoint of $\mathcal{L}$ in $L^2(\mathcal{M},\mu)$. Because $\mathcal{M}$ is compact and has no boundary, the domain $\mathcal{D}(\mathcal{L}^*)$ of $\mathcal{L}^*$ contains the smooth functions $C^\infty(\mathcal{M})$. Furthermore, $\mathcal{L}^*$ is again a second-order differential operator with the same principal symbol as $\mathcal{L}$. Therefore, $\mathcal{L}^*$ also satisfies the strong Hörmander condition of rank r. Thus, the subelliptic estimate as well as the hypoellipticity property also holds for $\mathcal{L}^*_{\mathrm{dis}}$. We, moreover, deduce that $\mathcal{L}$ is closable with closure $\overline{\mathcal{L}} = \mathcal{L}^{**} \subset \mathcal{L}_{\mathrm{dis}}$.

The following proposition recollects properties of $\mathcal{L}$ as a densely defined operator on the Hilbert space $L^2(\mathcal{M},\mu)$.

PROPOSITION 3. *There exists $c_0 > 0$ such that for $c > c_0$ the following holds*:

(i) *$\mathcal{L} - c$ is dissipative.*
(ii) *$(\mathcal{L} - c)(C^\infty(\mathcal{M}))$ is dense in $L^2(\mathcal{M},\mu)$.*
(iii) *$\overline{\mathcal{L}} - c$ is maximally dissipative.*
(iv) *$\overline{\mathcal{L}} - c$ is the generator of a contraction semigroup on $L^2(\mathcal{M},\mu)$.*
(v) *The resolvent $(\overline{\mathcal{L}} - c)^{-1}$ exists and is a compact operator on $L^2(\mathcal{M},\mu)$.*

PROOF. (i) Let us rewrite $\mathcal{L}$:

$$\mathcal{L}f = \sum_{i=1}^{I}[\mathrm{div}(\partial_{\mathcal{P}_i}(f)\partial_{\mathcal{P}_i}) - \mathrm{div}(\partial_{\mathcal{P}_i})\partial_{\mathcal{P}_i}(f)] + 2\partial_{\mathcal{Q}}(f).$$

Defining X to be the smooth vector field $2\partial_{\mathcal{Q}} - \sum_i \mathrm{div}(\partial_{\mathcal{P}_i})\partial_{\mathcal{P}_i}$, one has

$$\mathcal{L}f = \sum_{i=1}^{I}\mathrm{div}(\partial_{\mathcal{P}_i}(f)\,\partial_{\mathcal{P}_i}) + X(f).$$

For a real, smooth function f, the divergence theorem and estimate on the negative quadratic term gives

$$\langle f|\mathcal{L}f\rangle = \int_{\mathcal{M}} d\mu\left[-\sum_{i=1}^{I}\partial_{\mathcal{P}_i}(f)\,\partial_{\mathcal{P}_i}(f) + fX(f)\right] \le \int_{\mathcal{M}} d\mu\, fX(f).$$

Using $2fX(f) = X(f^2) = \mathrm{div}(f^2X) - f^2\,\mathrm{div}(X)$ and again the divergence theorem, it follows that

$$\langle f|\mathcal{L}f\rangle \le -\frac{1}{2}\int_{\mathcal{M}} d\mu\, \mathrm{div}(X)f^2 \le \frac{1}{2}\|\mathrm{div}(X)\|_\infty\|f\|_2^2. \tag{15}$$

As $\mathcal{L}$ is real, it follows that $\Re e\langle f|(\mathcal{L}-c)f\rangle \le 0$ for $f \in C^\infty(\mathcal{M})$ and $c > c_0$ where $c_0 = \frac{1}{2}\|\mathrm{div}(X)\|_\infty$. By definition, this means precisely that $\mathcal{L} - c$ is dissipative.

(ii) Let $h \in L^2(\mathcal{M},\mu)$ such that $\langle h|\mathcal{L}f - cf\rangle = 0$ for all $f \in C^\infty(\mathcal{M}) = \mathcal{D}(\mathcal{L})$. Then h is in the kernel of $\mathcal{L}^*_{\mathrm{dis}}$. By hypoellipticity, it follows that $h \in C^\infty(\mathcal{M})$. Therefore, $\langle h|\mathcal{L}h\rangle = c\|h\|_2^2$ contradicting (15) unless $h = 0$.

The statement (iii) means that there is no dissipative extension, which follows directly from (i) and (ii) by [5], Theorems 2.24, 2.25 and 6.4. Item (iv) follows from the same reference.

Concerning (v), the existence of the resolvent follows directly upon integration of the contraction semigroup. Its compactness follows from the subelliptic estimate (14) and the compact embedding of $H_s(\mathcal{M})$ into $L^2(\mathcal{M},\mu)$. $\square$

The next proposition is based on Bony's maximum principle for strong Hörmander operators [1], as well as standard Fredholm theory.

PROPOSITION 4. (i) *The kernel of* $\overline{\mathcal{L}}$ *consists of the constant functions on* $\mathcal{M}$.

(ii) *The kernel of* $\mathcal{L}^*$ *is one dimensional and spanned by a smooth function* ρ_0.

(iii) $\operatorname{Ran}\overline{\mathcal{L}} = (\ker\mathcal{L}^*)^\perp$ *and* $\operatorname{Ran}\mathcal{L}^* = (\ker\overline{\mathcal{L}})^\perp = (\ker\mathcal{L})^\perp$.

(iv) ρ_0 *is* μ*-almost surely positive.*

PROOF. (i) By Corollaire 3.1 of [1], a smooth function f which has a local maximum and for which $\mathcal{L}f = 0$ has to be constant on (the pathwise connected compact set) $\mathcal{M}$. If f lies in the kernel of the closure $\overline{\mathcal{L}} = \mathcal{L}^{**}$, then $\mathcal{L}_{\mathrm{dis}} f = 0$. As $\mathcal{L}$ is hypoelliptic, $f \in C^\infty(\mathcal{M})$ and therefore f is again constant.

(ii) Choose $c > c_0$ as in Proposition 3 and let $K = (\overline{\mathcal{L}} + c)^{-1}$. Then one has

$$\overline{\mathcal{L}} f = g \quad\Leftrightarrow\quad (\overline{\mathcal{L}} + c) f = cf + g \quad\Leftrightarrow\quad f = cKf + Kg$$
$$\Leftrightarrow\quad (\mathbf{1} - cK) f = Kg,$$

and similarly $\mathcal{L}^* f = g \Leftrightarrow (\mathbf{1} - cK^*) f = K^* g$. For $g = 0$, this implies $\ker\mathcal{L} = \ker(\mathbf{1} - cK)$ and $\ker\mathcal{L}^* = \ker(\mathbf{1} - \bar{c}K^*)$. By the Fredholm alternative (the index of $\mathbf{1} + cK$ is 0), the dimension of these two kernels are equal and by (i) hence, both one dimensional. The smoothness of the function in the kernel follows from the hypoellipticity of $\mathcal{L}^*$.

(iii) For $v \in \ker\mathcal{L}^* = \ker(\mathbf{1} - cK^*)$ and $\langle g|v\rangle = 0$, one has $0 = \langle g|v\rangle = \langle g|cK^*v\rangle = c\langle Kg|v\rangle$, therefore $g \in (\ker\mathcal{L}^*)^\perp$ implies $Kg \in \ker(\mathbf{1} - cK)^\perp$ and the Fredholm alternative states that $(\mathbf{1} - cK) f = Kg$ is solvable. Hence by the above, $\overline{\mathcal{L}} f = g$ is solvable. Therefore, $\operatorname{Ran}\overline{\mathcal{L}} = (\ker\mathcal{L}^*)^\perp$. The other equality is proved analogously.

(iv) Let $f \geq 0$ be smooth and suppose that $\int d\mu\, \rho_0 f = 0$. According to (ii), (iii) and hypoellipticity this implies that $f = \mathcal{L}F \geq 0$ for some smooth F. Again by Bony's maximum principle F is constant and therefore $f = 0$. Hence, for any nonvanishing positive function f one has $\int d\mu\, \rho_0 f > 0$. $\square$

Even though not relevant for the sequel, let us also prove the following.

PROPOSITION 5. $\mathcal{L}$ *generates a contraction semigroup in* $(C(\mathcal{M}), \|\cdot\|_\infty)$, *also called a Feller semigroup.*

PROOF. This will follow directly from the Hille–Yosida theorem [16], Theorem 19.11, once we verified that $(\mathcal{L}-c)C^\infty(\mathcal{M})$ is dense in $C(\mathcal{M})$ for some $c>0$ and that $\mathcal{L}$ satisfies the positive-maximum principle. The first property follows from the existence of the resolvent (Proposition 3) and the hypoellipticity. For the second, let a smooth f have a positive local maximum at some $x\in\mathcal{M}$. Then one only has to check $(\mathcal{L}f)(x)\le 0$, which follows because the first derivatives of f vanish, its second derivative is negative and the principal symbol is positive definite. □

One can rewrite (9) as $\lim_{\lambda\to 0}\frac{1}{2\lambda^2}(T_\lambda-\mathbf{1})f=\mathcal{L}f$ in $\|\cdot\|_\infty$ and for $f\in C^\infty(\mathcal{M})$. Hence, the above statement and [16], 19.28, implies directly the following approximation result of the Feller process by the discrete time Markov processes.

COROLLARY 2. *Let $e^{t\mathcal{L}}$ denote the Feller semigroup of Proposition 5. Then with convergence in $(C(\mathcal{M}),\|\cdot\|_\infty)$,*

$$\lim_{\lambda\to 0} T_\lambda^{[t/(2\lambda^2)]}f=e^{t\mathcal{L}}f.$$

Finally, let us note yet another representation of the generator $\mathcal{L}$ following from the two above, namely $\mathcal{L}=\lim_{N\to\infty}\frac{1}{2}N^\beta((T_{N^{-\alpha}})^N-\mathbf{1})$ where $\beta=2\alpha-1>0$ and with strong convergence.

3. Control of Birkhoff sum in the case $\mathcal{R}=1,\mathcal{P}=0$. The aim of this section is the proof of Theorem 1.

PROPOSITION 6. *Let $\mathcal{R}=\mathbf{1}$ and $\mathcal{P}=0$. The kernel of $\mathcal{L}^*_{\mathrm{dis}}$ is spanned by a nonnegative smooth function ρ_0 that is normalized by $\int_{\mathcal{M}}d\mu\,\rho_0=1$. For $f\in C^\infty(\mathcal{M})$,*

$$I_{\lambda,N}(f)=\int_{\mathcal{M}}d\mu\,\rho_0 f+\mathcal{O}\left(\frac{1}{N\lambda^2},\lambda\right).$$

PROOF. By hypoellipticity, the kernel of $\mathcal{L}^*_{\mathrm{dis}}$ coincides with the kernel of $\mathcal{L}^*$. First, we show $C^\infty(\mathcal{M})=\mathbb{C}1_{\mathcal{M}}+\mathcal{L}C^\infty(\mathcal{M})$. Indeed, let $f\in C^\infty(\mathcal{M})$. Set $C=\int_{\mathcal{M}}d\mu\,f\rho_0$ and $\hat f=f-C$. Then one has $\int_{\mathcal{M}}d\mu\,\hat f\rho_0=0$ and therefore $\hat f\in(\ker\mathcal{L}^*)^\perp=\operatorname{Ran}\overline{\mathcal{L}}$ by Proposition 4. By hypoellipticity, $\hat f\in\mathcal{L}(C^\infty(\mathcal{M}))$.

Now using Proposition 1 and the above decomposition

$$I_{\lambda,N}(f)=I_{\lambda,N}(\hat f+C)=C+I_{\lambda,N}(\mathcal{L}F)=C+\mathcal{O}(N^{-1}\lambda^{-2},\lambda),$$

one completes the proof. □

In order to prove Theorem 1, let us define the operators

$$\mathcal{L}^{(M)} f(x) = \frac{d^M}{d\lambda^M}\bigg|_{\lambda=0} (T_\lambda f)(x), \qquad f \in C^\infty(\mathcal{M}).$$

Then $\mathcal{L}^{(1)} = 0$ as $\mathcal{P}_{1,\sigma}$ is centered and $\mathcal{L}^{(2)} = \mathcal{L}$. Using (1), these operators can be written as

$$\mathcal{L}^{(M)} f = \mathbf{E}_\sigma\left(\sum_{m=0}^{M} \sum_{a_1+\cdots+a_m=M} \frac{M!}{m!} \partial_{\mathcal{P}_{a_1,\sigma}} \cdots \partial_{\mathcal{P}_{a_m,\sigma}} f\right).$$

Hence, $\mathcal{L}^{(M)}$ is a differential operator of order M. As $1_\mathcal{M} \in \ker \mathcal{L}^{(m)}$ and hence $\ker \mathcal{L} \subset \ker \mathcal{L}^{(m)}$ for all positive m, one obtains using Proposition 4(iii)

$$\operatorname{Ran} \mathcal{L}^{(m)*} \subset \left(\ker \mathcal{L}^{(m)}\right)^\perp \subset (\ker \mathcal{L})^\perp = \operatorname{Ran} \mathcal{L}^*.$$

Therefore, and as $\ker \mathcal{L}^*$ is one dimensional, the functions ρ_M for $M \in \mathbb{N}$ are iteratively and uniquely defined by

$$\mathcal{L}^* \rho_M = \sum_{m=1}^{M} \frac{2}{(m+2)!} \mathcal{L}^{(m+2)*} \rho_{M-m}, \qquad \int_\mathcal{M} d\mu\, \rho_M = 0, \tag{16}$$

with ρ_0 given by Proposition 6. By induction and hypoellipticity of $\mathcal{L}^*$, it follows that ρ_M is a smooth function for all M, therefore the right-hand side of (16) always exists. Now we can complete the following proof.

