The Annals of Probability
2021, Vol. 49, No. 3, 1450–1479
https://doi.org/10.1214/20-AOP1488

ADDITIVE FUNCTIONALS AS ROUGH PATHS

BY JEAN-DOMINIQUE DEUSCHEL[1], TAL ORENSHTEIN[2] AND NICOLAS PERKOWSKI[3]

[1]*Institut für Mathematik, TU Berlin,* deuschel@math.tu-berlin.de
[2]*Institut für Mathematik, TU Berlin and Weierstrass Institute Berlin,* orenshtein@wias-berlin.de
[3]*Institut für Mathematik, FU Berlin,* perkowski@math.fu-berlin.de

We consider additive functionals of stationary Markov processes and show that under Kipnis–Varadhan type conditions they converge in rough path topology to a Stratonovich Brownian motion, with a correction to the Lévy area that can be described in terms of the asymmetry (nonreversibility) of the underlying Markov process. We apply this abstract result to three model problems: First, we study random walks with random conductances under the annealed law. If we consider the Itô rough path, then we see a correction to the iterated integrals even though the underlying Markov process is reversible. If we consider the Stratonovich rough path, then there is no correction. The second example is a nonreversible Ornstein–Uhlenbeck process, while the last example is a diffusion in a periodic environment.

As a technical step, we prove an estimate for the p-variation of stochastic integrals with respect to martingales that can be viewed as an extension of the rough path Burkholder–Davis–Gundy inequality for local martingale rough paths of (In *Séminaire de Probabilités XLI* (2008) 421–438 Springer; In *Probability and Analysis in Interacting Physical Systems* (2019) 17–48 Springer; *J. Differential Equations* **264** (2018) 6226–6301) to the case where only the integrator is a local martingale.

1. Introduction. In recent years, there has been an increased interest in the link between homogenization and rough paths. It had been observed previously that homogenization often gives rise to nonstandard rough path limits [9, 21]. The more recent investigations were initiated with the work of Kelly and Melbourne [15–17] who study rough path limits of additive functionals of the form $\sqrt{n}\int_0^{n\cdot} f(Y_s)\,\mathrm{d}s$, where Y is a deterministic dynamical system with suitable mixing conditions. In that way, they are able to prove homogenization results for the convergence of deterministic multiscale systems of the type

$$\mathrm{d}X^n = \sqrt{n}b\big(X_t^n, Y_t^n\big)\,\mathrm{d}t,$$
$$\mathrm{d}Y^n = nf\big(Y_t^n\big)\,\mathrm{d}t,$$

for which under suitable conditions X^n converges to an autonomous stochastic differential equation. This line of research was picked up and extended, for example, by [1, 5, 8, 23, 24, 26]. More recent results also cover discontinuous limits [7].

Motivated by this problem, as well as by the aim of understanding the invariance principle for random walks in random environment in rough path topology, we want to study rough path invariance principles for additive functionals $\sqrt{n}\int_0^{n\cdot} f(X_s)\,\mathrm{d}s$ of Markov processes X in generic situations. If we are only interested in a central limit theorem at a fixed time, then there are of course many results of this type and many ways of showing them. See, for example, [27] for a recent and fairly general result. A particularly successful approach for

Received March 2020; revised September 2020.
MSC2020 subject classifications. 60L20, 60K37, 60F17, 82C41, 82B43.
Key words and phrases. Rough paths, homogenization, additive functionals of Markov processes, random walks among random conductances, Lépingle's Burkholder–Davis–Gundy inequality in p-variation, UCV condition.

proving such a central limit theorem and even the functional central limit theorem (invariance principle) is based on Dynkin's formula and martingale arguments, and it was developed by Kipnis and Varadhan [18] for reversible Markov processes and later extended to many other situations; see the nice monograph [19]. Here, we extend this approach to the rough path topology and we study some applications to model problems like random walks among random conductances, additive functionals of Ornstein–Uhlenbeck processes, and periodic diffusions.

This can also be seen as a complementary direction of research with respect to the recent advances in regularity structures [2–4], where the aim is to find generic convergence results for models associated to singular stochastic PDEs. In those works, the equations tend to be extremely complicated, but the approximation of the noise is typically quite simple (the prototypical example is just a mollification of the driving noise, but [4] also allow some stationary mixing random fields that converge to the space–time white noise by the central limit theorem). In our setting, the equation that we study is very simple (just a stochastic ODE), but the approximation of the noise is very complicated and (at least for us) it seems difficult to check whether the conditions of [4], Theorem 2.34, are satisfied for the kind of examples that we are interested in.

The most interesting model that we study here is probably the random walk among random conductances. Here, we distribute i.i.d. conductances $(\eta(\{x,y\}))_{x,y\in\mathbb{Z}^d:x\sim y}$ on the bonds of $\mathbb{Z}^d$ (where $x\sim y$ means that x and y are neighbors). Then we let a continuous time random walk move along $\mathbb{Z}^d$, with jump rate $\eta(\{x,y\})$ from x to y (resp., from y to x). We are interested in the large scale behavior, that is, we study $n^{-1/2}X_{nt}$ for $n\to\infty$. It is well known that the path itself converges in distribution under the annealed law to a Brownian motion B with an effective diffusion coefficient. Our contribution is to extend this convergence to the rough path topology, which allows us, for example, to understand the limit of discrete stochastic differential equations

$$\mathrm{d}Y_t^n = \sigma(Y_{t-}^n)\,\mathrm{d}X_t^n, \tag{1.1}$$

but also of SPDEs driven by X^n. And here we encounter a surprise: Even though X is in a certain sense reversible (more precisely the underlying Markov process of the environment as seen from the walker is reversible), the iterated integrals $\int_0^\cdot X_s^n\otimes \mathrm{d}X_s^n$ do not converge to $\int_0^\cdot B_s\otimes \mathrm{d}B_s$, but instead we see a correction: We have

$$\left(X^n, \int_0^\cdot X_{s-}^n\otimes \mathrm{d}X_s^n\right) \longrightarrow \left(B, \int_0^\cdot B_s\otimes \mathrm{d}B_s + \Gamma t\right),$$

where Γ is a correction given by

$$\begin{aligned}\Gamma &= \frac{1}{2}\langle B,B\rangle_1 - E_\pi\big[\eta(\{0,e_1\})\big]I_d \\ &= \frac{1}{2}\big(\langle B,B\rangle_1 - E_\pi\big[\eta(\{0,e_1\})+\eta(\{0,-e_1\})\big]I_d\big)\end{aligned}$$

for the law π of the random conductances. Of course, Γ vanishes if the conductances are deterministic (i.e., if π is a Dirac measure). But if the conductances are truly random, then typically the effective diffusion is not just given by the expected conductance, and in $d=1$ one can even show that this is never the case (see the discussion at the top of p. 89 of [19]). Therefore, Γ is typically nonzero, and the solution Y^n of (1.1) converges to the solution Y of

$$\mathrm{d}Y_t = \sigma(Y_t)\,\mathrm{d}B_t + \sum_{j,k,\ell}\partial_k\sigma_{\cdot j}(Y_t)\sigma_{k\ell}(Y_t)\Gamma_{\ell j}\,\mathrm{d}t.$$

If on the other hand we denote by $\tilde{X}^n$ the linear interpolation of the pure jump path X^n, then $(\tilde{X}^n, \int_0 \tilde{X}_s^n\,\mathrm{d}\tilde{X}_s^n)$ converges to the limit that we would naively expect, namely to the

Stratonovich rough path above B. From the point of view of stochastic calculus, this is a bit surprising: After all, there are stability results for Itô integrals [20], while the quadratic variation (i.e., the difference between Itô and Stratonovich integrals) is very unstable. In fact, we are not aware of any previous results of this type (naive limit for the Stratonovich rough path, correction for the Itô rough path), but it seems to be a generic phenomenon. The same effect appears for periodic diffusions, and we expect to see it for nearly all models treated in the monograph [19]. On the other hand, for ballistic random walks in random environment, after centering, a correction to the *Stratonovich* rough path is identified in terms of the expected stochastic area on a regeneration interval [23], Theorem 3.3 and [26], Theorem 1.5. Moreover, for random walks in deterministic periodic environments simple examples for nonvanishing corrections are available [24], Section 1.2, (or [23], Section 4.2). For processes that can be handled with the Kipnis–Varadhan approach, we generically expect to see a correction to the Stratonovich rough path if and only if the underlying Markov process is nonreversible.

Structure of the paper. In the next section, we introduce some basic notions from rough path theory. Section 3 presents our main result Theorem 3.3, the rough path invariance principle for additive functionals of stationary Markov processes, which holds under the same conditions as the abstract result in [19]. The proof is based on recent advances on Itô rough paths with jumps due to Friz and Zhang [12], on stability results for Itô integrals under the so-called UCV condition by Kurtz and Protter [20], on Lépingle's Burkholder–Davis–Gundy inequality in p-variation [22], and on repeated integrations by parts together with a new estimate on the p-variation of stochastic integrals (Proposition 3.8). In Section 4, we apply our abstract result to three model problems: random walks among random conductances, additive functionals of Ornstein–Uhlenbeck processes and periodic diffusions. Finally, Section 5 contains the proof of Proposition 3.8 which might be of independent interest.