PROOF OF THEOREM 1. The proof will be done by induction. The case $M = 0$ is contained in Proposition 6. For the step from $M - 1$ to M, we first need a Taylor expansion of higher order than done so far. As $\mathcal{P}_{1,\sigma}$ is centered and due to the compact support of $\mathcal{P}_{n,\sigma}$ and $\sum_{m \geq n} \lambda^{m-n} \mathcal{P}_{n,\sigma}$ [uniform for small λ by (2)], one obtains with uniform error bound

$$T_\lambda F(x) = F(x) + \frac{1}{2}\lambda^2 \mathcal{L} F(x) + \sum_{m=3}^{M+2} \frac{\lambda^m}{m!} \mathcal{L}^{(m)} F(x) + \mathcal{O}(\lambda^{M+3}),$$

which using the induction hypothesis implies for Birkhoff sums

$$\begin{aligned}
I_{\lambda,N}(\mathcal{L}F) &= \sum_{m=1}^{M} \frac{2\lambda^m}{(m+2)!} I_{\lambda,N}\big(\mathcal{L}^{(m+2)} F\big) + \mathcal{O}\left(\lambda^{M+1}, \frac{1}{\lambda^2 N}\right) \\
&= \sum_{m=1}^{M} \sum_{l=0}^{M-m} \frac{2\lambda^{l+m}}{(m+2)!} \int d\mu\, \rho_l \mathcal{L}^{(m+2)} F + \mathcal{O}\left(\lambda^{M+1}, \frac{1}{\lambda^2 N}\right) \\
&= \sum_{m=1}^{M} \sum_{l=1}^{m} \frac{2\lambda^m}{(l+2)!} \int d\mu\, \big(\mathcal{L}^{(l+2)*} \rho_{m-l}\big) F + \mathcal{O}\left(\lambda^{M+1}, \frac{1}{\lambda^2 N}\right) \\
&= \sum_{m=1}^{M} \lambda^m \int d\mu\, \rho_m (\mathcal{L}F) + \mathcal{O}\left(\lambda^{M+1}, \frac{1}{\lambda^2 N}\right).
\end{aligned}$$

The last step follows from the definition (16) of ρ_m. Now given any smooth function f, we can write it as $f = \int d\mu\, \rho_0 f + \mathcal{L}F$ and obtain

$$I_{\lambda,N}(f) = \int d\mu\, \rho_0 f + \sum_{m=1}^{M} \lambda^m \int d\mu\, \rho_m \mathcal{L}F + \mathcal{O}\left(\lambda^{M+1}, \frac{1}{\lambda^2 N}\right)$$
$$= \sum_{m=0}^{M} \lambda^m \int d\mu\, \rho_m f + \mathcal{O}\left(\lambda^{M+1}, \frac{1}{\lambda^2 N}\right),$$

where the last step follows from $\int d\mu\, \rho_m = 0$ for $m \geq 1$. □

4. Extension to lowest-order rotations. In this section, the lowest-order matrix $\mathcal{R}$ is an arbitrary rotation and $\mathbf{E}(\mathcal{P}_{1,\sigma}) = \mathcal{P}$ commutes with $\mathcal{R}$ and generates a rotation. For any $R \in \mathcal{G}$, let us consider the associated diffeomorphism $x \in \mathcal{M} \mapsto R \cdot x$ and its differential DR. Then the push-forward of functions $f : \mathcal{M} \to \mathbb{C}$ and vector fields $X = (X_x)_{x \in \mathcal{M}}$ are defined by

$$(R_* f)(x) = f(R^{-1} \cdot x), \qquad (R_* X)_{R \cdot x} = DR_x(X_x).$$

The pull-back is then $R^* = (R_*)^{-1}$. With this notation, $R_*(Xf) = (R_*X)(R_*f)$ and

$$R_*(\partial_P(R^* f)) = (R_* \partial_P) f = \partial_{RPR^{-1}} f.$$

Furthermore, we set $R_*(XY) = (R_*X)(R_*Y)$ for the composition of two vector fields X and Y.

Now let $\mathcal{L}$ be defined as in (11) [note that this is not equal to the right-hand side of (9)]. As $\mathcal{R}$ is a zeroth-order term in λ, the Birkhoff sums are to lowest order given by averages along the orbits of $\mathcal{R}$. Furthermore the expectation of the first-order term, $\lambda\mathcal{P}$, then leads to averages over the group $\langle\mathcal{P}\rangle$ to order λ. It is hence reasonable to expect that an averaged Kolmogorov operator has to be considered. In order to define it, recall that there are unique, normalized Haar measures on the compact groups $\langle\mathcal{R}\rangle$, $\langle\mathcal{P}\rangle$ and $\langle\mathcal{R},\mathcal{P}\rangle$. Averages with respect to these measures will be denoted by $\mathbf{E}_{\langle\mathcal{R}\rangle}$, $\mathbf{E}_{\langle\mathcal{P}\rangle}$ and $\mathbf{E}_{\langle\mathcal{R},\mathcal{P}\rangle}$; the integration variable will be R. As the Haar measure is defined by left invariance and the groups $\langle\mathcal{R}\rangle$ and $\langle\mathcal{P}\rangle$ commute by hypothesis, one has $\mathbf{E}_{\langle\mathcal{R},\mathcal{P}\rangle}(g(R)) = \mathbf{E}_{\langle\mathcal{P}\rangle}(\hat{g}(R)$ for $\hat{g}(\tilde{R}) = \mathbf{E}_{\langle\mathcal{R}\rangle}(g(\tilde{R}R))$ and any function g on $\langle\mathcal{R},\mathcal{P}\rangle$. Then set

$$\hat{\mathcal{L}} = \mathbf{E}_{\langle\mathcal{R},\mathcal{P}\rangle}(R_*\mathcal{L}) = \mathbf{E}_{\langle\mathcal{R},\mathcal{P}\rangle}\left(\sum_{i=1}^{I} \partial^2_{R\mathcal{P}_i R^{-1}} + \partial^2_{\mathcal{P}} + 2\partial_{R\mathcal{Q}R^{-1}}\right), \tag{17}$$

where $\mathcal{P}_i$ are obtained by decomposing the centered random variable $\mathcal{P}_{1,\sigma} - \mathcal{P}$ into a sum $\sum_i v_{i,\sigma}\mathcal{P}_i$ such that the real coefficients satisfy $\mathbf{E}(v_{i,\sigma}v_{i',\sigma}) = \delta_{i,i'}$ (cf. Appendix A). With this definition, we are able to prove a result similar to Proposition 1.

PROPOSITION 7. *Let $f \in C^\infty(\mathcal{M})$ and assume one of the following conditions to hold*:

(i) *$\mathcal{R}$ and $\mathcal{P}$ are Diophantine and $\mathcal{M} = \mathfrak{K}/\mathfrak{H}$ for compact Lie groups $\mathfrak{K}$ and $\mathfrak{H} \subset \mathfrak{K}$.*

(ii) *f consists of only low frequencies w.r.t. $\langle \mathcal{R}, \mathcal{P} \rangle$.*

Then one has

$$I_{\lambda,N}(\hat{\mathcal{L}} f) = \mathcal{O}\left(\frac{1}{N\lambda^2}, \lambda\right).$$

For the proof, we first need the following lemma.

LEMMA 2. *Let $f \in C^\infty(\mathcal{M})$, $f_0 = \mathbf{E}_{\langle \mathcal{R} \rangle}(R^* f)$ and $\tilde{f}_0 = \mathbf{E}_{\langle \mathcal{P} \rangle}(R^* f_0) = \mathbf{E}_{\langle \mathcal{R}, \mathcal{P} \rangle}(R^* f)$. If either* (i) *or* (ii) *as in the Proposition 7 holds*, *then*

$$f - f_0 = g - \mathcal{R}^* g, \qquad f_0 - \tilde{f}_0 = \partial_{\mathcal{P}} \mathbf{E}_{\langle \mathcal{R} \rangle}(R^* \tilde{g}), \tag{18}$$

for smooth functions $g, \tilde{g} \in C^\infty(\mathcal{M})$.

PROOF. The group $\langle \mathcal{R} \rangle$ is isomorphic to a torus $\mathbb{T}^{L_\mathcal{R}}$ with isomorphism $R_\mathcal{R}(\theta) \in \langle \mathcal{R} \rangle$. Furthermore we define $\hat{\theta}_\mathcal{R}$ by $\mathcal{R} = R_\mathcal{R}(\hat{\theta}_\mathcal{R})$. If f consists of only low frequencies w.r.t. $\langle \mathcal{R}, \mathcal{P} \rangle$, it can be written as finite sum of its Fourier coefficients

$$f = \sum_{\|j\| < J} f_j \qquad \text{where } f_j\bigl(R_\mathcal{R}(\theta) \cdot x\bigr) = e^{\imath j \cdot \theta} f_j(x),$$

where the Fourier coefficients are calculated as in (6). Now set

$$g = \sum_{0 < \|j\| < J} \frac{f_j}{1 - e^{\imath j \cdot \hat{\theta}_\mathcal{R}}}.$$

This is well defined because $\hat{\theta}_\mathcal{R}$ is irrational as it generates the whole torus. Then $g - \mathcal{R}^* g = \sum_{0 < \|j\| < J} f_j = f - f_0$.

As $\langle \mathcal{P} \rangle$ is an embedded subtorus in $\langle \mathcal{R}, \mathcal{P} \rangle$, f_0 consists of only low frequencies w.r.t. $\langle \mathcal{P} \rangle$. Let $R_\mathcal{P}(\theta)$ denote the isomorphism of $\mathbb{T}^{L_\mathcal{P}}$ with $\langle \mathcal{P} \rangle$ such that $e^{\lambda \mathcal{P}} = R_\mathcal{P}(\lambda \hat{\theta}_\mathcal{P})$. One can decompose $f_0 = \mathbf{E}_{\langle \mathcal{R} \rangle}((R)^* f)$ into a Fourier sum w.r.t. the group $\langle \mathcal{P} \rangle$:

$$f_0 = \sum_{\|j\| < J} \tilde{f}_j \qquad \text{where } \tilde{f}_j\bigl(R_\mathcal{P}(\theta) \cdot x\bigr) = e^{\imath j \cdot \theta} \tilde{f}_j(x).$$

Then

$$\tilde{g} = \sum_{0 < \|j\| < J} \frac{\tilde{f}_j}{\imath j \cdot \hat{\theta}_\mathcal{P}}$$

satisfies $\partial_{\mathcal{P}} \tilde{g} = f_0 - \tilde{f}_0$. Furthermore, $f_0 - \tilde{f}_0$ is invariant under $\mathcal{R}$ which commutes with $\mathcal{P}$, thus

$$f_0 - \tilde{f}_0 = \mathbf{E}_{\langle \mathcal{R} \rangle}(R^* \partial_{\mathcal{P}} \tilde{g}) = \partial_{\mathcal{P}} \mathbf{E}_{\langle \mathcal{R} \rangle}(R^* \tilde{g}).$$

In case (i), g and $\tilde{g}$ will be defined by the same formulas, but with infinite sums. Thus, we have to show that these sums are well defined and that they define smooth functions on $\mathcal{M}$. Let $p : \mathfrak{K} \to \mathcal{M}$ be the projection identifying $\mathcal{M}$ with $\mathfrak{K}/\mathfrak{H}$ and define the smooth class function $F(K, \theta) = f(R_{\mathcal{R}}(\theta) \cdot p(K))$ on the compact Lie group $\mathfrak{K} \times \mathbb{T}^{L_{\mathcal{R}}}$. We want to compare the Fourier series (6) of f w.r.t. $\mathcal{R}$ with the Fourier series of F as given by the Peter–Weyl theorem. By Theorem 5 in Appendix B, this Fourier series of F is given by

$$f\big(R_{\mathcal{R}}(\theta) \cdot p(K)\big) = F(K, \theta) = \sum_{a \in \mathcal{W}_+} \sum_{j \in \mathbb{Z}^{L_{\mathcal{R}}}} d(a) \operatorname{Tr}(\mathcal{F}F(a, j) \pi_a(K)) e^{\imath j \cdot \theta},$$

where $\mathcal{W}_+$ denotes the set of highest weight vectors of $\mathfrak{K}$, $\pi_a : \mathfrak{K} \to U(d(a))$ is the $d(a)$-dimensional, unitary representation of $\mathfrak{K}$ parameterized by a, and $\mathcal{F}F(a, j)$ is a $d(a) \times d(a)$ matrix given by

$$\mathcal{F}F(a, j) = \int_{\mathfrak{K}} dK \int_{\mathbb{T}^{L_{\mathcal{R}}}} d\theta\, F(K, \theta) \pi_a(K^{-1}) e^{-\imath j \cdot \theta}.$$

Here, $d\theta$ and dK denote the normalized Haar measures. Comparing this equation with (6), one obtains that the Fourier coefficients w.r.t. $\langle \mathcal{R} \rangle$ satisfy

$$f_j(p(K)) = \sum_{a \in \mathcal{W}_+} d(a) \operatorname{Tr}(\mathcal{F}F(a, j) \pi_a(K)).$$

Let $g_j(x) = (1 - e^{\imath j \cdot \hat{\theta}_{\mathcal{R}}})^{-1} f_j$ for $\|j\| > 0$. The next aim is to verify that the infinite sum $g = \sum_{\|j\|>0} g_j$ defines a smooth function on $\mathcal{M}$.

As F is smooth, the Fourier coefficients $\mathcal{F}F(a, j)$ are rapidly decreasing by [29] or Theorem 4 in Appendix B, meaning that $\lim_{\|(a,j)\| \to \infty} \|(a, j)\|^h \|\mathcal{F}F(a, j)\| = 0$ for any natural h. Here, one may choose some norm for which $\|(a, j)\| \geq \|j\|$ and $\|\mathcal{F}F(a, j)\|$ denotes the Hilbert–Schmidt norm. As $\mathcal{R}$ is Diophantine, $|e^{\imath j \cdot \hat{\theta}_{\mathcal{R}}} - 1| \geq C\|j\|^{-s} \geq C\|(a, j)\|^{-s}$ for some natural s and the coefficients $\mathcal{F}G(a, j) = (1 - e^{\imath j \cdot \hat{\theta}_{\mathcal{R}}})^{-1} \mathcal{F}F(a, j)$ defined for $\|j\| > 0$ are still rapidly decreasing. Therefore,

$$G(K, \theta) = \sum_{\|j\|>0} \sum_{a \in \mathcal{W}_+} d(a) \operatorname{Tr}(\mathcal{F}G(a, j) \pi_a(K)) e^{\imath j \cdot \theta} = \sum_{\|j\|>0} g_j(p(K)) e^{\imath j \cdot \theta}$$

is a smooth function and the series converges absolutely and uniformly by Theorem 4. Setting $\theta = 0$, this implies that $\sum_{\|j\|>0} g_j$ converges uniformly to a smooth function g on $\mathcal{M}$ satisfying $g - \mathcal{R}^* g = \sum_{\|j\|>0} f_j = f - f_0$.

As before, we write $f_0 = \mathbf{E}_{\langle \mathcal{R} \rangle}(R^* f)$ as sum of Fourier coefficients w.r.t. $\langle \mathcal{P} \rangle$, so $f_0 = \sum_j \tilde{f}_j$, and let $\tilde{g}_j = (\imath j \cdot \hat{\theta}_{\mathcal{P}})^{-1} \tilde{f}_j$ for $\|j\| > 0$. Consider the function

$\tilde{F}(K,\theta) = f_0(R_{\mathcal{P}}(\theta) \cdot p(K))$ on $\mathfrak{K} \times \mathbb{T}^{L_{\mathcal{P}}}$, just as above define the Fourier coefficients $\mathcal{F}\tilde{F}(a, j)$ for $a \in \mathcal{W}_+$, $j \in \mathbb{Z}^{L_{\mathcal{P}}}$ and let $\mathcal{F}\tilde{G}(a, j) = (\imath j \cdot \hat{\theta}_{\mathcal{P}})^{-1}\mathcal{F}\tilde{F}(a, j)$. As $|j \cdot \hat{\theta}_{\mathcal{P}}| \geq C\|j\|^{-s} \geq C\|(a, j)\|^{-s}$ the coefficients $\mathcal{F}\tilde{G}(a, j)$ are rapidly decreasing, the series

$$\tilde{G}(K,\theta) = \sum_{a\in\mathcal{W}_+} \sum_{j\in\mathbb{Z}^{L_{\mathcal{P}}}} d(a)\,\mathrm{Tr}(\mathcal{F}\tilde{G}(a, j)\pi_a(K))e^{\imath j\cdot\theta} = \sum_{\|j\|>0} \tilde{g}_j(p(K))e^{\imath j\cdot\theta}$$

converges absolutely and $\tilde{G}$ is smooth. Thus, $\tilde{g} = \sum_{\|j\|>0} \tilde{g}_j$ exists, is smooth and

$$\partial_{\mathcal{P}}\tilde{g} = \frac{d}{d\lambda}\bigg|_{\lambda=0} \sum_{\|j\|>0} \tilde{g}_j e^{\imath\lambda j\cdot\hat{\theta}_{\mathcal{P}}} = \sum_{\|j\|>0} \tilde{f}_j = f_0 - \tilde{f}_0.$$

As $f_0 - \tilde{f}_0$ is $\mathcal{R}$-invariant one obtains also $\partial_{\mathcal{P}}\mathbf{E}_{\langle\mathcal{R}\rangle}(R^*\tilde{g}) = f_0 - \tilde{f}_0$. □

LEMMA 3. *If either* (i) *or* (ii), *as in Proposition* 7 *holds*, *one has*

$$I_{\lambda,N}(f) = I_{\lambda,N}\big(\mathbf{E}_{\langle\mathcal{R},\mathcal{P}\rangle}(R^*f)\big) + \mathcal{O}\left(\lambda, \frac{1}{\lambda N}\right).$$

PROOF. Similarly as in the proof of Proposition 1, a Taylor expansion gives

$$\mathbf{E}_\sigma F(\mathcal{T}_{\lambda,\sigma} \cdot x) = \mathcal{R}^*F(x) + \lambda\partial_{\mathcal{P}}\mathcal{R}^*F(x) + \frac{\lambda^2}{2}\mathcal{L}\mathcal{R}^*F(x) + \mathcal{O}(\lambda^3),$$

where the error term is uniform in x. For Birkhoff sums, this implies

$$I_{\lambda,N}(F - \mathcal{R}^*F) = \lambda I_{\lambda,N}(\partial_{\mathcal{P}}\mathcal{R}^*F) + \frac{\lambda^2}{2}I_{\lambda,N}(\mathcal{L}\mathcal{R}^*F) + \mathcal{O}\left(\lambda^3, \frac{1}{N}\right). \tag{19}$$

Using this for $F = g$, it therefore follows that $I_{\lambda,N}(f - f_0) = I_{\lambda,N}(g - \mathcal{R}^*g) = \mathcal{O}(\lambda, N^{-1})$. The function $F = \mathbf{E}_{\langle\mathcal{R}\rangle}(R^*\tilde{g})$ is $\mathcal{R}^*$-invariant, so that the left-hand side of (19) vanishes, and it follows that

$$I_{\lambda,N}(f_0 - \tilde{f}_0) = I_{\lambda,N}\big(\partial_{\mathcal{P}}\mathbf{E}_{\langle\mathcal{R}\rangle}(R^*\tilde{g})\big) = \mathcal{O}\left(\lambda, \frac{1}{\lambda N}\right).$$

Combining both estimates completes the proof. □

As an immediate consequence, one obtains the following.

COROLLARY 3. *The derivative $dR_{\mathcal{R},\mathcal{P}}$ of the isomorphism $R_{\mathcal{R},\mathcal{P}}:\mathbb{T}^{L_{\mathcal{R},\mathcal{P}}}$ gives an isomorphism from $\imath\mathbb{R}^{L_{\mathcal{R},\mathcal{P}}}$ to the Lie algebra $\mathfrak{r}$ of $\langle\mathcal{R},\mathcal{P}\rangle$. Let $\mathcal{Q}_1,\dots,\mathcal{Q}_{L_{\mathcal{R},\mathcal{P}}}$ be the images of the standard orthonormal basis. Then one has* $\exp(2\pi \times \mathcal{Q}_i) = 1$ *and the $\mathcal{Q}_i$ span $\mathfrak{r}$. If either* (i) *or* (ii) *as in Proposition* 7 *holds*, *one has*

$$I_{\lambda,N}(\partial_{\mathcal{Q}_i}(f)) = \mathcal{O}\left(\lambda, \frac{1}{\lambda N}\right) \qquad \textit{implying } I_{\lambda,N}\left(\sum_{i=1}^{L_{\mathcal{R},\mathcal{P}}} \partial^2_{\mathcal{Q}_i}(f)\right) = \mathcal{O}\left(\lambda, \frac{1}{\lambda N}\right).$$

PROOF. First, note that $\partial_{\mathcal{Q}_i} f$ consists of only low frequencies w.r.t. $\langle \mathcal{R}, \mathcal{P} \rangle$ whenever f does. By Lemma 3, it is sufficient to prove $\mathbf{E}_{\langle \mathcal{R}, \mathcal{P} \rangle}(R^*(\partial_{\mathcal{Q}_i} f)) = 0$. This can be easily checked to be true as $\int_0^1 dt \exp(2\pi t \mathcal{Q}_i)^*(\partial_{\mathcal{Q}_i} f) = 0$. $\square$

The following lemma is only needed for the proof of Theorem 2 under hypothesis (ii).

LEMMA 4. *For any Lie algebra element $P \in \mathfrak{g}$, smooth function f on $\mathcal{M}$ and any $x \in \mathcal{M}$, the map $\langle \mathcal{R}, \mathcal{P} \rangle \to \mathbb{C}$, $R \mapsto \partial^i_{RPR^{-1}} f(x)$, $i \in \mathbb{N}$, is a trigonometric polynomial on $\langle \mathcal{R}, \mathcal{P} \rangle$ with uniformly bounded coefficients and uniform degree in $x \in \mathcal{M}$ (depending on i though). This implies that the function $\mathcal{L}(\mathbf{E}_{\langle \mathcal{R}, \mathcal{P} \rangle}(R^* f))$ consists of only low frequencies w.r.t. $\langle \mathcal{R}, \mathcal{P} \rangle$.*

PROOF. As stated above, $\langle \mathcal{R}, \mathcal{P} \rangle \subset \mathcal{G} \subset \mathrm{GL}(L, \mathbb{C})$ is isomorphic to $\mathbb{T}^{L_{\mathcal{R},\mathcal{P}}}$ and the isomorphism is denoted by $R_{\mathcal{R},\mathcal{P}}(\theta) \in \langle \mathcal{R}, \mathcal{P} \rangle$. Furthermore, this group lies in some maximal torus of $\mathrm{GL}(L, \mathbb{C})$. As all maximal tori are conjugate to each other, so that by exchanging $\mathcal{G}$ with some conjugate subgroup in $\mathrm{GL}(L, \mathbb{C})$ one may assume $\langle \mathcal{R}, \mathcal{P} \rangle$ to be diagonal, that is, it consists of diagonal matrices $R(\theta) = \mathrm{diag}(e^{\imath \varphi_1(\theta)}, \dots, e^{\imath \varphi_L(\theta)})$. Beneath the $\varphi_1(\theta), \dots, \varphi_L(\theta)$ there are maximally $L_{\mathcal{R},\mathcal{P}}$ rationally independent, and each is a linear combination with integer coefficients of $\theta_1, \dots, \theta_{L_{\mathcal{R},\mathcal{P}}}$. Hence, any trigonometric polynomial in $\varphi(\theta)$ is a trigonometric polynomial in θ (possibly of higher degree), that is a trigonometric polynomial on $\langle \mathcal{R}, \mathcal{P} \rangle$.