Notation. For two families $(a_i)_{i\in I}$, $(b_i)_{i\in I}$ of real numbers indexed by I the notation $a_i \lesssim b_i$ means that $a_i \le cb_i$ for every $i \in I$ where $c \in (0,\infty)$ is a constant. Let $\Delta_T := \{s,t \in [0,T] : s \le t\}$ for $T > 0$. We interpret any function $X : [0,T] \to \mathbb{R}^d$ also as a function on Δ_T via $X_{s,t} := X_t - X_s$, $(s,t) \in \Delta_T$. For a metric space (E,d), we write $C([0,T],E)$, respectively, $D([0,T],E)$ for the continuous, respectively, càdlàg functions from $[0,T]$ to E. A function $X : \Delta_T \to E$ is called continuous, respestively, càdlàg if for all $s \in [0,T)$ the map $t \mapsto X_{s,t}$ on $[s,T]$ is continuous, respectively, càdlàg, and we write $C(\Delta_T,E)$, respectively, $D(\Delta_T,E)$ for the corresponding function spaces.

2. Elements of rough path theory.

Here, we recall some basic elements of rough path theory for Itô rough paths with jumps. See [12] for much more detail.

Let us write $\|X\|_{\infty,[0,T]} := \sup_{t\in[0,T]} |X_t|$ (resp., $\|\mathbb{X}\|_{\infty,[0,T]} := \sup_{(s,t)\in\Delta_T} |\mathbb{X}_{s,t}|$) to denote the uniform norm of $X \in D([0,T],\mathbb{R}^d)$ (resp., $\mathbb{X} \in D(\Delta_T,\mathbb{R}^{d\times d})$). For $0 < p < \infty$ and a normed space $(E,|\cdot|_E)$, we define the p-variation of $\Xi : \Delta_T \to E$ (and so in particular of $\Xi : [0,T] \to E$) by

$$\|\Xi\|_{p,[0,T]} := \Bigl(\sup_{\mathcal{P}} \sum_{[s,t]\in\mathcal{P}} |\Xi_{s,t}|^p\Bigr)^{1/p} \in [0,+\infty], \tag{2.1}$$

where the supremum is taken over all finite partitions $\mathcal{P}$ of $[0,T]$ and the summation is over all intervals $[s,t] \in \mathcal{P}$. Note that for any $0 < p \le q < \infty$, we have that $\|\Xi\|_{q,[0,T]} \le \|\Xi\|_{p,[0,T]}$.

DEFINITION 2.1 (p-variation rough path space). For $p \in [2,3)$, the space $D_{p\text{-var}}([0,T], \mathbb{R}^d \times \mathbb{R}^{d\times d})$ (resp., $C_{p\text{-var}}([0,T],\mathbb{R}^d \times \mathbb{R}^{d\times d})$) of càdlàg (resp., continuous) p-variation rough paths is defined by the subspace of all functions $(X, \mathbb{X}) \in D([0,T],\mathbb{R}^d) \times D(\Delta_T, \mathbb{R}^{d\times d})$ satisfying Chen's relation, that is,

$$\mathbb{X}_{r,t} - \mathbb{X}_{r,s} - \mathbb{X}_{s,t} = X_{r,s} \otimes X_{s,t} \tag{2.2}$$

for $0 \le r \le s \le t \le T$, and

$$\|(X,\mathbb{X})\|_{p,[0,T]} := |X_0| + \|X\|_{p,[0,T]} + \|\mathbb{X}\|_{p/2,[0,T]}^{1/2} < \infty. \tag{2.3}$$

The p-variation Skorohod distance on $D_{p\text{-var}}([0,T],(\mathbb{R}^d,\mathbb{R}^{d\times d}))$ is

$$\begin{aligned}\sigma_{p,[0,T]}\big((X,\mathbb{X}),(Y,\mathbb{Y})\big) := \inf_{\lambda\in\Lambda_T}\big\{|\lambda| \vee \big(\|X - Y\circ\lambda\|_{p,[0,T]} \\ + \|\mathbb{X} - \mathbb{Y}\circ(\lambda,\lambda)\|_{p/2,[0,T]}\big)\big\},\end{aligned}$$

where Λ_T are the strictly increasing bijective functions from $[0,T]$ onto itself, and $|\lambda| = \sup_{t\in[0,T]}|\lambda(t) - t|$. The uniform Skorohod distance is defined similarly, except with the p-variation respectively $p/2$-variation distance replaced by the uniform distance; see [12], Section 5, for details.

For $X, Y \in D([0,T],\mathbb{R}^d)$ we use the notation $\int_0^t Y_{s-} \otimes \mathrm{d}X_s$ for the left-point Riemann integral, that is,

$$\int_0^t Y_{s-}\otimes \mathrm{d}X_s := \int_{(0,t]} Y_{s-}\otimes \mathrm{d}X_s := \lim_{n\to\infty}\Big\{\sum_{[u,v]\in\mathcal{P}_n} Y_u \otimes (X_v - X_u)\Big\},$$

whenever this limit is well defined along an implicitly fixed sequence of partitions $(\mathcal{P}_n)$ of $[0,t]$ with mesh size going to zero. Note that if X is a semimartingale and Y is adapted to the same filtration, then this definition coincides with the Itô integral. We remark also that the iterated integrals

$$\mathbb{X}_{s,t} := \int_s^t X_{s,u-}\otimes \mathrm{d}X_u = \int_{(s,t]} X_{s,u-}\otimes \mathrm{d}X_u,$$

satisfy Chen's relation (2.2). Moreover, so do $\tilde{\mathbb{X}}_{s,t} := \mathbb{X}_{s,t} + (t-s)\Gamma$, for any fixed matrix Γ.

REMARK 2.2. Note that by Chen's relation $\mathbb{X}_{s,t} = \mathbb{X}_{0,t} - \mathbb{X}_{0,s} - X_{0,s}\otimes X_{s,t}$ whenever $0 \le s \le t \le T$ and, therefore,

$$\begin{aligned}\|\mathbb{X} - \mathbb{Y}\|_{\infty,[0,T]} = \sup_{0\le s<t\le T}|\mathbb{X}_{s,t} - \mathbb{Y}_{s,t}| \lesssim \|\mathbb{X}_{0,\cdot} - \mathbb{Y}_{0,\cdot}\|_{\infty,[0,T]} \\ + \big(\|X_{0,\cdot}\|_{\infty,[0,T]} \vee \|Y_{0,\cdot}\|_{\infty,[0,T]}\big)\|X_{0,\cdot} - Y_{0,\cdot}\|_{\infty,[0,T]}.\end{aligned}$$

Consequently, the uniform resp. Skorohod distance of the (one-parameter) paths $(X_{0,\cdot}, \mathbb{X}^n_{0,\cdot})$ and $(Y_{0,\cdot}, \mathbb{Y}_{0,\cdot})$ controls the uniform resp. Skorohod distance of $(X,\mathbb{X})$ and $(Y,\mathbb{Y})$.

The following lemma by [12] will be useful in the sequel.

LEMMA 2.3. *Let $(Z^n,\mathbb{Z}^n)$ be a sequence of càdlàg rough paths and let $p < 3$. Assume that there exists a càdlàg rough path $(Z,\mathbb{Z})$ such that $(Z^n,\mathbb{Z}^n) \to (Z,\mathbb{Z})$ in distribution in the Skorohod (resp., uniform) topology and that the family of real valued random variables $(\|(Z^n,\mathbb{Z}^n)\|_{p,[0,T]})_n$ is tight. Then $(Z^n,\mathbb{Z}^n) \to (Z,\mathbb{Z})$ in distribution in the p'-variation Skorohod (resp., uniform) topology for all $p' \in (p,3)$.*

PROOF. This follows from a simple interpolation argument, see the proof of Theorem 6.1 in [12]. □

Invariance principles for rough path sequences guarantee the convergence of the solutions to rough differential equations where the noise is approximated by the path sequence. Moreover, whenever the second level (the first order "iterated integrals") of the rough path has a correction, the limiting path solves a drift-modified rough equation defined explicitly in terms of the correction. More precisely, [12], Theorem 6.1 and Proposition 6.9, proved the following. Let (Z^n) be a sequence of semimartingales and assume that $(Z^n, \int_0^\cdot Z^n_{s-}\,\mathrm{d}Z^n_s)$ converges in distribution in p-variation Skorohod (resp., uniform) distance to a rough path $(Z, \mathbb{Z})$, where Z is a semimartingale and $\mathbb{Z}_{s,t} = \int_s^t Z_{s,r-} \otimes \mathrm{d}Z_r + \Gamma \times (t-s)$ for $\Gamma \in \mathbb{R}^{d\times d}$. Then the solutions (Y^n) of

$$\mathrm{d}Y^n_t = \sigma(Y^n_{t-})\,\mathrm{d}Z^n_t, \quad Y^n_0 = y,$$

converge in distribution in the Skorohod (resp., uniform) topology to the solution Y of

$$\mathrm{d}Y_t = \sigma(Y_{t-})\,\mathrm{d}(Z,\mathbb{Z})_t = \sigma(Y_{t-})\,\mathrm{d}Z_t + \sum_{j,k,\ell} \partial_k \sigma_{\cdot j}(Y_t)\sigma_{k\ell}(Y_t)\Gamma_{\ell j}\,\mathrm{d}t, \quad Y_0 = y,$$

where $\mathrm{d}(Z,\mathbb{Z})_t$ denotes rough path integration and $\mathrm{d}Z_t$ is just the Itô integral.