On $\mathfrak{g} \subset \mathrm{gl}(L, \mathbb{C})$, consider the usual real scalar product $\Re e\, \mathrm{Tr}(P^* Q) = \Re e \sum_{a,b} \overline{P_{ab}} Q_{ab}$, where P_{ab} denotes the entries of the matrix P. Let $M = \dim_{\mathbb{R}}(\mathfrak{g})$ and $B^1, \dots, B^M \in \mathfrak{g}$ be some orthonormal basis for $\mathfrak{g}$ w.r.t. this scalar product. If $R = \mathrm{diag}(e^{\imath \varphi_1}, \dots, e^{\imath \varphi_L}) \in \langle \mathcal{R}, \mathcal{P} \rangle$ and $P \in \mathfrak{g}$, then one has

$$
\begin{aligned}
RPR^{-1} &= \sum_{m=1}^{M} \sum_{a,b=1}^{L} \Re e(\overline{B^m_{ab}} (RPR^{-1})_{ab}) B^m \\
&= \sum_{m=1}^{M} \sum_{a,b=1}^{L} \Re e\big(\overline{B^m_{ab}} P_{ab} e^{\imath(\varphi_a - \varphi_b)}\big) B^m,
\end{aligned}
$$

and therefore

$$
\partial^i_{RPR^{-1}} f = \left(\sum_{m=1}^{M} \sum_{a,b=1}^{L} \Re e\big(\overline{B^m_{ab}} P_{ab} e^{\imath(\varphi_a - \varphi_b)}\big) \partial_{B^m} \right)^i f
$$

is a trigonometric polynomial in φ. Thus by definition of $\mathcal{L}$, the map $R \mapsto R_*(\mathcal{L}(R^* f)) = (R_* \mathcal{L}) f$ is a trigonometric polynomial on $\langle \mathcal{R}, \mathcal{P} \rangle$, and therefore also $R \mapsto R_*(\mathcal{L} \hat{f})$ for $\hat{f} = \mathbf{E}_{\langle \mathcal{R}, \mathcal{P} \rangle}(R^* f)$. But this means precisely that $\mathcal{L} \hat{f}$ consists of only low frequencies w.r.t. $\langle \mathcal{R}, \mathcal{P} \rangle$. $\square$

PROOF OF PROPOSITION 7. As $\hat{\mathcal{L}} = \mathbf{E}_{\langle\mathcal{R},\mathcal{P}\rangle}(R_*\mathcal{L})$, it follows for $R \in \langle\mathcal{R},\mathcal{P}\rangle$ that $(R_*\hat{\mathcal{L}})f = \hat{\mathcal{L}}f = (R^*\hat{\mathcal{L}})f$. This implies $R^*(\hat{\mathcal{L}}f) = \hat{\mathcal{L}}(R^*f)$ and $\mathbf{E}_{\langle\mathcal{R},\mathcal{P}\rangle}(R^*(\hat{\mathcal{L}}f)) = \hat{\mathcal{L}}(\mathbf{E}_{\langle\mathcal{R},\mathcal{P}\rangle}(R^*f))$. Hence, the Fourier coefficients of $\hat{\mathcal{L}}f$ are given by

$$(\hat{\mathcal{L}}f)_j = \hat{\mathcal{L}}(f_j). \tag{20}$$

Therefore, $\hat{\mathcal{L}}f$ consists of only low frequencies w.r.t. $\langle\mathcal{R},\mathcal{P}\rangle$ whenever f does. Furthermore, one obtains for $\hat{f} = \mathbf{E}_{\langle\mathcal{R},\mathcal{P}\rangle}(R^*f)$ the following equalities:

$$\mathbf{E}_{\langle\mathcal{R},\mathcal{P}\rangle}(R^*(\hat{\mathcal{L}}f)) = \hat{\mathcal{L}}\hat{f} = \mathbf{E}_{\langle\mathcal{R},\mathcal{P}\rangle}(R_*(\mathcal{L}(R^*\hat{f}))) = \mathbf{E}_{\langle\mathcal{R},\mathcal{P}\rangle}(R_*(\mathcal{L}\hat{f})).$$

Now $\mathcal{L}\hat{f}$ consists of only low frequencies by Lemma 4.

Thus, applying Lemma 3 twice [the hypothesis are given either by hypothesis (i) of Proposition 7 or by (ii) and Lemma 4]. One obtains

$$I_{\lambda,N}(\hat{\mathcal{L}}f) = I_{\lambda,N}\big(\mathbf{E}_{\langle\mathcal{R},\mathcal{P}\rangle}(R^*(\hat{\mathcal{L}}f))\big) + \mathcal{O}\left(\lambda, \frac{1}{\lambda N}\right) = I_{\lambda,N}(\mathcal{L}\hat{f}) + \mathcal{O}\left(\lambda, \frac{1}{\lambda N}\right).$$

As $\mathcal{R}^*\hat{f} = \hat{f}$ and $\partial_{\mathcal{P}}\hat{f} = 0$, equation (19) for $F = \hat{f}$ implies

$$I_{\lambda,N}(\mathcal{L}\hat{f}) = \mathcal{O}\left(\frac{1}{\lambda^2 N}, \lambda\right),$$

which combined with the above finishes the proof. $\square$

After these preparations, the proof of Theorem 2 is analogous to the case $\mathcal{R} = \mathbf{1}$.

PROOF OF THEOREM 2. Consider the Markov process on $\mathcal{M}$ induced by the random family

$$\mathcal{T}_{\lambda,\hat{\sigma}} = \exp(\lambda\mathcal{P}_{1,\hat{\sigma}} + \lambda^2\mathcal{P}_{2,\hat{\sigma}}),$$

where $\hat{\sigma} = (\sigma, R, \alpha, \beta, i) \in \hat{\Sigma} = \Sigma \times \langle\mathcal{R},\mathcal{P}\rangle \times \{-1, 1\} \times \{-1, 1\} \times \{1, \dots, L_{\mathcal{R},\mathcal{P}}\}$ and $\mathcal{P}_{1,\hat{\sigma}} = (R\mathcal{P}_{1,\sigma}R^{-1} - \mathcal{P}) + \alpha\mathcal{P} + \beta\mathcal{Q}_i$, $\mathcal{P}_{2,\hat{\sigma}} = R\mathcal{P}_{2,\sigma}R^{-1}$. The $\mathcal{Q}_i$ are defined as in Corollary 3. $\hat{\Sigma}$ is equipped with the probability measure $\mathbf{p} \times dR \times \frac{1}{2}(\delta_{-1} + \delta_1) \times \frac{1}{2}(\delta_{-1} + \delta_1) \times \frac{1}{L_{\mathcal{R},\mathcal{P}}}(\delta_1 + \cdots + \delta_{L_{\mathcal{R},\mathcal{P}}})$ where dR denotes the Haar measure on $\langle\mathcal{R},\mathcal{P}\rangle$. Let us define $\tilde{\mathcal{L}} = \hat{\mathcal{L}} + \sum_{i=1}^{L_{\mathcal{R},\mathcal{P}}} \partial^2_{\mathcal{Q}_i}$. As $\mathbf{E}_{\hat{\sigma}}(\mathcal{P}_{1,\hat{\sigma}}) = 0$,

$$\tilde{\mathcal{L}} = \sum_{i=1}^{L_{\mathcal{R},\mathcal{P}}} \partial^2_{\mathcal{Q}_i} + \mathbf{E}_{\langle\mathcal{R},\mathcal{P}\rangle}\mathbf{E}_\sigma(\partial^2_{R\mathcal{P}_{1,\sigma}R^{-1}-\mathcal{P}} + \partial^2_{\mathcal{P}} + 2\partial_{R\mathcal{P}_{2,\sigma}R^{-1}}) = \mathbf{E}_{\hat{\sigma}}(\partial^2_{\mathcal{P}_{1,\hat{\sigma}}} + 2\partial_{\mathcal{P}_{2,\hat{\sigma}}})$$

and $\mathrm{span}(\mathrm{supp}(\mathcal{P}_{1,\hat{\sigma}})) = \mathrm{span}(\mathrm{supp}(\overline{\mathbf{p}}), \mathfrak{r})$, this new process leads to the operator $\tilde{\mathcal{L}} = \hat{\mathcal{L}} + \sum_{i=1}^{L_{\mathcal{R},\mathcal{P}}} \partial^2_{\mathcal{Q}_i}$ instead of $\mathcal{L}$ and the whole analysis done for $\mathcal{L}$ in the case $\mathcal{R} = \mathbf{1}$, $\mathcal{P} = 0$ is applicable to $\tilde{\mathcal{L}}$ now due to the hypothesis of Theorem 2. In particular, $\tilde{\mathcal{L}}$ and $\tilde{\mathcal{L}}^*$ are hypoelliptic operators, the kernel of $\tilde{\mathcal{L}}$ consists of the constant functions and the kernel of $\tilde{\mathcal{L}}^*$ is one-dimensional and spanned by a normalized, smooth function $\rho_0 \geq 0$. Furthermore, $C^\infty(\mathcal{M}) = \mathbb{C}1_{\mathcal{M}} + \tilde{\mathcal{L}}C^\infty(\mathcal{M})$ and hence for any smooth function f and $C = \int_{\mathcal{M}} d\mu\, \rho_0 f$, there is a smooth function g such that $f = C + \tilde{\mathcal{L}}g$.

Assume f consists of only low frequencies, that is, $f_j = 0$ for $\|j\| > J$. Then by (20) one obtains for frequencies $\|j\| > 0$ that $f_j = (f - C)_j = \tilde{\mathcal{L}}g_j$ and hence $\tilde{\mathcal{L}}g_j = 0$ for $\|j\| \geq J$. Therefore g_j is constant, which means $g_j = 0$ as $\|j\| > J > 0$ and g consists of only low frequencies if f does. Hence, Proposition 7 implies for both cases (i) and (ii) the first statement of Theorem 2:

$$I_{\lambda,N}(f) = C + I_{\lambda,N}(\tilde{\mathcal{L}}g) = C + \mathcal{O}\left(\lambda, \frac{1}{\lambda^2 N}\right).$$

To see that the measure $\rho_0\mu$ is $\langle \mathcal{R}, \mathcal{P}\rangle$-invariant, let again f be any smooth function. As mentioned above, there exists $g \in C^\infty(\mathcal{M})$ and $C \in \mathbb{C}$ such that $\tilde{\mathcal{L}}g = f - C$. For all $R \in \langle \mathcal{R}, \mathcal{P}\rangle$, this implies $\tilde{\mathcal{L}}R^*g = R^*\tilde{\mathcal{L}}g = R^*f - C$ and hence $f - R^*f = \tilde{\mathcal{L}}(g - R^*g) \in (\ker \tilde{\mathcal{L}}^*)^\perp$ which gives

$$\int_{\mathcal{M}} d\mu\, \rho_0(f - R^*f) = 0.$$

This is precisely the stated invariance property of the measure $\rho_0\mu$. □

5. An application to random Jacobi matrices.

5.1. *Randomly coupled wires.* Here, we consider a family H_λ of random Jacobi matrices with matrix entries of the form

$$(H_\lambda \psi)_n = -\psi_{n+1} - \psi_{n-1} + \lambda W_{\sigma_n} \psi_n, \qquad \psi = (\psi_n)_{n\in\mathbb{Z}} \in (\mathbb{C}^L)^{\times\mathbb{Z}},$$

where the $(W_{\sigma_n})_{n\in\mathbb{Z}}$ are independently drawn from an ensemble of Hermitian $L \times L$ matrices, for which all the entries $W_{i,j} \in \mathbb{C}$, $1 \leq i < j \leq L$, and $W_{k,k} \in \mathbb{R}$, $1 \leq k \leq L$, are independent and centered random variables with variances satisfying

$$\mathbf{E}(W_{i,j}^2) = 0, \qquad \mathbf{E}(|W_{i,j}|^2) = 1, \qquad \mathbf{E}(W_{k,k}^2) = 1. \tag{21}$$

This is equivalent to having $\mathbf{E}(W_{i,j} W_{k,l}) = \delta_{i,l}\delta_{j,k}$. This model is relevant for the quantum mechanical description of a disordered wire, consisting of L identical subwires (all described by a one-dimensional discrete Laplacian) which are pairwise coupled by random hopping elements having random magnetic phases. Moreover, within each wire there is a random potential of the Anderson type. This is similar to a model considered by Wegner [30] and Dorokhov [8]. We are interested in the weak coupling limit of small randomness. Next, we show how this model leads to a question which fits the framework of the main theorems of this work.

For a given fixed energy $E \in (-2, 2)$, the associated transfer matrices [2, 20] are

$$\hat{\mathcal{T}}^E_{\lambda,\sigma} = \begin{pmatrix} \lambda W_\sigma - E\mathbf{1} & -\mathbf{1} \\ \mathbf{1} & \mathbf{0} \end{pmatrix}.$$

Let us introduce the symplectic form $\mathcal{J}$, the Lorentz form $\mathcal{G}$ and the Cayley transformation $\mathcal{C}$ by

$$\mathcal{J} = \begin{pmatrix} 0 & -\mathbf{1} \\ \mathbf{1} & 0 \end{pmatrix}, \qquad \mathcal{G} = \begin{pmatrix} \mathbf{1} & 0 \\ 0 & -\mathbf{1} \end{pmatrix}, \qquad \mathcal{C} = \frac{1}{\sqrt{2}} \begin{pmatrix} \mathbf{1} & -\imath\mathbf{1} \\ \mathbf{1} & \imath\mathbf{1} \end{pmatrix}.$$

Then the transfer matrix $\hat{\mathcal{T}}^E_{\lambda,\sigma}$ is in the Hermitian symplectic group, namely it satisfies $\hat{\mathcal{T}}^* \mathcal{J} \hat{\mathcal{T}} = \mathcal{J}$. Hence, its Cayley transform $\mathcal{C}\hat{\mathcal{T}}^E_{\lambda,\sigma}\mathcal{C}^*$ is in the generalized Lorentz group $\mathrm{U}(L, L)$ of signature (L, L) consisting by definition of the complex $2L \times 2L$ matrices $\hat{\mathcal{T}}$ satisfying $\mathcal{T}^*\mathcal{G}\mathcal{T} = \mathcal{G}$. As a first step, let bring the transfer matrix in its normal form (this corresponds to a change of conjugation as in the proof of Lemma 4). Setting $E = -2\cos(k)$ and

$$\mathcal{N} = \frac{1}{\sqrt{\sin(k)}} \begin{pmatrix} \sin(k)\mathbf{1} & 0 \\ -\cos(k)\mathbf{1} & \mathbf{1} \end{pmatrix},$$

where $|E| < 2$, $\sin(k) \neq 0$, it is a matter of computation to verify

$$\mathcal{T}_{\lambda,\sigma} = \mathcal{C}\mathcal{N}\hat{\mathcal{T}}^E_{\lambda,\sigma}\mathcal{N}^{-1}\mathcal{C}^* = \mathcal{R}_k e^{\lambda\mathcal{P}_\sigma} \in \mathrm{U}(L, L),$$

where

$$\mathcal{R}_k = \begin{pmatrix} e^{-\imath k}\mathbf{1} & 0 \\ 0 & e^{\imath k}\mathbf{1} \end{pmatrix}, \qquad \mathcal{P}_\sigma = \frac{\imath}{2\sin(k)} \begin{pmatrix} W_\sigma & W_\sigma \\ -W_\sigma & -W_\sigma \end{pmatrix}. \tag{22}$$

Note that the group generated by $\mathcal{R}_k$ is a subgroup of the group consisting of all $\mathcal{R}_\theta$ for $\theta \in \mathbb{T}$. Furthermore, $\mathbf{E}(W_\sigma) = 0$.