3. Additive functionals as rough paths.

Here, we present our abstract convergence result for additive functionals of stationary Markov processes. We place ourselves in the context of Chapter 2 in [19]: Let $(X_t)_{t\geq 0}$ be a càdlàg Markov process in a filtration satisfying the usual conditions, with values in a Polish space E, and let π be a stationary probability measure for X and $X_0 \sim \pi$. We assume that the transition semigroup of X can be extended to a strongly continuous contraction semigroup $(T_t)_{t\geq 0}$ on $L^2(\pi)$. We write $\mathcal{L}$ for the infinitesimal generator of (T_t) and we assume that π is ergodic for $\mathcal{L}$, that is, that F is π-almost surely constant whenever $\pi(\{\mathcal{L}F = 0\}) = 1$. We also assume that there exists a common core $\mathcal{C}$ for $\mathcal{L}$ and $\mathcal{L}^*$, where $\mathcal{L}^*$ is the $L^2(\pi)$-adjoint of $\mathcal{L}$, and that $\mathcal{C}$ contains the constant functions. We write

$$\mathcal{L}_S = \frac{1}{2}(\mathcal{L} + \mathcal{L}^*) \quad \text{and} \quad \mathcal{L}_A = \frac{1}{2}(\mathcal{L} - \mathcal{L}^*).$$

NOTATION 1. We write $\mathbb{P}$ or $\mathbb{P}_\pi$ (and $\mathbb{E}$ or $\mathbb{E}_\pi$) for the distribution of the stationary process $(X_t)_{t\geq 0}$ on the Skorohod space $D(\mathbb{R}_+, E)$. The notation E_π is reserved for the integration with respect to π on the space E.

DEFINITION 3.1. The space $\mathcal{H}^1$ is defined as the completion of $\mathcal{C}$ with respect to the norm

$$\|F\|^2_{\mathcal{H}^1} := E_\pi[F(-\mathcal{L})F] = E_\pi[F(-\mathcal{L}_S)F],$$

or more precisely we identify $F, G \in \mathcal{C}$ if $\|F - G\|_{\mathcal{H}^1} = 0$, and $\mathcal{H}^1$ is the completion of the equivalence classes. The space $\mathcal{H}^{-1}$ is the dual of $\mathcal{H}^1$: We define for $F \in \mathcal{C}$,

$$\|F\|^2_{\mathcal{H}^{-1}} := \sup_{\substack{G\in\mathcal{C}:\\ \|G\|_{\mathcal{H}^1}\leq 1}} E_\pi[FG]^2 = \sup_{G\in\mathcal{C}}\{2E_\pi[FG] - \|G\|^2_{\mathcal{H}^1}\}$$

and then $\mathcal{H}^{-1}$ is the completion of $\{F \in \mathcal{C} : \|F\|_{\mathcal{H}^{-1}} < \infty\}$ with respect to $\|\cdot\|_{\mathcal{H}^{-1}}$. We define $\langle\cdot,\cdot\rangle_{\mathcal{H}^1}$ and $\langle\cdot,\cdot\rangle_{\mathcal{H}^{-1}}$ to be the naturally defined scalar products on $\mathcal{H}^1$ and $\mathcal{H}^{-1}$. If F takes values in $\mathbb{R}^d$ we also write $F \in \mathcal{H}^1(\mathbb{R}^d)$, $F \in \mathcal{H}^{-1}(\mathbb{R}^d)$ or $F \in L^2(\pi, \mathbb{R}^d)$, etc, whenever it is the case coordinatewise. The corresponding norms are then defined, for example, as $\|F\|^2_{\mathcal{H}^1} = \sum_{i=1}^d \|F^i\|^2_{\mathcal{H}^1}$, with scalar product $\langle F, G\rangle_{\mathcal{H}^1} = \sum_{i=1}^d \langle F^i, G^i\rangle_{\mathcal{H}^1}$.

Note that if $E_\pi[F] \neq 0$ then we can take $G = \lambda \in \mathcal{C}$ for $\lambda \in \mathbb{R}$ so that $\|G\|_{\mathcal{H}^1} = 0$ and by sending $\lambda \to \pm\infty$ we see that $\|F\|_{\mathcal{H}^{-1}} = \infty$. Therefore, we get that $E_\pi[F] = 0$ for all $F \in L^2(\pi) \cap \mathcal{H}^{-1}$.

Let now $F \in \mathcal{H}^{-1}(\mathbb{R}^d) \cap L^2(\pi, \mathbb{R}^d)$. Our aim is to derive a scaling limit for the (absolutely continuous) rough path $(Z^n, \mathbb{Z}^n)$, where

$$Z^n_{s,t} := \frac{1}{\sqrt{n}} \int_{ns}^{nt} F(X_r)\,\mathrm{d}r, \qquad \mathbb{Z}^n_{s,t} := \int_s^t Z^n_{s,r} \otimes \mathrm{d}Z^n_r,$$

and the integration with respect to Z^n_r is in the Riemann–Stieltjes sense. Let us first recall the following result.

LEMMA 3.2 ([19], Theorem 2.33). *Assume that π is ergodic for $\mathcal{L}^*$. Let $F \in L^2(\pi, \mathbb{R}^d) \cap \mathcal{H}^{-1}(\mathbb{R}^d)$ and assume that the solution Φ_λ to the resolvent equation $(\lambda - \mathcal{L})\Phi_\lambda = F$ with $\lambda > 0$ satisfies*

$$\lim_{\lambda \to 0} \big(\sqrt{\lambda}\|\Phi_\lambda\|_{L^2(\pi)} + \|\Phi_\lambda - \Phi\|_{\mathcal{H}^1}\big) = 0 \tag{3.1}$$

for some $\Phi \in \mathcal{H}^1(\mathbb{R}^d)$. Then $(Z^n)_n$ converges in distribution in $C([0,T], \mathbb{R}^d)$ to a Brownian motion with covariance matrix

$$\langle B, B\rangle_t = 2t\big(\langle \Phi^k, \Phi^\ell\rangle_{\mathcal{H}^1}\big)_{1 \le k,\ell \le d} = 2t \lim_{\lambda \to 0}\big(\langle \Phi^k_\lambda, \Phi^\ell_\lambda\rangle_{\mathcal{H}^1}\big)_{1 \le k,\ell \le d}.$$

Our aim is to extend Lemma 3.2 to the rough path topology. Our main result is the following.

THEOREM 3.3. *Let $p > 2$. Under the assumptions of Lemma 3.2 the process $(Z^n, \mathbb{Z}^n)$ converges weakly to*

$$\left(B_t, \int_0^t B_s \otimes \circ\,\mathrm{d}B_s + \Gamma t\right)_{t \ge 0} \tag{3.2}$$

in the (uniform) p-variation topology on $C_{p\text{-var}}([0,T], \mathbb{R}^d \times \mathbb{R}^{d\times d})$, where B is the same Brownian motion as in Lemma 3.2,

$$\int_0^t B_s \otimes \circ\,\mathrm{d}B_s := \int_0^t B_s \otimes \mathrm{d}B_s + \frac{1}{2}\langle B, B\rangle_t$$

denotes Stratonovich integration, and Γ is given by the following limit, which exists:

$$\Gamma = \lim_{\lambda \to 0} E_\pi[\Phi_\lambda \otimes \mathcal{L}_A \Phi_\lambda].$$

For the rest of the section, we shall assume without further mention that the conditions of Theorem 3.3 are satisfied.

REMARK 3.4. As Z^n is of finite variation and *absolutely continuous* the iterated integral $\int_0^t Z^n_s \otimes \mathrm{d}Z^n_s$ "wants" to converge to the Stratonovich integral, and Γ describes the area correction. Note that $\Gamma = 0$ if $\mathcal{L}$ is symmetric, that is, if X is reversible, so in that case we indeed obtain the Stratonovich rough path over B.

REMARK 3.5. In Lemma 3.2 and Theorem 3.3, the ergodicity of π with respect to $\mathcal{L}^*$ is only needed for proving the tightness of $(Z^n, \mathbb{Z}^n)$ in the uniform topology. This is relatively subtle because we need tightness of certain martingales M^Ψ for which we only know that $\mathbb{E}[\langle M^\Psi\rangle_t - \langle M^\Psi\rangle_s] \lesssim |t-s|$, which is insufficient to apply Kolmogorov's continuity criterion. If we can show $\mathbb{E}[|\langle M^\Psi\rangle_t - \langle M^\Psi\rangle_s|^{1+\delta}] \lesssim |t-s|^{1+\delta}$ for some $\delta > 0$ and for the martingales M^Ψ of Lemma 3.6 below, then we do not need the ergodicity of π with respect to $\mathcal{L}^*$ (although we do need ergodicity with respect to $\mathcal{L}$).

The strategy for proving Theorem 3.3 is to apply Lemma 2.3, which separates the convergence proof into two problems: Showing tightness of the p-variation norms of $(Z^n, \mathbb{Z}^n)$ and showing convergence in the Skorohod topology. For the latter, we follow a similar strategy as in [19] and combine it with tools from rough paths and a simple integration by parts formula.

Let us formally sketch how the correction Γ arises, under the assumption that we can solve the Poisson equation $-\mathcal{L}\Phi = F$ (which is, e.g., the case if X has a spectral gap and $E_\pi[F] = 0$). In that case, we have

$$Z_t^n = \frac{1}{\sqrt{n}}\Phi(X_0) - \frac{1}{\sqrt{n}}\Phi(X_{nt}) + M_t^n$$

for a sequence of martingales (M^n). Therefore,

$$\int_0^t Z_s^n \otimes \mathrm{d}Z_s^n = \int_0^t \big(\Phi(X_0) - \Phi(X_{ns})\big) \otimes F(X_{ns})\,\mathrm{d}s + \int_0^t M_s^n \otimes \mathrm{d}Z_s^n.$$

By the ergodic theorem, the first term on the right-hand side converges to

$$t\big(\Phi(X_0) \otimes E_\pi[F] - E_\pi[\Phi \otimes F]\big) = tE_\pi[\Phi \otimes \mathcal{L}\Phi].$$

To understand the remaining contribution, we use integration by parts: Since Z^n is of finite variation, we have

$$\int_0^t M_s^n \otimes \mathrm{d}Z_s^n = M_t^n \otimes Z_t^n - \int_0^t Z_s^n \otimes \mathrm{d}M_s^n.$$

Since X is stationary, we have $\|Z^n - M^n\|_{L^2(\pi)} = O(n^{-1/2})$, and (Z^n, M^n) converges jointly to (B, B). The martingale sequence (M^n) satisfies the "UCV condition" (see Definition 3.14 and the discussion below it for details) and, therefore, $\int_0^t Z_s^n \otimes \mathrm{d}M_s^n \to \int_0^t B_s \otimes \mathrm{d}B_s$. After passing to the limit, we apply integration by parts once more and deduce that

$$\int_0^t M_s^n \otimes \mathrm{d}Z_s^n \to B_t \otimes B_t - \int_0^t B_s \otimes \mathrm{d}B_s = \int_0^t B_s \otimes \mathrm{d}B_s + \langle B, B\rangle_t.$$

So overall

$$\begin{aligned}\int_0^t Z_s^n \otimes \mathrm{d}Z_s^n &\to \int_0^t B_s \otimes \circ\,\mathrm{d}B_s + \frac{1}{2}\langle B, B\rangle_t + tE_\pi[\Phi \otimes \mathcal{L}\Phi] \\ &= \int_0^t B_s \otimes \circ\,\mathrm{d}B_s + tE_\pi\big[\Phi \otimes (-\mathcal{L}_S)\Phi\big] + tE_\pi[\Phi \otimes \mathcal{L}\Phi] \\ &= \int_0^t B_s \otimes \circ\,\mathrm{d}B_s + tE_\pi[\Phi \otimes \mathcal{L}_A\Phi].\end{aligned}$$

3.1. *Tightness of the p-variation norms.* Throughout this section, we will often use the following representation of additive functionals.

LEMMA 3.6. *Let $\Psi \in \mathcal{C}(\mathbb{R}^m)$. Then we have for $T > 0$,*

$$\int_0^t \mathcal{L}_S\Psi(X_s)\,\mathrm{d}s = \frac{1}{2}\big(M_t^\Psi + \hat{M}_T^\Psi - \hat{M}_{T-t}^\Psi\big), \quad t \in [0, T], \tag{3.3}$$

where M^Ψ is a martingale and $\hat{M}^\Psi$ is a martingale with respect to the backward filtration $\hat{\mathcal{F}}_t = \sigma(\hat{X}_s : s \le t)$, where $\hat{X}_s := X_{T-s}$, such that

$$\mathbb{E}\big[\langle M^\Psi\rangle_t\big] = \mathbb{E}\big[\langle \hat{M}^\Psi\rangle_t\big] = 2E_\pi\big[\Psi \otimes (-\mathcal{L}_S)\Psi\big]t = 2t\big(\langle \Psi^k, \Psi^\ell\rangle_{\mathcal{H}^1}\big)_{1\le k,\ell\le m} \tag{3.4}$$

for $t \in [0, T]$. Assume that π is ergodic for $\mathcal{L}^*$. Then under the rescaling $T \to nT$ and $M_t^{\Psi,n} = n^{-1/2}M_{nt}^\Psi$ and similarly for $\hat{M}^{\Psi,n}$ both processes converge in distribution in

$D([0,T],\mathbb{R}^m)$ *to a Wiener process. If* $G, H \in L^2(\pi,\mathbb{R}^m)$ *and* $A_{s,t} = \int_s^t \int_s^{r_1} G(X_{r_2})\,\mathrm{d}r_2 \otimes H(X_{r_1})\,\mathrm{d}r_1$ *for* $0 \le s < t \le T$, *then*

$$A_{s,t} = \frac{1}{2}\int_s^t \int_s^{r_1} G(X_{r_2})\,\mathrm{d}r_2 \otimes \mathrm{d}M_{r_1}^{\Psi} - \frac{1}{2}\int_{T-t}^{T-s}\int_{T-r_1}^{t} G(X_{r_2})\,\mathrm{d}r_2 \otimes \mathrm{d}\hat{M}_{r_1}^{\Psi} \tag{3.5}$$

$$+ \frac{1}{2}\int_s^t G(X_r)\,\mathrm{d}r \otimes \big(\hat{M}_{T-s}^{\Psi} - \hat{M}_{T-t}^{\Psi}\big)$$

$$+ \int_s^t \int_s^{r_1} G(X_{r_2})\,\mathrm{d}r_2 \otimes (H - \mathcal{L}_S\Psi)(X_{r_1})\,\mathrm{d}r_1. \tag{3.6}$$

PROOF. The representation (3.3) is obtained, for example, by applying Dynkin's formula to $\Psi(X)$ and $\Psi(\hat{X})$ on $[0,u]$, $u \in [0,T]$, and then computing $M_t^{\Psi} + \hat{M}_T^{\Psi} - \hat{M}_{T-t}^{\Psi}$. If Ψ^2 is in the domain of $\mathcal{L}$, then also (3.4) follows from Dynkin's formula; otherwise, we use an approximation argument (see p. 35 of [19]). For the convergence of $M^{\Psi,n}$ and $\hat{M}^{\Psi,n}$, see the proof of Theorem 2.32/2.33 in [19]. The representation for $A_{s,t}$ follows by writing the integral against $\hat{M}_{T-\cdot}^{\Psi}$ as a limit of Riemann sums—note that $\int_0^{\cdot} G(X_r)\,\mathrm{d}r$ is continuous and of finite variation, so the integral is defined pathwise and we do not need to worry about quadratic covariations or the difference between forward and backward integral. □

In order to show that $\|(Z^n,\mathbb{Z}^n)\|_{p,[0,T]}$ is a tight sequence of real valued random variables, we first recall the following estimate.

LEMMA 3.7 ([14], Corollary 3.5). *Let* $G \in L^2(\pi) \cap \mathcal{H}^{-1}$ *and* $T > 0$ *and* $p > 2$. *Then*

$$\mathbb{E}\bigg[\bigg\|\int_0^{\cdot} G(X_s)\,\mathrm{d}s\bigg\|_{\infty,[0,T]}^2\bigg] \le \mathbb{E}\bigg[\bigg\|\int_0^{\cdot} G(X_s)\,\mathrm{d}s\bigg\|_{p,[0,T]}^2\bigg] \lesssim T\|G\|_{\mathcal{H}^{-1}}^2.$$

PROOF. The inequality from the left is immediate from the definition of the norms. For the second estimate, see Corollary 3.5 in [14]. This corollary is written for the specific process studied in [14], but the proof carries over verbatim to the general situation considered here. □

In particular, we get

$$\mathbb{E}\Big[\sup_{t\le T}|Z_t^n|^2\Big] \le \mathbb{E}\big[\|Z^n\|_{p,[0,T]}^2\big] \lesssim T\|F\|_{\mathcal{H}^{-1}}^2. \tag{3.7}$$

To bound $\|\mathbb{Z}^n\|_{p,[0,T]}$, we need the following auxiliary result, which is the core technical result of this section and which replaces the Burkholder–Davis–Gundy inequality for local martingale rough paths of [6, 12] in the case where only the integrator is a local martingale.

PROPOSITION 3.8. *Let* $(Y_t)_{t\in[0,T]}$ *be a predictable càdlàg process with* $Y_0 = 0$ *and such that* $\mathbb{E}[\|Y\|_{p,[0,T]}^2] < \infty$ *for some* $p > 2$ *and let* $(N_t)_{t\in[0,T]}$ *be a càdlàg local martingale with* $\mathbb{E}[\langle N\rangle_T] < \infty$. *Define* $A_{s,t} := \int_s^t Y_{s,r}\,\mathrm{d}N_r$. *Then for any* $q > 4p/(p+2) > 2$ *and* $0 < \varepsilon < 1/2$

$$\mathbb{E}\big[\|A\|_{q/2,[0,T]}^{1-\varepsilon}\big] \lesssim \big(1 + \mathbb{E}\big[\|Y\|_{p,[0,T]}^2\big]^{1/2}\big)\mathbb{E}\big[\langle N\rangle_T\big]^{(1-\varepsilon)/2}.$$

To not disrupt the flow of reading, we give the proof in Section 5 below (see, in particular, the more precise result in Proposition 5.2).

REMARK 3.9. After we submitted the first version of this paper to the arXiv, a preprint by Friz and Zorin-Kranich [11] appeared in which, among others, they derive an improved version of the last estimate which has the correct scaling.