The group $\mathrm{U}(L, L)$ naturally acts on the Grassmanian flag manifold $\mathcal{M}$ of $\mathcal{G}$-isotropic subspaces of $\mathbb{C}^{2L}$ [2]. In order to describe the flag manifold, let us introduce the set of isotropic frames

$$\mathbb{I} = \{\Phi \in \mathrm{Mat}(2L \times L, \mathbb{C}) : \Phi^*\Phi = \mathbf{1};\, , \Phi^*\mathcal{G}\Phi = 0\}.$$

One readily checks that each $\Phi \in \mathbb{I}$ is of the from $\Phi = 2^{-1/2}\binom{U}{V}$ with $U, V \in \mathrm{U}(L)$. Hence, $\mathbb{I} \cong \mathrm{U}(L) \times \mathrm{U}(L)$ and it has a natural measure given by the product of the Haar measures. The column vectors of Φ then generate a flag. Two isotropic frames Φ_1 and Φ_2 span the same flag if and only if there is an upper triangular $L \times L$ matrix S such that $\Phi_1 = \Phi_2 S$. Due to the above, S is also unitary so that it has to be a diagonal unitary. These diagonal unitaries can be identified with the torus $\mathbb{T}^L$ and thus $\mathbb{I}$ is a $\mathbb{T}^L$-cover of the flag manifold, namely $\mathcal{M} = \mathbb{I}/\mathbb{T}^L = \mathrm{U}(L) \times \mathrm{U}(L)/\mathbb{T}^L$. Consequently $\mathcal{M}$ is a symmetric space and it also carries a natural measure μ. The group action of $\mathrm{U}(L, L)$ on $\mathrm{U}(L) \times \mathrm{U}(L)$ is given

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \cdot \begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} AU + BV \\ CU + DV \end{pmatrix} S, \tag{23}$$

where S is an upper triangular matrix such that $(AU + BV)S$ is unitary; then automatically also $(CU + DV)S$ is unitary. This also defines an action on the quotient $\mathcal{M}$ and one readily checks that μ is invariant under the action of the subgroup $\mathrm{U}(L, L) \cap \mathrm{U}(2L)$.

Let us recall how the general framework of the Introduction is applied in the present situation: the Lie group is $\mathcal{G} = \mathrm{U}(L, L)$ acting on the compact flag manifold $\mathcal{M}$ by (23); equation (22) shows that the rotation is $\mathcal{R} = \mathcal{R}_k$ and the random perturbation $\mathcal{P}_{1,\sigma} = \mathcal{P}_\sigma$, while $\mathcal{P}_{n,\sigma} = 0$ for $n \geq 2$. Objects of interest are now the L positive Lyapunov exponents $\gamma_{l,\lambda}(E)$, $l = 1, \dots, L$ [2]. It can be shown that

$$\sum_{l=1}^{p} \gamma_{l,\lambda}(E) = \lim_{N\to\infty} \mathbf{E}\frac{1}{N}\sum_{n=1}^{N} f_{p,\lambda}(x_n) = \lim_{N\to\infty} I_{\lambda,N}(f_{p,\lambda}), \tag{24}$$

where x_n is the Markoff process on the compact manifold $\mathcal{M}$ and $f_{p,\lambda}$ will be defined next. Actually, we may also consider the action on the cover $\mathbb{I}$ and then $f_{p,\lambda}$ is a class function, defined for $\Phi = (\phi_1, \dots, \Phi_L) \in \mathbb{I}$ by

$$\begin{aligned} f_{p,\lambda}(\Phi) &= \mathbf{E}_\sigma \log(\|\mathcal{T}_{\lambda,\sigma}\phi_1 \wedge \cdots \wedge \mathcal{T}_{\lambda,\sigma}\phi_p\|_{\Lambda^p\mathbb{C}^{2L}}) \\ &= \mathbf{E}_\sigma \det_p(\mathbf{1}_{p\times L}\Phi^*\mathcal{T}^*_{\lambda,\sigma}\mathcal{T}_{\lambda,\sigma}\Phi\mathbf{1}_{L\times p}), \end{aligned}$$

for $1 \leq p \leq L$, where $\mathbf{1}_{p\times L} = (\mathbf{1}, 0)$ is a $p \times L$ matrix and $\mathbf{1}_{L\times p} = \mathbf{1}^*_{p\times L}$. Hence, $\gamma_{l,\lambda}(E)$ are all given by a Birkhoff sum. Applying Theorem 2, one obtains the following.

PROPOSITION 8. *As long as $E = 2\cos(k) \neq 0$ and $|E| < 2$, the lowest-order approximation $\rho_0\mu$ of the invariant measure is the Haar measure on $\mathcal{M}$, that is, $\rho_0 = 1$. The pth greatest Lyapunov exponent $\gamma_p(E)$ is then given by*

$$\gamma_p(E) = \lambda^2\frac{1 + 2(L-p)}{8\sin^2(k)} + \mathcal{O}(\lambda^3). \tag{25}$$

For $L = 1$, (25) is proved in [20]. At the band center $E = 0$, the methods below show that the lowest-order invariant measure is not the Haar measure. In the case $L = 1$, the measure was explicitly calculated in [26]. A formula similar to (25) was obtained in [8]. It shows, in particular, that the Lyapunov spectrum is equidistant. Distinctness of the Lyapunov exponents can also be deduced from the Goldscheid–Margulis criterion. The first step of the proof is to expand $f_{p,\lambda}$ w.r.t. λ for any p. To deal with the expectation values, the following identities are useful.

LEMMA 5. *Let $P, Q \in \mathrm{Mat}(L, \mathbb{C})$. Then one has*

$$\mathbf{E}(W_\sigma) = 0, \qquad \mathbf{E}(W_\sigma^2) = L\mathbf{1}, \qquad \mathbf{E}(\mathrm{Tr}(PW_\sigma)\,\mathrm{Tr}(QW_\sigma)) = \mathrm{Tr}(PQ),$$
$$\mathbf{E}(W_\sigma PW_\sigma) = \mathrm{Tr}(P)\mathbf{1}, \qquad \mathbf{E}(W_\sigma Q\overline{W}_\sigma) = Q^t.$$

Using this, some calculatory effort leads to

$$f_{p,\lambda}(\Phi) = \frac{\lambda^2}{8\sin^2(k)} F_p(\Phi) + \mathcal{O}(\lambda^3), \tag{26}$$

where, setting $\Phi = 2^{-1/2}\binom{U}{V}$, the class function F_p is defined by

$$\begin{aligned} F_p(\Phi) = 2Lp + L\,\mathrm{Tr}\big(\mathbf{1}_{L;p}(V^*U + U^*V)\big) + \tfrac{1}{2}[\mathrm{Tr}(\mathbf{1}_{L;p}(U^*V))]^2 \\ + \tfrac{1}{2}[\mathrm{Tr}(\mathbf{1}_{L;p}V^*U)]^2 - p^2, \end{aligned}$$

where $\mathbf{1}_{L;p} = \mathbf{1}_{L\times p}\mathbf{1}_{p\times L}$ is the projection on the first p entries in $\mathbb{C}^L$. Therefore,

$$\sum_{l=1}^{p} \gamma_{l,\lambda} = \frac{\lambda^2}{8\sin^2(k)} \lim_{N\to\infty} I_{\lambda,N}(F_p) + \mathcal{O}(\lambda^3). \tag{27}$$

Note that F is a polynomial of second degree in the entries of (U, V), and hence consists of only low frequencies w.r.t. to $\langle \mathcal{R}_k \rangle$ as $\mathcal{R}_\theta\binom{U}{V} = \binom{e^{-\imath\theta}U}{e^{\imath\theta}V}$. Thus, in order to apply Theorem 2 we just need to check the coupling hypothesis.

5.2. *Verifying the coupling hypothesis for Theorem 2.* First, we introduce a connected, transitively acting subgroup $\mathcal{U} \subset \mathcal{G}$ such that the space $\mathfrak{v}$ as defined in Theorem 2 fulfills $\mathfrak{u} \subset \mathfrak{v}$, where $\mathfrak{u}$ is the Lie-algebra of $\mathcal{U}$. Then $\mathcal{U}$ is also a subgroup of the group $\mathcal{V}$ as defined in Theorem 2 and $\mathcal{V}$ acts transitively as required. Set

$$\mathcal{U} = \{\mathrm{diag}(U, V) : U, V \in \mathrm{U}(L) \text{ and } UV \in \mathrm{SU}(L)\} \subset \mathrm{U}(L, L).$$

Its Lie algebra is given by

$$\mathfrak{u} = \{\mathrm{diag}(u, v) : u, v \in \mathrm{u}(L), \mathrm{Tr}(u + v) = 0\}.$$

Now the action of $\mathcal{U}$ via (23) on $\mathbb{I}$ is not transitive, but it is indeed transitive on the quotient $\mathcal{M} = \mathbb{I}/\mathbb{T}^L$.

PROPOSITION 9. *The Lie algebra* $\mathfrak{u}$ *is contained in the Lie algebra* $\mathfrak{v}$ *generated by the set* $\{\mathcal{R}\mathcal{P}\mathcal{R}^{-1} : \mathcal{R} \in \langle \mathcal{R}_k \rangle, \mathcal{P} \in \mathrm{supp}(\mathcal{P}_\sigma)\}$, *where* $\mathcal{P}_\sigma$ *is given in* (22).

PROOF. We obtain

$$\mathcal{R}_k\mathcal{P}_\sigma\mathcal{R}_k^{-1} = \frac{\imath}{2\sin(k)} \begin{pmatrix} W_\sigma & e^{-2\imath k}W_\sigma \\ -e^{2\imath k}W_\sigma & -W_\sigma \end{pmatrix}.$$

Hence,

$$-2\cos(2k)\mathcal{P}_\sigma + \mathcal{R}_k\mathcal{P}_\sigma\mathcal{R}_k^{-1} + \mathcal{R}_k^{-1}\mathcal{P}_\sigma\mathcal{R}_k = \frac{1-\cos(2k)}{\sin(k)} \begin{pmatrix} \imath W_\sigma & \mathbf{0} \\ \mathbf{0} & -\imath W_\sigma \end{pmatrix}.$$

Therefore, the space $\mathfrak{v}$ contains all matrices $\begin{pmatrix} \imath W & \mathbf{0} \\ \mathbf{0} & -\imath W \end{pmatrix}$ where $W = W^*$. The commutator of such two matrices is $\begin{pmatrix} [\imath V, \imath W] & \mathbf{0} \\ \mathbf{0} & [\imath V, \imath W] \end{pmatrix}$, hence also obtained in $\mathfrak{v}$. As $\mathrm{su}(L)$ is a simple Lie-algebra and $\imath V$ and $\imath W$ are arbitrary elements of $\mathrm{u}(L)$, the commutators $[\imath V, \imath W]$ contain any element of $\mathrm{su}(L)$. Therefore, taking linear combinations of these terms shows that $\mathfrak{u} \subset \mathfrak{v}$. $\square$

Thus, Theorem 2 applies and equation (25) follows readily from (27) once one has shown that $\rho_0 \mu$ is the Haar measure on $\mathcal{M} = \mathrm{U}(L) \times \mathrm{U}(L)/\mathbb{T}^L$ for $E \neq 0$. Furnishing $\mathcal{M}$ with a left invariant metric, the Haar measure is the volume measure so that we have to show $\rho_0 = C 1_{\mathcal{M}}$ with some normalization constant C. This is equivalent to verifying that $\hat{\mathcal{L}}^* 1_{\mathcal{M}} = 0$. Using $\partial_P^* = -\partial_P - \mathrm{div}(\partial_P)$ and the special form $\hat{\mathcal{L}} = \mathbf{E}_{\langle \mathcal{R} \rangle} \mathbf{E}_\sigma (\partial^2_{R \mathcal{P}_\sigma R^{-1}})$ of the Fokker–Planck operator in the present situation, one gets

$$
\begin{aligned}
\hat{\mathcal{L}}^* 1_{\mathcal{M}} &= \mathbf{E}_{\langle \mathcal{R} \rangle} \mathbf{E}_\sigma \big(([\partial_{R \mathcal{P}_\sigma R^{-1}} + \mathrm{div}(\partial_{R \mathcal{P}_\sigma R^{-1}})]^2) 1_{\mathcal{M}} \big) \\
&= \mathbf{E}_{\langle \mathcal{R} \rangle} \mathbf{E}_\sigma \big(\partial_{R \mathcal{P}_\sigma R^{-1}} (\mathrm{div}(\partial_{R \mathcal{P}_\sigma R^{-1}})) + (\mathrm{div}(\partial_{R \mathcal{P}_\sigma R^{-1}}))^2 \big).
\end{aligned}
\tag{28}
$$

In order to calculate this further, one needs a formula for the divergence of a vector field $\partial_{\mathcal{P}}$, which is the object of the next section.