COROLLARY 3.10. *Let $G, H \in \mathcal{H}^{-1} \cap L^2(\pi)$ and set*

$$A_{s,t} = \int_s^t \int_s^{r_1} G(X_{r_2})\,\mathrm{d}r_2 H(X_{r_1})\,\mathrm{d}r_1.$$

Then we have for all $p > 2$ and $T > 0$ and $\varepsilon \in (0, 1/2)$,

$$\mathbb{E}[\|A\|_{p/2,[0,T]}^{1-\varepsilon}] \lesssim (1 + T^{1/2}\|G\|_{\mathcal{H}^{-1}})(1 + T^{1/2}\|H\|_{\mathcal{H}^{-1}}).$$

PROOF. Lemma 3.6 shows that

$$\begin{aligned} A_{s,t} = {}& \frac{1}{2}\int_s^t \int_s^{r_1} G(X_{r_2})\,\mathrm{d}r_2\,\mathrm{d}M^{\Psi}_{r_1} - \frac{1}{2}\int_{T-s}^{T-t}\int_{T-r_1}^{t} G(X_{r_2})\,\mathrm{d}r_2\,\mathrm{d}\hat{M}^{\Psi}_{r_1} \\ & + \frac{1}{2}\int_s^t G(X_r)\,\mathrm{d}r\big(\hat{M}^{\Psi}_{T-s} - \hat{M}^{\Psi}_{T-t}\big) \\ & + \int_s^t \int_s^{r_1} G(X_{r_2})\,\mathrm{d}r_2 (H - \mathcal{L}_S\Psi)(X_{r_1})\,\mathrm{d}r_1. \end{aligned} \tag{3.8}$$

The first two terms on the right-hand side will be controlled with Proposition 3.8 and Lemma 3.7. The third term (3.8) is bounded by

$$\left|\frac{1}{2}\int_s^t G(X_r)\,\mathrm{d}r\big(\hat{M}^{\Psi}_{T-s} - \hat{M}^{\Psi}_{T-t}\big)\right| \lesssim \left\|\int_0^{\cdot} G(X_r)\,\mathrm{d}r\right\|_{p,[s,t]} \|\hat{M}^{\Psi}\|_{p,[T-t,T-s]}, \tag{3.9}$$

and the fourth term by

$$\begin{aligned} & \left|\int_s^t \int_s^{r_1} G(X_{r_2})\,\mathrm{d}r_2 (H - \mathcal{L}_S\Psi)(X_{r_1})\,\mathrm{d}r_1\right| \\ & \quad \lesssim \sup_{r\in[s,t]}\left|\int_0^r G(X_{r_2})\,\mathrm{d}r_2\right| \int_s^t \big|(H - \mathcal{L}_S\Psi)(X_{r_1})\big|\,\mathrm{d}r_1. \end{aligned} \tag{3.10}$$

Recall also Lépingle's p-variation Burkholder–Davis–Gundy inequality (see Theorem A.2), and note that $\mathbb{E}[[M^{\Psi}]_T] = \mathbb{E}[\langle M^{\Psi}\rangle_T]$ which can be easily seen by stopping the local martingale $[M^{\Psi}] - \langle M^{\Psi}\rangle$ and then applying monotone convergence. Combining Proposition 3.8 with (3.8)–(3.10), we obtain

$$\begin{aligned} & \mathbb{E}[\|A\|_{p/2,[0,T]}^{1-\varepsilon}] \\ & \lesssim \left(1 + \mathbb{E}\left[\left\|\int_0^{\cdot} G(X_r)\,\mathrm{d}r\right\|_{p,[0,T]}^2\right]^{1/2}\right)\big(1 + |\mathbb{E}[\langle M^{\Psi}\rangle_T]|^{1/2} \\ & \quad + |\mathbb{E}[\langle \hat{M}^{\Psi}\rangle_T]|^{1/2}\big) + \mathbb{E}\left[\left\|\int_0^{\cdot} G(X_r)\,\mathrm{d}r\right\|_{p,[0,T]}^2\right]^{(1-\varepsilon)/2} \mathbb{E}[\|\hat{M}^{\Psi}\|_{p,[0,T]}^2]^{(1-\varepsilon)/2} \\ & \quad + \mathbb{E}\left[\sup_{r\in[0,T]}\left|\int_0^r G(X_{r_2})\,\mathrm{d}r_2\right|^2\right]^{(1-\varepsilon)/2} \mathbb{E}\left[\left(\int_0^T \big|(H - \mathcal{L}_S\Psi)(X_{r_1})\big|\,\mathrm{d}r_1\right)^2\right]^{(1-\varepsilon)/2} \\ & \lesssim \left(1 + \mathbb{E}\left[\left\|\int_0^{\cdot} G(X_r)\,\mathrm{d}r\right\|_{p,[0,T]}^2\right]^{1/2}\right)\big(1 + T^{1/2}\|\Psi\|_{\mathcal{H}^1} + T\|H - \mathcal{L}_S\Psi\|_{L^2(\pi)}\big) \\ & \lesssim (1 + T^{1/2}\|G\|_{\mathcal{H}^{-1}})\big(1 + T^{1/2}\|\Psi\|_{\mathcal{H}^1} + T\|H - \mathcal{L}_S\Psi\|_{L^2(\pi)}\big) \end{aligned}$$

for $0 < \varepsilon < \frac{1}{2}$, where the last step follows from Lemma 3.7. Now we take $\Psi = \Phi^H_\lambda$ as the solution to the Poisson equation $(\lambda - \mathcal{L}_S)\Phi^H_\lambda = -H$. Note that in general $\Phi^H_\lambda \notin \mathcal{C}$, but we can approximate Φ^H_λ with functions in $\mathcal{C}$ and get the same estimate. Then standard estimates for the solution of the resolvent equation (see equation (2.15) in [19]) give $\|\Phi^H_\lambda\|_{\mathcal{H}^1} + \sqrt{\lambda}\|\Phi^H_\lambda\|_{L^2(\pi)} \lesssim \|H\|_{\mathcal{H}^{-1}}$, and since $H - \mathcal{L}_S\Phi^H_\lambda = \lambda\Phi^H_\lambda$ we can send $\lambda \to 0$ to deduce the claimed estimate. □

COROLLARY 3.11. *The sequence* $(\|Z^n\|_{p,[0,T]}\|, \mathbb{Z}^n\|_{p/2,[0,T]})_n$ *is tight for all* $T > 0$, $p > 2$.

PROOF. By equation (3.7), it suffices to show that $\mathbb{E}[\|\mathbb{Z}^n\|^{1-\varepsilon}_{p/2,[0,T]}] \le C$ for all n and some $\varepsilon \in (0, 1)$. But this follows from Corollary 3.10: We set $G = H = n^{-1/2}F$ and replace T with nT to obtain for any $\varepsilon \in (0, 1/2)$,

$$\begin{aligned}\mathbb{E}[\|\mathbb{Z}^n\|^{1-\varepsilon}_{p/2,[0,T]}] &\lesssim (1 + (nT)^{1/2}\|n^{-1/2}F\|_{\mathcal{H}^{-1}})(1 + (nT)^{1/2}\|n^{-1/2}F\|_{\mathcal{H}^{-1}}) \\ &= (1 + T^{1/2}\|F\|_{\mathcal{H}^{-1}})^2.\end{aligned}$$

□

3.2. *Convergence of the full path.* To prove tightness of the p-variation norms, we worked with the forward-backward decomposition of Lemma 3.6. But since the process $\hat{M}^\Psi$ from that lemma is only a martingale in the backward filtration, this decomposition is not useful for identifying the limit. So here we work instead with the following decomposition based on the resolvent equation.

LEMMA 3.12. *For* $\lambda > 0$, *we write* Φ_λ *for the solution of the resolvent equation* $(\lambda - \mathcal{L})\Phi_\lambda = F$. *Then*

$$\lambda\|\Phi_\lambda\|^2_{L^2(\pi)} + \|\Phi_\lambda\|^2_{\mathcal{H}^1} \le \|F\|^2_{\mathcal{H}^{-1}} \tag{3.11}$$

and there exists a martingale $M^{\lambda,1}$ *with* $M^{\lambda,1}_0 = 0$ *and with* $\mathbb{E}[\langle M^{\lambda,1}\rangle_t] = 2E_\pi[\Phi_\lambda \otimes (-\mathcal{L}_S)\Phi_\lambda]t$, *such that*

$$\int_0^t F(X_s)\,\mathrm{d}s = \Phi_\lambda(X_0) - \Phi_\lambda(X_t) + \int_0^t \lambda\Phi_\lambda(X_s)\,\mathrm{d}s + M^{\lambda,1}_t =: R^{\lambda,1}_t + M^{\lambda,1}_t.$$

PROOF. This formally follows by applying Dynkin's formula to Φ_λ, and to make it rigorous if $\Phi_\lambda \otimes \Phi_\lambda \notin \mathrm{dom}(\mathcal{L})$ one can use an approximation argument (see p. 35 of [19]). □

We write $M^{\lambda,n}_t := n^{-1/2}M^{\lambda,1}_{nt}$ and $R^{\lambda,n}_t := n^{-1/2}R^{\lambda,1}_{nt}$.

LEMMA 3.13. *Assume* (3.1). *Then there exist processes* $R^n, M^n \in D(\mathbb{R}_+, \mathbb{R}^d)$ *such that for all* $T > 0$ *and* $n \in \mathbb{N}$,

$$\lim_{\lambda\to 0}\Big\{\mathbb{E}\Big[\sup_{t\le T}|M^n_t - M^{\lambda,n}_t|^2\Big] + \mathbb{E}\Big[\sup_{t\le T}|R^n_t - R^{\lambda,n}_t|^2\Big]\Big\} = 0.$$

Moreover, M^n *is a martingale with* $\mathbb{E}[\langle M^n\rangle_t] = 2t\lim_{\lambda\to 0} E_\pi[\Phi_\lambda \otimes (-\mathcal{L}_S)\Phi_\lambda]$.

PROOF. This is all shown in [19]; see Lemma 2.9 and (2.26) therein. □

The following notion was introduced by Kurtz–Protter [20].

DEFINITION 3.14 (UCV condition). Let $(X^n)_{n\ge 1} \subset D([0, T], \mathbb{R})$ be a sequence of càdlàg local martingales. We say that $(X^n)_{n\ge 0}$ satisfies the *Uniformly Controlled Variation* (UCV) condition if

$$\sup_n \mathbb{E}[[X^n]_T] < \infty.$$

Strictly speaking, this is a very particular special case of the definition by Kurtz and Protter, who are much more permissive and consider general semimartingales rather than local martingales, and they allow for localization with stopping times as well as truncation of large jumps. But here we only need the special case above.

The celebrated result of Kurtz–Protter [20] guarantees the convergence in the Skorohod topology of the stochastic integrals of a sequence of càdlàg local martingales satisfying the UCV condition. Before we state it, recall that a sequence of processes $(Y^n)_{n\in\mathbb{N}}$ in $D(\mathbb{R}_+,\mathbb{R}^d)$ is called *C-tight* if it is tight in the Skorohod topology and all limit points are continuous processes.

THEOREM 3.15 ([20], Theorem 2.2). *Let $(X^n,Y^n)_{n\geq 1}\subset D([0,T],\mathbb{R}^2)$ be converging in probability in the Skorohod topology (or jointly in distribution) to a pair $(X,Y)\in D([0,T],\mathbb{R}^2)$, where X^n, Y^n are adapted to some given filtrations $\mathcal{F}^n$. Suppose that $(X^n)_{n\geq 1}$ is a sequence of local martingales which satisfies the UCV condition. Then X is a semimartingale in a filtration with respect to which Y is adapted and $(X^n, Y^n, \int_0^\cdot Y^n_{s-}\,\mathrm{d}X^n_s)$ converges to $(X,Y,\int_0^\cdot Y_{s-}\,\mathrm{d}X_s)$ as $n\to\infty$ in probability (or weakly) in $D([0,T],\mathbb{R}^3)$. In particular, if in addition $\int_0^\cdot Y_{s-}\,\mathrm{d}X_s\in C([0,T],\mathbb{R})$, then $\int_0^\cdot Y^n_{s-}\,\mathrm{d}X^n_s$ is C-tight.*

Note that the assumption that $(X^n,Y^n)_{n\geq 1}$ converges as a pair in $D([0,T],\mathbb{R}^2)$ is stronger than the joint convergence of $(X^n)_{n\geq 1}$ and $(Y^n)_{n\geq 1}$ in $D([0,T],\mathbb{R})$. In particular, by the continuous mapping theorem it implies, for example, the convergence of X^nY^n to XY in $D([0,T],\mathbb{R})$. This leads to the following application.

COROLLARY 3.16. *Let $(X^n,Y^n)_{n\geq 1}\subset D([0,T],\mathbb{R}^2)$ satisfy the same assumptions as in Theorem 3.15. If in addition $(Y_n)_{n\geq 1}$ is a sequence of semimartingales and $(X^n,Y^n,[X^n,Y^n])$ converges to (X,Y,A) in probability (or jointly in distribution) in $D([0,T],\mathbb{R}^3)$, where Y is a semimartingale and A is an adapted càdlàg process of finite variation, then*

$$\left(X^n, Y^n, \int_0^\cdot X^n_{s-}\,\mathrm{d}Y^n_s, \int_0^\cdot Y^n_{s-}\,\mathrm{d}X^n_s\right)\longrightarrow\left(X, Y, \int_0^\cdot X_{s-}\,\mathrm{d}Y_s+[X,Y]-A, \int_0^\cdot Y_{s-}\,\mathrm{d}X_s\right)$$

in probability (or weakly) in $D([0,T],\mathbb{R}^4)$. In particular, if in addition $\int_0^\cdot X_{s-}\,\mathrm{d}Y_s\in C([0,T],\mathbb{R})$, then $\int_0^\cdot X^n_{s-}\,\mathrm{d}Y^n_s$ is C-tight.

PROOF. Using integration by parts, we have

$$\int_0^\cdot X^n_{s-}\,\mathrm{d}Y^n_s = X^n_\cdot Y^n_\cdot - X^n_0Y^n_0-\int_0^\cdot Y^n_{s-}\,\mathrm{d}X^n_s-[X^n,Y^n],$$

so that the claim follows from the Kurz–Protter result together with another integration by parts:

$$X_\cdot Y_\cdot - X_0Y_0-\int_0^\cdot Y_{s-}\,\mathrm{d}X_s - A = \int_0^\cdot X_{s-}\,\mathrm{d}Y_s+[X,Y]-A. \qquad \square$$

The following corollary completes the proof of Theorem 3.3.

COROLLARY 3.17. *Under the assumptions of Theorem 3.3 the process $(Z^n,\mathbb{Z}^n)$ converges in distribution in the p-variation topology on $C(\mathbb{R}_+,\mathbb{R}^d\oplus\mathbb{R}^{d\otimes d})$ to*

$$\left(B_t, \int_0^t B_s\otimes\circ\,\mathrm{d}B_s+\Gamma t\right)_{t\geq 0}, \tag{3.12}$$

where B is a d-dimensional Brownian motion with covariance

$$2t \lim_{\lambda\to 0} E_\pi[\Phi_\lambda \otimes (-\mathcal{L}_S)\Phi_\lambda] = 2t \lim_{\lambda\to 0} \langle \Phi_\lambda, \otimes \Phi_\lambda\rangle_{\mathcal{H}^1},$$

and where

$$\Gamma = \lim_{\lambda\to 0} E_\pi[\Phi_\lambda \otimes \mathcal{L}_A \Phi_\lambda].$$

PROOF. Let $Z^n = M^n + R^n$ as above. In Theorem 2.32 of [19], it is shown that both (M^n) and (Z^n) converge in distribution in the Skorohod topology on $D(\mathbb{R}_+, \mathbb{R}^d)$ to a Brownian motion B with covariance $\langle B, B\rangle_t = 2t \lim_{\lambda\to 0} E_\pi[\Phi_\lambda \otimes (-\mathcal{L}_S)\Phi_\lambda]$. Therefore, both Z^n and M^n are C-tight, and thus also R^n is C-tight. It is shown in Proposition 2.8 of [19] that $\mathbb{E}[|R^n_t|^2] \to 0$ for each fixed $t \ge 0$, which together with the C-tightness gives the convergence of R^n to zero in distribution in $C(\mathbb{R}_+, \mathbb{R}^d)$ (and thus in probability because the limit is deterministic). Since $Z^n = M^n + R^n$, this gives the joint convergence of (Z^n, M^n, R^n) in distribution in $C(\mathbb{R}_+, \mathbb{R}^{3d})$ to $(B, B, 0)$. By the "moreover" part of Lemma 3.13 M^n satisfies UCV. Also, using the absolute continuity of Z^n we have that $[M^n, Z^n] = 0$ a.s. and, therefore, Corollary 3.16 is applicable and we deduce the joint convergence

$$\left(Z^n, M^n, \int_0^\cdot M^n_s \otimes \mathrm{d}Z^n_s\right) \to \left(B, B, \int_0^\cdot B_s \otimes \mathrm{d}B_s + \langle B, B\rangle\right).$$

It remains to study the term $\int_0^\cdot R^n_s \otimes \mathrm{d}Z^n_s$. We claim that for all $T > 0$,

$$\lim_{n\to\infty} \mathbb{E}\left[\sup_{t\le T}\left|\int_0^t \left(R^n_s + n^{-1/2}\Phi_{n^{-1}}(X_{ns})\right) \otimes \mathrm{d}Z^n_s\right|\right] = 0. \tag{3.13}$$

Indeed, first note that

$$R^n_s - R^{n^{-1},n}_s = M^{n^{-1},n}_s - M^n_s.$$

As Z^n is absolutely continuous, integration by parts implies that

$$\begin{aligned} &\int_0^t \left(R^n_s - R^{n^{-1},n}_s\right) \otimes \mathrm{d}Z^n_s \\ &\quad = \int_0^t Z^n_s \otimes \mathrm{d}\left(M^n_s - M^{n^{-1},n}_s\right) - \left(M^n_t - M^{n^{-1},n}_t\right) \otimes Z^n_t. \end{aligned} \tag{3.14}$$