5.3. *Divergence of vector fields.* Let

$$
\mathcal{P} = \begin{pmatrix} A & B \\ B^* & D \end{pmatrix} \in \mathrm{u}(L, L), \qquad A^* = -A, \qquad D^* = -D.
$$

The aim of this section is to calculate the divergence of the vector field $\partial_{\mathcal{P}}$ on $\mathcal{M}$. It can be lifted to a vector field on $\mathbb{I} \cong \mathrm{U}(L) \times \mathrm{U}(L)$. At the point (U, V), $\partial_{\mathcal{P}}$ is given by the path

$$
\begin{aligned}
t \mapsto \big(&U(\mathbf{1} + t[U^* A U + U^* B V + S]), V(\mathbf{1} + t[V^* D U + V^* B^* U + S]) \big) \\
&+ \mathcal{O}(t^2).
\end{aligned}
\tag{29}
$$

The upper triangular matrix S is determined by the fact that it has reals on the diagonal such that $U^* A U + U^* B V + S$ is in the Lie algebra $\mathrm{u}(L)$. This leads to $S + S^* = -U^* B V - V^* B^* U$. In order to calculate $S - S^*$, let us define the following $\mathbb{R}$-linear function on $\mathrm{Mat}(L, \mathbb{C})$,

$$
w(A) = \sum_{j<k} [E_{j,k}(A + A^*)^t E_{j,k} - E_{k,j}(A + A^*)^t E_{k,j}],
\tag{30}
$$

where $E_{j,k}$ is the matrix with a one at position j, k and a zero elsewhere. One obtains $S - S^* = w(-U^* B V) = -w(U^* B V) \in \mathrm{u}(L)$. Hence, the path defining $\partial_{\mathcal{P}}$ at (U, V) as in (29) is given by

$$
\begin{aligned}
\exp(t\mathcal{P}) \cdot (U, V) = \big(&U\big(\mathbf{1} + t\big[U^* A U + \tfrac{1}{2}(U^* B V - V^* B^* U) - \tfrac{1}{2} w(U^* B V)\big]\big), \\
&V\big(\mathbf{1} + t\big[V^* D V + \tfrac{1}{2}(V^* B^* U - U^* B V) - \tfrac{1}{2} w(U^* B V)\big]\big)\big).
\end{aligned}
$$

Hence, we associate to the induced (lifted) vector field the function $P(U, V) = (U^*AU + \frac{1}{2}(U^*BV - V^*BU) - \frac{1}{2}w(U^*BV), U^*DU + \frac{1}{2}(V^*B^*U - U^*BV) - \frac{1}{2}w(U^*BV)])$.

This vector field induces a projected vector field $\partial_{\mathcal{P}}$ on $\mathcal{M}$ and we want to calculate its divergence on $\mathcal{M}$. The natural metric on $\mathrm{u}(L) \times \mathrm{u}(L)$ induced by the Killing form on $\mathrm{u}(2L)$ is given by $\langle (u, v)|(\tilde{u}, \tilde{v})\rangle = \mathrm{Tr}(u^*\tilde{u} + v^*\tilde{v})$. The Lie algebra $\mathfrak{h}$ of $\mathfrak{H}$ consists of the elements $(\imath\Phi, \imath\Phi)$ for diagonal, real matrices Φ. An orthonormal basis (u_i, v_i) for $\mathfrak{h}^\perp$ in $\mathrm{u}(L) \times \mathrm{u}(L)$ is given by the matrices $\frac{1}{\sqrt{2}}(E_{j,k} - E_{k,j}, \mathbf{0})$, $\imath\frac{1}{\sqrt{2}}(E_{j,k} + E_{k,j}, \mathbf{0})$, $\frac{1}{\sqrt{2}}(\mathbf{0}, E_{j,k} - E_{k,j})$, $\imath\frac{1}{\sqrt{2}}(\mathbf{0}, E_{j,k} + E_{k,j})$ and $\imath\frac{1}{\sqrt{2}}(E_{j,j}, -E_{j,j})$ for $1 \le j < k \le L$. The derivative w.r.t. to the left-invariant vector field on $\mathrm{U}(L) \times \mathrm{U}(L)$ defined by (u_i, v_i) will be denoted by $\delta_{(u_i,v_i)}$. According to (32) in Appendix C the divergence $\mathrm{div}(\partial_{\mathcal{P}})$ on $\mathcal{M}$ is given by

$$\begin{aligned}
&\sum_i \delta_{(u_i,v_i)}\langle (u_i^*, v_i^*)|P(U, V)\rangle \\
&\qquad = \sum_i \delta_{(u_i,v_i)}\Big(\mathrm{Tr}(u_i^*U^*AU + v_i^*V^*DV) \\
&\qquad\qquad - \frac{1}{2}\mathrm{Tr}\Big((u_i + v_i)^* w(U^*BV) \\
&\qquad\qquad\qquad + \frac{1}{2}\mathrm{Tr}\big((u_i - v_i)^*(U^*BV - V^*B^*U)\big)\Big)\Big).
\end{aligned}$$

Now as $u_i^* = -u_i$, one obtains

$$\delta_{(u_i,v_i)}\mathrm{Tr}(u_i^*U^*AU) = \mathrm{Tr}\big(u_i^*(u_i^*U^*AU + U^*AUu_i)\big) = \mathrm{Tr}\big(U^*AU(u_i^2 - u_i^2)\big) = 0.$$

Thus, one has $\sum_i \delta_{(u_i,v_i)}\mathrm{Tr}(u_i^*U^*AU) = 0$ and analogously $\sum_i \delta_{(u_i,v_i)}\mathrm{Tr}(v_i^*V^* \times DV) = 0$. Next, consider $\sum_i \delta_{(u_i,v_i)}\mathrm{Tr}((u_i + v_i)w(U^*BV))$. It is easy to check that for $j \neq k$ one has $\sum_i u_iE_{j,k}\bar{v}_i = \sum_i \bar{u}_iE_{j,k}v_i = 0$ and $\sum_i \bar{u}_iE_{j,k}u_i = \sum_i u_iE_{j,k}\bar{u}_i = E_{k,j}$. The same holds with v_i and u_i exchanged. From these equations, the cyclicity of the trace and the definition of w one obtains after some calculatory effort

$$\frac{1}{2}\sum_i \delta_{(u_i,v_i)}\mathrm{Tr}\big((u_i + v_i)w(U^*BV)\big) = \sum_{j<k}\mathrm{Tr}\big((E_{k,k} - E_{j,j})(U^*BV + V^*B^*U)\big).$$

The remaining term in $\mathrm{div}(\partial_{\mathcal{P}})$ is given by

$$\begin{aligned}
&\frac{1}{2}\sum_i \delta_{u_i,v_i}\mathrm{Tr}\big((u_i^* - v_i^*)(U^*BV - V^*B^*U)\big) \\
&\qquad = \frac{1}{2}\sum_i \mathrm{Tr}\big((v_i - u_i)^2(U^*BV + V^*B^*U)\big).
\end{aligned}$$

As $\sum_i (v_i - u_i)^2 = -2L\mathbf{1}$, it follows that

$$\operatorname{div}(\partial_{\mathcal{P}}) = 2\Re e\,\mathrm{Tr}(CU^* BV), \tag{31}$$

where $C = -L\mathbf{1} + \sum_{j<k}(E_{k,k} - E_{j,j}) = \sum_{j=1}^{L}(2j-1-2L)E_{j,j}$. Note that $\operatorname{div}(\partial_{\mathcal{P}})$ is in fact a function on $\mathcal{M}$, that is, it is independent on the choice of the preimage (U, V) because C is a diagonal matrix.

5.4. *Volume measure to lowest order.* For $E \neq 0$, we now want to show $\hat{\mathcal{L}}1_{\mathcal{M}} = 0$ using (28). As the group $\langle \mathcal{R}_k \rangle$ is a closed subgroup of the torus consisting of all $\mathcal{R}_\theta$ for $\theta \in \mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$, the Haar measure of $\langle \mathcal{R}_k \rangle$ can be considered as a probability measure on $\mathbb{T}$. Expectations w.r.t. to this measure with integration variable $\theta \in \mathbb{T}$ will be denoted by $\mathbf{E}_\theta$. Then for any function f on $\langle \mathcal{R}_k \rangle$, one has $\mathbf{E}_{\mathcal{R}}(f(\mathcal{R})) = \mathbf{E}_\theta(f(\mathcal{R}_\theta))$.

LEMMA 6. *Away from the band center $E \neq 0$, one has*

$$\mathbf{E}_\theta(e^{\pm 2\imath\theta}) = 0, \qquad \mathbf{E}_\theta(e^{\pm 4\imath\theta}) = 0.$$

PROOF. If k is an irrational angle, that is, $\frac{k}{2\pi}$ is irrational, then the closed group generated by $\mathcal{R}_k$ is just the set of all $\mathcal{R}_\theta$ and the measure $\mathbf{E}_\theta$ is the Haar measure of the torus $\mathbb{T}$ implying $\mathbf{E}_\theta(e^{\pm 2\imath\theta}) = \mathbf{E}_\theta(e^{\pm 4\imath\theta}) = 0$. If k is a rational angle, then the closed group generated by $\mathcal{R}_k$ is finite and consists of all $\mathcal{R}_\theta$ such that $e^{\imath\theta}$ is a sth root of 1 for some natural s. The Haar measure is just the point measure giving each point the same mass. As $\sin(k) \neq 0$, we get $s > 2$ which gives $\mathbf{E}_\theta(e^{\pm 2\imath\theta}) = 0$. Similarly, as long as $s \neq 4$ one also obtains $\mathbf{E}(e^{\pm 4\imath\theta}) = 0$. If $s = 4$ which means $k = \pi/2$ and $E = 0$, then $\mathbf{E}_\theta(e^{4\imath\theta}) = 1$. $\square$

Define $A_\sigma = B_\sigma = \frac{\imath W_\sigma}{2\sin^2(k)}$ and $D_\sigma = -A_\sigma$. Then

$$\mathcal{R}_\theta \mathcal{P}_\sigma \mathcal{R}_\theta^{-1} = \begin{pmatrix} A_\sigma & e^{-2\imath\theta} B_\sigma \\ e^{2\imath\theta} B_\sigma^* & D_\sigma \end{pmatrix}.$$

From now on, we assume $E \neq 0$. First, consider the term $[\operatorname{div}(\partial_{\mathcal{R}_\theta \mathcal{P}_\sigma \mathcal{R}_\theta^{-1}})]^2$ appearing in (28). By (31), it is equal to

$$e^{-4\imath\theta}\,\mathrm{Tr}(CU^* B_\sigma V)^2 + e^{4\imath\theta}\,\mathrm{Tr}(CV^* B_\sigma^* U)^2 + 2\,\mathrm{Tr}(CU^* B_\sigma V)\,\mathrm{Tr}(CV^* B_\sigma^* U).$$

By Lemmas 6 and 5, one obtains

$$\mathbf{E}_{\mathcal{R}}\mathbf{E}_\sigma(\operatorname{div}(\partial_{\mathcal{R}\mathcal{P}_\sigma\mathcal{R}^{-1}})(U,V))^2 = \frac{1}{2\sin^2 k}\,\mathrm{Tr}(VCU^*UCV^*) = \frac{\mathrm{Tr}(C^2)}{2\sin^2(k)}.$$