But

$$\mathbb{E}\Big[\sup_{t\le T}\left|M^n_t - M^{n^{-1},n}_t\right|^2\Big] \lesssim \|\Phi - \Phi_{n^{-1}}\|^2_{\mathcal{H}^1} \to 0,$$

and by equation (3.7) we know that $\sup_n \mathbb{E}[\sup_{t\le T}|Z^n_t|^2] \lesssim 1$. Hence, the second term on the right-hand side can be controlled with the Cauchy–Schwarz inequality. Using the Burkholder–Davis–Gundy inequality and the Cauchy–Schwarz inequality, we bound the integral term in (3.14) as follows:

$$\begin{aligned} &\mathbb{E}\left[\sup_{t\le T}\left|\int_0^t Z^n_s \otimes \mathrm{d}\left(M^n_s - M^{n^{-1},n}_s\right)\right|\right] \\ &\quad \lesssim \mathbb{E}\left[\left|\int_0^T |Z^n_s|^2 \,\mathrm{d}\left|\left[M^n - M^{n^{-1},n}\right]\right|_s\right|^{1/2}\right] \\ &\quad \lesssim \mathbb{E}\Big[\sup_{t\le T}|Z^n_t| \times \left|\left[M^n - M^{n^{-1},n}\right]\right|_T^{1/2}\Big] \\ &\quad \lesssim \mathbb{E}\Big[\sup_{t\le T}|Z^n_t|^2\Big]^{1/2} \mathbb{E}\Big[\sup_{t\le T}\left|M^n_t - M^{n^{-1},n}_t\right|^2\Big]^{1/2} \to 0, \end{aligned}$$

where in the last line we also used the "inverse direction" of the Burkholder–Davis–Gundy inequality. Therefore,

$$\mathbb{E}\Big[\sup_{t\le T}\Big|\int_0^t \big(R_s^n - R_s^{n^{-1},n}\big)\otimes \mathrm{d}Z_s^n\Big|\Big] \to 0.$$

The remaining term in (3.13) involves only the continuous finite variation process $R_s^{n^{-1},n} + n^{-1/2}\Phi_{n^{-1}}(X_{ns})$, so that we can apply Lemma A.1 to obtain

$$\begin{aligned}
&\limsup_{n\to\infty}\mathbb{E}\Big[\sup_{t\le T}\Big|\int_0^t \big(R_s^{n^{-1},n} + n^{-1/2}\Phi_{n^{-1}}(X_{ns})\big)\otimes \mathrm{d}Z_s^n\Big|\Big]\\
&\quad= \limsup_{n\to\infty}\mathbb{E}\Big[\sup_{t\le T}\Big|\frac{1}{\sqrt{n}}\int_0^{nt}\big(R_{n^{-1}s}^{n^{-1},n} + n^{-1/2}\Phi_{n^{-1}}(X_s)\big)\otimes F(X_s)\,\mathrm{d}s\Big|\Big]\\
&\quad\lesssim \limsup_{n\to\infty}\mathbb{E}\Big[\sup_{t\le nT}|R_{n^{-1}t}^{n^{-1},n} + n^{-1/2}\Phi_{n^{-1}}(X_t))\,\mathrm{d}s|^2\Big]^{1/2} T^{1/2}\|F\|_{\mathcal{H}^{-1}}.
\end{aligned}$$

To bound the expectation on the right-hand side, note that

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{t\le T}\big|R_t^{n^{-1},n} + n^{-1/2}\Phi_{n^{-1}}(X_{nt})\big|^2\Big]\\
&\quad\lesssim \mathbb{E}\big[\big|n^{-1/2}\Phi_{n^{-1}}(X_0)\big|^2\big] + \mathbb{E}\Big[\sup_{t\le T}\Big|n^{-1/2}\int_0^{nt} n^{-1}\Phi_{n^{-1}}(X_s)\,\mathrm{d}s\Big|^2\Big]\\
&\quad\lesssim n^{-1}\|\Phi_{n^{-1}}\|_{L^2(\pi)}^2 + T^2 n^{-1}\|\Phi_{n^{-1}}\|_{L^2(\pi)}^2\\
&\quad= (1+T^2)n^{-1}\|\Phi_{n^{-1}}\|_{L^2(\pi)}^2,
\end{aligned}$$

and since according to assumption (3.1) the right-hand side vanishes for $n\to\infty$, we deduce (3.13). Therefore, it suffices to study the limit of $\int_0^t n^{-1/2}\Phi_{n^{-1}}(X_{ns})\otimes \mathrm{d}Z_s^n = n^{-1}\int_0^{nt}\Phi_{n^{-1}}(X_{ns})\otimes F(X_{ns})\,\mathrm{d}s$. Let $\lambda>0$, then

$$\begin{aligned}
\mathbb{E}\Big[\sup_{t\le T}\Big|n^{-1}\int_0^{nt}(\Phi_{n^{-1}}-\Phi_\lambda)(X_{ns})\otimes F(X_{ns})\,\mathrm{d}s\Big|\Big] &\le T E_\pi\big[\big|(\Phi_{n^{-1}}-\Phi_\lambda)\otimes F\big|\big]\\
&\le T\|\Phi_{n^{-1}}-\Phi_\lambda\|_{\mathcal{H}^1}\|F\|_{\mathcal{H}^{-1}},
\end{aligned}$$

and by assumption the right-hand side converges to $T\|\Phi-\Phi_\lambda\|_{\mathcal{H}^1}\|F\|_{\mathcal{H}^{-1}}$, which goes to zero for $\lambda\to 0$. Moreover, by the ergodic theorem the term $n^{-1}\int_0^{nt}\Phi_\lambda(X_{ns})\otimes F(X_{ns})\,\mathrm{d}s$ converges almost surely and in $L^1(\mathbb{P})$ to $tE_\pi[\Phi_\lambda\otimes F]$. By Lemma A.3 in the Appendix, this convergence is even uniform in $t\in[0,T]$: to get the required uniform integrability note that

$$\sup_{t\in[0,T]}\Big|n^{-1}\int_0^{nt}\Phi_\lambda(X_{ns})\otimes F(X_{ns})\,\mathrm{d}s\Big| \le n^{-1}\int_0^{nT}\big|\Phi_\lambda(X_{ns})\otimes F(X_{ns})\big|\,\mathrm{d}s,$$

and the right-hand side converges in L^1 by the ergodic theorem. Now it suffices to send $\lambda\to 0$ to deduce that $t\mapsto \int_0^t R_s^n\otimes \mathrm{d}Z_s^n$ converges to the deterministic limit $t\mapsto t\lim_{\lambda\to 0}E_\pi[\Phi_\lambda\otimes F]$ in probability in $C(\mathbb{R}_+,\mathbb{R}^{d\times d})$. Consequently,

$$\begin{aligned}
(Z^n,\mathbb{Z}^n) &\to \Big(B, \int_0^t B_s\otimes \mathrm{d}B_s + \langle B,B\rangle_t - t\lim_{\lambda\to 0}E_\pi[\Phi_\lambda\otimes F]\Big)\\
&= \Big(B, \int_0^t B_s\otimes\circ\,\mathrm{d}B_s + \frac{1}{2}\langle B,B\rangle_t - t\lim_{\lambda\to 0}E_\pi[\Phi_\lambda\otimes F]\Big),
\end{aligned}$$

and finally we have

$$\lim_{\lambda\to 0} E_\pi[\Phi_\lambda \otimes F] = \lim_{\lambda\to 0} E_\pi\big[\Phi_\lambda \otimes (\lambda - \mathcal{L})\Phi_\lambda\big] = \lim_{\lambda\to 0} E_\pi\big[\Phi_\lambda \otimes (-\mathcal{L})\Phi_\lambda\big]$$

because $\sqrt{\lambda}\Phi_\lambda \to 0$ in $L^2(\pi)$. The limit on the left-hand side exists because Φ_λ converges in $\mathcal{H}^1$ and $F \in \mathcal{H}^{-1}$, and thus also the limit on the right-hand side exists. Moreover, $\frac{1}{2}\langle B, B\rangle_t = t \lim_{\lambda\to 0} E_\pi[\Phi_\lambda \otimes (-\mathcal{L}_S)\Phi_\lambda]$ and since $\mathcal{L} - \mathcal{L}_S = \mathcal{L}_A$ we get the claimed form $\Gamma = \lim_{\lambda\to 0} E_\pi[\Phi_\lambda \otimes \mathcal{L}_A\Phi_\lambda]$ (and in particular this limit exists). □

4. Applications. To illustrate the applicability of our results, we derive here scaling limits in the rough path topology for three classes of models, random walks with random conductances, Ornstein–Uhlenbeck process with divergence free drift, and diffusions with periodic coefficients.

4.1. *Random walks with random conductances.* We place ourselves in the setting of Chapter 3.1 of [19] or [25]. Namely, let

$$\eta = \big\{\eta\big(\{x, y\}\big) = \eta\big(\{y, x\}\big) : x, y \in \mathbb{Z}^d, |x - y| = 1\big\}$$

be a set of numbers with $0 < c \le \eta(\{x, y\}) \le C$ for all x, y and let us write X^η for the continuous time random walk in $\mathbb{Z}^d$ with $X^\eta_0 = 0$ and that jumps from x to y (resp., from y to x) with rate $\eta(\{x, y\})$. Since the rates are bounded from above this random walk exists for all times. We interpret $\eta(\{x, y\})$ as the *conductance* on the bond $\{x, y\}$. To simplify notation, we will write

$$\eta(x, y) = \eta(y, x) = \eta\big(\{x, y\}\big)$$

from now on. We are interested in the situation where $(\eta(\{x, y\}))_{|x-y|=1}$ is an i.i.d. family of random variables (and each $\eta(x, y)$ still takes values in $[c, C]$).