Next, we need to calculate the average of $\partial_{\mathcal{R}_\theta \mathcal{P}_\sigma \mathcal{R}_\theta^{-1}} \operatorname{div}(\partial_{\mathcal{R}_\theta \mathcal{P}_\sigma \mathcal{R}_\theta^{-1}})$ which equals

$$\begin{aligned}\Re e\,\mathrm{Tr}\big(&e^{-2\imath\theta} 2CU^*(A_\sigma^* B_\sigma + B_\sigma D_\sigma)V + C(V^* B_\sigma^* B_\sigma V + U^* B_\sigma B_\sigma^* U) \\ &- e^{-4\imath\theta} 2U^* B_\sigma V U^* B_\sigma V - e^{-2\imath\theta} U^* B_\sigma V (C w_{\theta,\sigma}^* + w_{\theta,\sigma} C)\big),\end{aligned}$$

where $w_{\theta,\sigma} = w(e^{-2\imath\theta} U^* B_\sigma V)$ and w is defined as in (30). Averaging over $\langle \mathcal{R}_k \rangle$ and σ one gets by Lemma 6 and Lemma 5 that $\mathbf{E}_{\mathcal{R}} \mathbf{E}_\sigma (\partial_{\mathcal{R}\mathcal{P}_\sigma\mathcal{R}^{-1}} \operatorname{div}(\partial_{\mathcal{R}\mathcal{P}_\sigma\mathcal{R}^{-1}}))$ is equal to

$$\frac{L \operatorname{Tr}(C)}{2\sin^2(k)} - \mathbf{E}_\theta \mathbf{E}_\sigma \Re e\big(e^{-2\imath\theta} \operatorname{Tr}\big(U^* B_\sigma V (C w_{\theta,\sigma}^* + w_{\theta,\sigma} C)\big)\big).$$

The last term with $w_{\theta,\sigma}$ consists of terms of the form $e^{-4\imath\theta} \operatorname{Tr}(U^* B_\sigma V E_{k,j} (U^* \times BV)^t E_{k,j} C)$ and $\operatorname{Tr}(U^* B_\sigma V E_{j,k} U^t \overline{B}_\sigma \overline{V} E_{j,k} C)$. The latter one gives $\frac{1}{4\sin^2(k)} \times \operatorname{Tr}(U^* U E_{k,j} V^t \overline{V} E_{j,k} C) = \frac{1}{4\sin^2(k)} \operatorname{Tr}(E_{k,k} C)$ after averaging over σ. Therefore and by a similar result for the term with $w_{\theta,\sigma}^*$ as well as the definition of C, one obtains

$$\begin{aligned}
\mathbf{E}_\theta \mathbf{E}_\sigma \operatorname{Tr}\big(e^{-2\imath\theta} U^* B_\sigma V (w_{\theta,\sigma} C + C w_{\theta,\sigma}^*)\big) &= \frac{\sum_{j<k} \operatorname{Tr}((E_{k,k} - E_{j,j})C)}{2\sin^2(k)} \\
&= \frac{\operatorname{Tr}((C + L\mathbf{1})C)}{2\sin^2(k)}.
\end{aligned}$$

Putting everything together one has

$$\mathbf{E}_{\mathcal{R}} \mathbf{E}_\sigma (\operatorname{div}(\operatorname{div}(\partial_{\mathcal{R}\mathcal{P}_\sigma\mathcal{R}^{-1}}) \partial_{\mathcal{R}\mathcal{P}_\sigma\mathcal{R}^{-1}})) = \frac{\operatorname{Tr}(C^2) + L \operatorname{Tr}(C) - \operatorname{Tr}((C + L\mathbf{1})C)}{2\sin^2(k)} = 0.$$

Therefore the lowest-order invariant measure $\rho_0 \mu$ on $\mathcal{M}$ is given by the Haar measure.

APPENDIX A: VECTOR-VALUED RANDOM VARIABLES

LEMMA 7. *Let $a = (a_1, \dots, a_n)^t : \Sigma \to \mathbb{R}^n$ be a centered, vector-valued random variable on a probability space $(\Sigma, \mathbf{p})$, and each $a_k \in L^2(\Sigma, \mathbf{p})$. Then there exist a linear decomposition $a = \sum_i v_i b_i$ over finitely many fixed vectors $b_i \in \mathbb{R}^n$ with coefficient v_i which are centered random variables $v_i \in L^2(\Sigma, \mathbf{p})$ that are uncorrelated $\mathbf{E}(v_i v_{i'}) = \mathbf{E}(v_i^2) \delta_{i,i'}$.*

PROOF. One can assume that the random variables a_k as elements on $L^2(\Sigma, \mathbf{p})$ are linearly independent [otherwise one takes a basis for the vector space $\operatorname{span}(\operatorname{supp}(a))$ and rewrites the random variable a as vector using this basis]. Let us introduce $\lambda_{k,j}$ for $k > j$ and write the Ansatz $v_k = a_k + \sum_{i=1}^{k-1} \lambda_{k,i} a_i$. Inverting the matrix form of these equations gives

$$\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \lambda_{2,1} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \lambda_{n,1} & \cdots & \lambda_{n,n-1} & 1 \end{pmatrix}^{-1} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}.$$

Hence, one can write a as a sum $\sum_k v_k b_k$ where the b_k's are the vectors of the inverted matrix. The v_k's are pairwise uncorrelated, if $\mathbf{E}(v_k a_i) = 0$ for all $i < k$, as this implies $\mathbf{E}(v_k v_i) = 0$ for all $i < k$. Now $\mathbf{E}(v_k a_i) = 0$ for $i = 1, \dots, k-1$ is guaranteed if

$$-\begin{pmatrix} \mathbf{E}(a_k a_1) \\ \vdots \\ \mathbf{E}(a_k a_{k-1}) \end{pmatrix} = \begin{pmatrix} \mathbf{E}(a_1 a_1) & \cdots & \mathbf{E}(a_1 a_{k-1}) \\ \vdots & \ddots & \vdots \\ \mathbf{E}(a_{k-1} a_1) & \cdots & \mathbf{E}(a_{k-1} a_{k-1}) \end{pmatrix} \begin{pmatrix} \lambda_{k,1} \\ \vdots \\ \lambda_{k,k-1} \end{pmatrix}.$$

If the appearing matrix is invertible, one can resolve this equation to get $\lambda_{k,i}$ for all $i < k$. So it remains to show that this matrix is invertible which is equivalent to the property that the columns are linearly independent. Now let $\xi_i \in \mathbb{R}$ such that

$$\sum_{i=1}^{k-1} \xi_i \mathbf{E}(a_j a_i) = \mathbf{E}\left(a_j \sum_{i=1}^{k-1} \xi_i a_i \right) = 0$$

for all $j = 1, \dots, k$. The vector $\sum_{1 \le i \le k-1} \xi_i a_i$ is then orthogonal in $L^2(\Sigma, \mu)$ to any vector in the subspace spanned by $a_1, \dots, a_{k-1}$ and it therefore has to be zero. As the random variables a_i are linearly independent, one gets $\xi_i = 0$ for all $i = 1, \dots, k-1$. $\square$

APPENDIX B: FOURIER SERIES ON COMPACT LIE GROUPS

First, let us summarize some facts about the representation theory of compact Lie groups. All this is well known and proofs can be found in the literature, for example, [3], but we need to introduce the notation for the proof of Theorem 5.

Let $\mathfrak{K}$ be a compact Lie group equipped with its normalized Haar measure and let $\mathfrak{T} \subset \mathfrak{K}$ be some maximal torus $\mathfrak{T} \cong \mathbb{T}^r$, where r is called the rank of $\mathfrak{K}$. The continuous irreducible representations of the torus $\mathfrak{T}$ are given by the characters, that is, the homomorphisms into the group $S^1 = \mathrm{U}(1) \subset \mathbb{C}$. Let us denote them by $X^*(\mathfrak{T})$. They form a $\mathbb{Z}$-module isomorphic to the lattice $\mathbb{Z}^r$ and hence $X^*(\mathfrak{T})$ is a lattice in the vector space $\mathcal{V} = \mathbb{R} \otimes_{\mathbb{Z}} X^*(\mathfrak{T})$, the tensor product over the ring $\mathbb{Z}$. This is an abstract description of the fact, that the characters of the torus $\mathbb{T}^r$ are given by the maps $\theta \in \mathbb{T}^r \mapsto e^{\imath j \cdot \theta}$ for a fixed $j \in \mathbb{Z}^r$. In this case, $\mathcal{V} = \mathbb{R}^r$.

Define some $\mathrm{Ad}_{\mathfrak{K}}$-invariant scalar product on the Lie algebra $\mathfrak{k}$ of $\mathfrak{K}$, where $\mathrm{Ad}_{\mathfrak{K}}$ denotes the adjoint representation, and adopt $\mathcal{V}$ with an scalar product $\langle \cdot, \cdot \rangle$ such that the norm of $a \in X^*(\mathfrak{T})$ coincides with the operator norm of the derivative da acting on $\mathfrak{t}$, the Lie algebra of $\mathfrak{T}$.

Let $\mathfrak{p}$ be the orthogonal complement in $\mathfrak{k}$ of $\mathfrak{t}$, the Lie algebra of $\mathfrak{T}$. Then the group $\mathfrak{T}$ acts on the complexification $\mathfrak{p}_{\mathbb{C}} = \mathbb{C} \otimes_{\mathbb{R}} \mathfrak{p}$ by the adjoint representation and linearity. This representation of $\mathfrak{T}$ can be decomposed into irreducible continuous representations, which means $\mathfrak{p}_{\mathbb{C}} = \bigoplus_{a \in \Phi} \mathfrak{p}_a$ where $\mathfrak{p}_a$ is the set of $P \in \mathfrak{p}_{\mathbb{C}}$ such that $\mathrm{Ad}_{\mathrm{T}}(P) = a(T) P$ for all $T \in \mathfrak{T}$. One can show that the spaces $\mathfrak{p}_a$ are one-dimensional complex vector spaces. The appearing characters $a \in \Phi \subset X^*(\mathfrak{T})$ are called roots of $\mathfrak{K}$. If $a \in \Phi$ is a root, then also $-a \in \Phi$. Note that the character $-a$ as a map on $\mathfrak{T}$ is given by $(-a)(T) = (a(T))^{-1}$.

One can divide the vector space $\mathcal{V}$ in an upper half space and a lower half space in such a way that there is no root on the boundary. A root in the upper half space is then called a positive root. The set of vectors $v \in \mathcal{V}$ that satisfy $\langle v, a\rangle \geq 0$ for all positive roots a is a so-called positive Weyl chamber $\mathcal{C}_+$. An element of the lattice $X^*(\mathfrak{T})$ lying in the positive Weyl chamber is called a highest weight. The set of highest weights will be denoted by $\mathcal{W}_+$. There is a one-to-one correspondence between the irreducible representations and the highest weight vectors.

THEOREM 3. *Any irreducible* (*unitary*) *representation of* $\mathfrak{K}$ *induces* (*by restriction*) *a representation of* $\mathfrak{T}$, *which when decomposed into irreducible representations of* $\mathfrak{T}$ *contains exactly one highest weight* $a \in \mathcal{W}_+$. *For any highest weight vector* $a \in \mathcal{W}_+$, *there is exactly one irreducible representation of* $\mathfrak{K}$ *containing* a.

Let $\pi_a : \mathfrak{K} \to \mathrm{U}(d(a))$ for $a \in \mathcal{W}_+$ be the corresponding irreducible unitary representation of dimension $d(a)$. By Schur orthogonality and the Peter–Weyl theorem the matrix coefficients $\pi_a(K)_{k,l}$, where $1 \leq k, l \leq d(a)$, of these representations, considered as functions on $\mathfrak{K}$, form an orthogonal basis for $L^2(\mathfrak{K})$. The L^2 norm of such a matrix coefficient is $d(a)^{-1/2}$. Therefore, the orthogonal projection of f onto the space spanned by the matrix coefficients of the irreducible representation π_a is given by

$$\begin{aligned}\sum_{k,l=1}^{d(a)} \int_{\mathfrak{K}} d\tilde{K} (f(\tilde{K})\overline{\pi_a(\tilde{K})_{k,l}})\pi_a(K)_{k,l} &= \sum_{k,l=1}^{d(a)} \int_{\mathfrak{K}} d\tilde{K} (f(\tilde{K})\pi_a(\tilde{K}^{-1})_{l,k})\pi_a(K)_{k,l} \\ &= d(a)\operatorname{Tr}(\mathcal{F}f(a)\pi_a(K)),\end{aligned}$$

where

$$\mathcal{F}f(a) = \int_{\mathfrak{K}} dK\, f(\tilde{K})\pi_a(K^{-1}).$$

Hence Schur orthogonality and the Peter–Weyl theorem imply the following.

COROLLARY 4. *Let* $f \in L^2(\mathfrak{K})$, *then one obtains with convergence in* $L^2(\mathfrak{K})$

$$f(K) = \sum_{a \in \mathcal{W}_+} d(a)\operatorname{Tr}(\mathcal{F}f(a)\pi_a(K)).$$

As shown in [29], one can characterize the smooth functions on $\mathfrak{K}$ by their Fourier series.

THEOREM 4. *A function* f *on* $\mathfrak{K}$ *is smooth if and only if its Fourier coefficients are rapidly decreasing*, *which means that*

$$\forall h > 0 : \lim_{\|a\| \to \infty} \|a\|^h \|\mathcal{F}f(a)\| = 0.$$

Here $\|\mathcal{F}f(a)\|$ denotes the Hilbert–Schmidt norm. If this is fulfilled, then the Fourier series converges absolutely in the supremum norm on $\mathfrak{K}$.

Note that the definition of $\mathcal{F}f(a)$ to be rapidly decreasing is independent of the chosen norm on $\mathcal{W}_+ \subset \mathcal{V}$.

Now let us consider the compact group $\mathfrak{K} \times \mathbb{T}^L$ with the maximal torus $\mathfrak{T} \times \mathbb{T}^L$ and its Lie algebra $\mathfrak{t} \times \mathbb{R}^L$. The characters of this torus also factorize by $X^*(\mathfrak{T} \times \mathbb{T}^L) = X^*(\mathfrak{T}) \times \mathbb{Z}^L$. As $\{\mathbf{1}\} \times \mathbb{T}^L$ lies in the center, the direct product of the scalar product on $\mathfrak{k}$ and the canonical scalar product on $\mathbb{R}^L$ give a scalar product on $\mathfrak{k} \times \mathbb{R}^L$ that is invariant under the adjoint representation of the group $\mathfrak{K} \times \mathbb{T}^L$. Therefore, the induced scalar product on the vector space $\mathcal{V} \times \mathbb{R}^L$ spanned by the characters also factorizes.

As the adjoint representation of $\{\mathbf{1}\} \times \mathbb{T}^L$ is trivial, the roots of $\mathfrak{K} \times \mathbb{T}^L$ consist of elements $(a, 0)$ where a is a root of $\mathfrak{K}$. Therefore, the positive roots of $\mathfrak{K} \times \mathbb{T}^L$ are simply the positive roots of $\mathfrak{K}$ and, as the scalar product on $\mathcal{V} \times \mathbb{R}^L$ factorizes, the positive Weyl chamber for $\mathfrak{K} \times \mathbb{T}^L$ is given by $\mathcal{C}_+ \times \mathbb{R}^L$. Hence, the highest weight vectors are given by $\mathcal{W}_+ \times \mathbb{Z}^L$.

Now for $a \in \mathcal{W}_+$ the mapping $(K, \theta) \mapsto \pi_a(K)e^{\imath j \cdot \theta}$ is an irreducible representation of $\mathfrak{K} \times \mathbb{T}^L$ which contains the highest weight vector (a, j) and by Theorem 3, it is the unique one containing this weight. Thus, we have shown the following.

THEOREM 5. *The highest weight vectors of $\mathfrak{K} \times \mathbb{T}^L$ are given by $\mathcal{W}_+ \times \mathbb{Z}^L$, where $\mathcal{W}_+$ are the highest weight vectors of $\mathfrak{K}$. The irreducible representation parameterized by $(a, j) \in \mathcal{W}_+ \times \mathbb{Z}^L$ is given by*

$$\pi_{(a,j)}(K, \theta) = \pi_a(K)e^{\imath j \cdot \theta}.$$

Hence, the Fourier series of F is given by

$$F(K, \theta) = \sum_{a \in \mathcal{W}_+} \sum_{j \in \mathbb{Z}^L} d(a) \operatorname{Tr}(\mathcal{F}F(a, j)\pi_a(K))e^{\imath j \cdot \theta}$$

with convergence in $L^2(\mathfrak{K} \times \mathbb{T}^L)$, where

$$\mathcal{F}F(a, j) = \int_{\mathfrak{K}} dK \int_{\mathbb{T}^L} d\theta\, F(K, \theta)\pi_a(K^{-1})e^{-\imath j \cdot \theta}.$$

APPENDIX C: DIVERGENCE OF VECTOR FIELDS

Let $\mathfrak{H} \subset \mathfrak{K}$ be some compact subgroups of the unitary group $\mathrm{U}(L)$ and let $\mathcal{M} = \mathfrak{K}/\mathfrak{H}$ be the homogeneous quotient and $\pi : \mathfrak{K} \to \mathcal{M}$. On the Lie algebra $\mathrm{u}(L)$ and hence on the Lie algebra $\mathfrak{k}$ of $\mathfrak{K}$, the Killing form $(u, v) = \operatorname{Tr}(u^* v)$ defines a bi-invariant metric. At each point $K \in \mathfrak{K}$, the Lie algebra $\mathfrak{h}$ of $\mathfrak{H}$ form the vertical vectors, that is, the kernel of the differential of π. Hence, the tangent space at $\pi(K)$ can be identified with the horizontal vectors, $\mathfrak{h}^\perp$, the orthogonal complement of $\mathfrak{h}$ in $\mathfrak{k}$. This identification depends on the choice of K. Two horizontal lifts of some tangent vector on $\mathcal{M}$ to two different preimages differ by a conjugation and therefore have the same length due to the invariance of the metric. Thus, there is a unique metric on $\mathcal{M}$ such that the projection $\pi : \mathfrak{K} \to \mathcal{M}$ is a Riemannian submersion. This metric is invariant under the action of $\mathfrak{K}$.

Let S_i be some orthonormal basis for $\mathfrak{h}^\perp$, then the push forward, $\pi_*(S_i)$ forms an orthonormal basis at $\pi(K)$. (This basis vectors may differ for two different preimages.) Let X be some smooth vector field on $\mathcal{M}$ and denote the horizontal lift to $\mathfrak{K}$ by $\hat{X}$ which then is also smooth. As π is a Riemannian submersion, the covariant derivative of X with respect to $\pi_*(S_i)$ is given by $\pi_*(\nabla_{S_i}\hat{X})$. Let (B_j) denote some orthonormal basis of $\mathfrak{k}$ and identify B_j with the left invariant vector field. Furthermore, we identify any vector field Y with a function $Y : \mathfrak{K} \to \mathfrak{k}$ such that the vector at K is given by the path $K\exp(tY(K))$. With $\nabla_S\hat{X}$, we denote the covariant derivative of the vector field $\hat{X}$ and with $\delta_S\hat{X}$ the derivative of the function w.r.t. to the left-invariant vector field S. Then one has

$$\nabla_S\hat{X} = \sum_j \nabla_S \operatorname{Tr}(B_j^*\hat{X})B_j = \sum_j \left[\operatorname{Tr}(B_j^*\hat{X})\frac{1}{2}[S, B_j] + \delta_S\hat{X}\right].$$

If g denotes the metric on $\mathcal{M}$, then the divergence of X at $\pi(K)$ is given by

$$\operatorname{div}(X)\circ\pi = \sum_i g(\pi_*(S_i), \nabla_{\pi_* S_i} X)\circ\pi = \sum_i \operatorname{Tr}(S_i^*\nabla_{S_i}\hat{X}),$$

where we used that S_i is horizontal so that $g(\pi_*(S_i), \pi_*(Y)) = \operatorname{Tr}(S_i^*Y)$ for all Y. Using the identity above and the fact that $S_i^* = -S_i$ which implies $\operatorname{Tr}(S_i^*[S_i, B_j]) = 0$, the expression reduces to

$$\operatorname{div}(X)\circ\pi = \sum_i \delta_{S_i}\operatorname{Tr}(S_i^*\hat{X}). \tag{32}$$

As $\operatorname{Tr}(S_i^*Y) = 0$ for any vertical vector $Y \in \mathfrak{h}$, the lifted vector field $\hat{X}$ does not need to be horizontal for the last equation to hold.

Acknowledgments. We thank A. Bendikov for many helpful discussions and hints to the literature at an early stage of this work, and the referee for a number of helpful suggestions.

REFERENCES

[1] BONY, J.-M. (1969). Principe du maximum, inégalite de Harnack et unicité du problème de Cauchy pour les opérateurs elliptiques dégénérés. *Ann. Inst. Fourier* (*Grenoble*) **19** 277–304. MR0262881

[2] BOUGEROL, P. and LACROIX, J. (1985). *Products of Random Matrices with Applications to Schrödinger Operators*. *Progress in Probability and Statistics* **8**. Birkhäuser, Boston, MA. MR886674

[3] BUMP, D. (2004). *Lie Groups*. *Graduate Texts in Mathematics* **225**. Springer, New York. MR2062813

[4] CAMPANINO, M. and KLEIN, A. (1990). Anomalies in the one-dimensional Anderson model at weak disorder. *Comm. Math. Phys.* **130** 441–456. MR1060385

[5] DAVIES, E. B. (1980). *One-Parameter Semigroups*. *London Mathematical Society Monographs* **15**. Academic Press, London. MR591851

[6] DOLGOPYAT, D. and KRIKORIAN, R. (2007). On simultaneous linearization of diffeomorphisms of the sphere. *Duke Math. J.* **136** 475–505. MR2309172

[7] DORLAS, T. C. and PULÉ, J. V. (2004). The invariant measures at weak disorder for the two-line Anderson model. *Rev. Math. Phys.* **16** 639–673. MR2079121

[8] DOROKHOV, O. N. (1988). Solvable model of multichannel localization. *Phys. Rev. B* **37** 10526–10541.

[9] FURSTENBERG, H. (1963). Noncommuting random products. *Trans. Amer. Math. Soc.* **108** 377–428. MR0163345

[10] GOL'DSHEĬD, I. Y. and MARGULIS, G. A. (1989). Lyapunov exponents of a product of random matrices. *Uspekhi Mat. Nauk* **44** 13–60. MR1040268

[11] GUIVARCH, Y. and RAUGI, A. (1985). Frontière de Furstenberg, propriétés de contraction et théorèmes de convergence. *Z. Wahrsch. Verw. Gebiete* **69** 187–242. MR779457

[12] HELGASON, S. (1978). *Differential Geometry*, *Lie Groups*, *and Symmetric Spaces*. *Pure and Applied Mathematics* **80**. Academic Press, New York. MR514561

[13] HIRSCH, M. W. (1976). *Differential Topology*. *Graduate Texts in Mathematics* **33**. Springer, New York. MR0448362

[14] HÖRMANDER, L. (1967). Hypoelliptic second order differential equations. *Acta Math.* **119** 147–171. MR0222474

[15] JERISON, D. and SÁNCHEZ-CALLE, A. (1987). Subelliptic, second order differential operators. In *Complex Analysis*, *III* (*College Park*, *Md.*, 1985–86). *Lecture Notes in Math.* **1277** 46–77. Springer, Berlin. MR922334

[16] KALLENBERG, O. (2002). *Foundations of Modern Probability*, 2nd ed. Springer, New York. MR1876169

[17] KHANIN, K., LOPES DIAS, J. and MARKLOF, J. (2007). Multidimensional continued fractions, dynamical renormalization and KAM theory. *Comm. Math. Phys.* **270** 197–231. MR2276445

[18] LIAO, M. (2004). *Lévy Processes in Lie Groups*. *Cambridge Tracts in Mathematics* **162**. Cambridge Univ. Press, Cambridge. MR2060091

[19] MARKLOF, J., TOURIGNY, Y. and WOŁOWSKI, L. (2008). Explicit invariant measures for products of random matrices. *Trans. Amer. Math. Soc.* **360** 3391–3427. MR2386231

[20] PASTUR, L. and FIGOTIN, A. (1992). *Spectra of Random and Almost-Periodic Operators*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **297**. Springer, Berlin. MR1223779

[21] RISKEN, H. (1984). *The Fokker–Planck Equation*: *Methods of Solution and Applications*. *Springer Series in Synergetics* **18**. Springer, Berlin. MR749386

[22] RÖMER, R. and SCHULZ-BALDES, H. (2010). The random phase property and the Lyapunov spectrum for disordered multi-channel systems. *J. Stat. Phys.* **140** 122–153.

[23] ROTHSCHILD, L. P. and STEIN, E. M. (1976). Hypoelliptic differential operators and nilpotent groups. *Acta Math.* **137** 247–320. MR0436223

[24] SADEL, C. and SCHULZ-BALDES, H. (2007). Scaling diagram for the localization length at a band edge. *Ann. H. Poincaré* **8** 1595–1621. MR2374952

[25] SCHULZ-BALDES, H. (2004). Perturbation theory for Lyapunov exponents of an Anderson model on a strip. *Geom. Funct. Anal.* **14** 1089–1117. MR2105954

[26] SCHULZ-BALDES, H. (2007). Lyapunov exponents at anomalies of $\mathrm{SL}(2,\mathbb{R})$-actions. In *Operator Theory*, *Analysis and Mathematical Physics*. *Operator Theory*: *Advances and Applications* **174** 159–172. Birkhäuser, Basel. MR2330832

[27] SCHULZ-BALDES, H. (2007). Rotation numbers for Jacobi matrices with matrix entries. *Math*. *Phys*. *Electron*. *J*. **13** 40. MR2366118

[28] SIMON, B. and TAYLOR, M. (1985). Harmonic analysis on $\mathrm{SL}(2,\mathbb{R})$ and smoothness of the density of states in the one-dimensional Anderson model. *Comm*. *Math*. *Phys*. **101** 1–19. MR814540

[29] SUGIURA, M. (1971). Fourier series of smooth functions on compact Lie groups. *Osaka J*. *Math*. **8** 33–47. MR0294571

[30] WEGNER, F. (1979). Disordered system with n orbitals per site: $n=\infty$ limit. *Phys*. *Rev*. *B* **19** 783–792.

DEPARTMENT MATHEMATIK
UNIVERSITÄT ERLANGEN–NÜRNBERG
BISMAKRSTRASSE 1 1/2
ERLANGEN
GERMANY
E-MAIL: schuba@mi.uni-erlangen.de