4.1.1. *Scaling limit for the Itô rough path.* Let us write π for the distribution of η and write $X^\eta_{t-} = \lim_{s\uparrow t} X^\eta_s$ and then

$$\mathbb{X}^\eta_{s,t} = \int_s^t X^\eta_{s,r-} \otimes \mathrm{d}X^\eta_r.$$

We also define

$$X^{\eta,n}_t = n^{-1/2} X^\eta_{nt}, \qquad \mathbb{X}^{\eta,n}_{s,t} = \int_s^t X^{\eta,n}_{s,r-} \otimes \mathrm{d}X^{\eta,n}_r.$$

Our aim is to show an invariance principle in the rough path topology for $(X^{\eta,n}, \mathbb{X}^{\eta,n})$ under the *annealed measure*

$$\int \mathbb{E}\big[f\big(X^\eta\big)\big]\pi(\mathrm{d}\eta).$$

The corresponding annealed invariance principle for X^η in the Skorohod topology is established in Chapter 3.1 of [19]. The approach there is based on writing X^η as an additive functional of a certain Markov process plus a martingale, and on applying Lemma 3.2 to the additive functional. The Markov process is the “environment as seen from the walker”: For $x \in \mathbb{Z}^d$ let us write

$$\tau_x \eta(y, z) = \eta(y + x, z + x),$$

and then we define

$$\eta_t := \tau_{X^\eta_t} \eta,$$

which is a càdlàg process with values in the compact space $[c, C]^{E^d}$ equipped with the product topology, where $E^d = \{\{x, y\} : x, y \in \mathbb{Z}^d, |x - y| = 1\}$ are the bonds in $\mathbb{Z}^d$. We write

$$\mathcal{F}_t = \sigma\big(X_s^\eta \vee \eta : s \le t\big),$$

so that (η_t) is adapted to $(\mathcal{F}_t)$. In the following, all martingales are with respect to $(\mathcal{F}_t)$ and the annealed measure, unless explicitly stated otherwise.

LEMMA 4.1 ([19], Lemma 3.1). *The process $(\eta_t)_{t\ge 0}$ is Markovian with respect to $(\mathcal{F}_t)$, with generator*

$$\mathcal{L}F(\eta) = \sum_{\substack{y \in \mathbb{Z}^d: \\ |y|=1}} \eta(0, y)\big(F(\tau_y \eta) - F(\eta)\big)$$

and with reversible and ergodic invariant distribution π.

In Lemma 3.1 of [19], the filtration with respect to which the Markov property holds is not specified, but (a slight modification of) their proof shows that we can take $(\mathcal{F}_t)$ and not just the canonical filtration of (η_t).

Let us define the *local drift* $F : [c, C]^{E^d} \to \mathbb{R}^d$ by

$$F(\eta) = \sum_{|y|=1} y\eta(0, y).$$

It is shown on page 86 of [19] that there exists a càdlàg martingale $(N_t)_{t\ge 0}$ such that

$$X_t^\eta = N_t + \int_0^t F(\eta_s)\,\mathrm{d}s =: N_t + Z_t, \tag{4.1}$$

and, therefore, $X_t^{\eta,n} = N_t^n + Z_t^n$ with the obvious definition of the rescaled processes N^n and Z^n. The idea is now to apply the invariance principle for additive functionals to Z^n and to apply the martingale central limit theorem to N^n. Recall that (η_t) is reversible, so by the discussion in Chapter 2.7.1 in [19] we have $F \in L^2(\pi) \cap \mathcal{H}^{-1}$ and the assumptions of Theorem 3.3 are satisfied. Of course, we also have to understand the joint convergence of (N^n, Z^n), and for that purpose on page 88 of [19] the predictable quadratic covariation between N^n and the martingale $M^{\lambda,n}$ from the decomposition of Lemma 3.12 is derived, namely for $a, b \in \mathbb{R}$,

$$\langle aN^n + bM^{n,\lambda}, aN^n + bM^{n,\lambda}\rangle_t \tag{4.2}$$

$$= \sum_{|y|=1} \frac{1}{n} \int_0^{nt} \eta_s(0, y)\big(ay + b\big(\Phi_\lambda(\tau_y \eta_s) - \Phi_\lambda(\eta_s)\big)\big)^{\otimes 2}\,\mathrm{d}s \tag{4.3}$$

A simple adaptation of Theorem 3.2 in [19] now leads to the following.

LEMMA 4.2. *Under the annealed measure the pair (N^n, Z^n) converges in distribution in the Skorohod topology on $D(\mathbb{R}_+, \mathbb{R}^{2d})$ to a $2d$-dimensional Brownian motion (B^N, B^Z) such that for $a, b \in \mathbb{R}$,*

$$\begin{aligned}&\langle aB^N + bB^Z, aB^N + bB^Z\rangle_t \\ &\quad = t \lim_{\lambda\to 0} \sum_{|y|=1} E_\pi\big[\eta(0, y)\big(ay + b\big(\Phi_\lambda(\tau_y \eta) - \Phi_\lambda(\eta)\big)\big)^{\otimes 2}\big].\end{aligned} \tag{4.4}$$

Moreover, the sequence of processes (N^n) satisfies the UCV condition.

Combining this result with Theorem 3.3, we easily obtain the following convergence in rough path topology.

THEOREM 4.3. *The process $(X^{\eta,n}, \mathbb{X}^{\eta,n})$ converges in distribution in the p-variation rough path topology to*

$$\left(B, \left(\int_0^t B_s \otimes \mathrm{d}B_s + \Gamma t\right)_{t\geq 0}\right),$$

where B is a Brownian motion with covariance

$$\langle B, B\rangle_t = t \lim_{\lambda\to 0} \sum_{|y|=1} E_\pi\big[\eta(0,y)\big(y + \big(\Phi_\lambda(\tau_y\eta) - \Phi_\lambda(\eta)\big)\big)^{\otimes 2}\big],$$

and where for the unit matrix I_d and the vector $e_1 = (1, 0, \dots, 0) \in \mathbb{Z}^d$,

$$\Gamma = \frac{1}{2}\langle B, B\rangle_1 - E_\pi\big[\eta(0, e_1)\big] I_d.$$

PROOF. Using (4.2) together with the arguments from the proof of Corollary 3.17, it is not hard to strengthen Lemma 4.2 to obtain the joint convergence

$$\big(N^n, Z^n, \mathbb{Z}^n_{0,\cdot}\big) \longrightarrow \left(B^N, B^Z, \int_0^\cdot B^Z_s \otimes \mathrm{d}B^Z_s + \frac{1}{2}\langle B^Z, B^Z\rangle\right).$$

Since the limit is continuous the triple is even C-tight and, therefore, also $X^{\eta,n} = N^n + Z^n$ converges in distribution in the Skorohod topology to $B = B^Z + B^N$, and the convergence is jointly with $(N^n, Z^n, \mathbb{Z}^n_{0,\cdot})$. The iterated integrals of $X^{\eta,n}$ are given by

$$\mathbb{X}^{\eta,n}_{0,t} = \int_0^t X^{\eta,n}_{s-} \otimes \mathrm{d}N^n_s + \int_0^t N^n_{s-} \otimes \mathrm{d}Z^n_s + \mathbb{Z}^n_{0,t}. \tag{4.5}$$

Recall from Lemma 4.2 that N^n satisfies the UCV property. Since Z^n is continuous and of finite variation, we get from Theorem 3.15 and Corollary 3.16 the joint convergence

$$\begin{aligned}
&\left(N^n, Z^n, \mathbb{Z}^n_{0,\cdot}, X^{\eta,n}, \int_0^\cdot X^n_{s-} \otimes \mathrm{d}N^n_s, \int_0^\cdot N^n_{s-} \otimes \mathrm{d}Z^n_s\right) \\
&\quad \to \left(B^N, B^Z, \int_0^\cdot B^Z_s \otimes \mathrm{d}B^Z_s + \frac{1}{2}\langle B^Z, B^Z\rangle, B,\right. \\
&\qquad \left.\int_0^\cdot B_s \otimes \mathrm{d}B^N_s, \int_0^\cdot B^N_s \otimes \mathrm{d}B^Z_s + \langle B^N, B^Z\rangle\right).
\end{aligned}$$

Since all the limiting processes are continuous the tuple is C-tight and the joint convergence extends to sums of the entries, so from (4.5) we get

$$\begin{aligned}
\big(X^{\eta,n}, \mathbb{X}^{\eta,n}_{0,\cdot}\big) &\to \left(B, \int_0^\cdot B_s \otimes \mathrm{d}B_s + \frac{1}{2}\langle B^Z, B^Z\rangle + \langle B^N, B^Z\rangle\right) \\
&= \left(B, \int_0^\cdot B_s \otimes \mathrm{d}B_s + \frac{1}{2}\langle B, B\rangle - \frac{1}{2}\langle B^N, B^N\rangle\right)
\end{aligned}$$

and by (4.4) the last term on the right-hand side is given by

$$\begin{aligned}
-\frac{1}{2}\langle B^N, B^N\rangle_t &= -\frac{t}{2} \sum_{|y|=1} E_\pi\big[\eta(0,y) y \otimes y\big] \\
&= -\frac{t}{2} E_\pi\big[\eta(0, e_1)\big] \sum_{|y|=1} y \otimes y \\
&= -t E_\pi\big[\eta(0, e_1)\big] I_d.
\end{aligned}$$

To complete the proof, it remains to show tightness of the p-variation. Since (3.1) holds in the reversible case (see [19], Section 2.7.1) Theorem 3.3 implies that $(\|Z^n\|_{p,[0,T]} + \|\mathbb{Z}^n\|_{p/2,[0,T]})_n$ is tight. For the first level of the rough path, we have $\|X^{\eta,n}\|_{p,[0,T]} \leq \|N_n\|_{p,[0,T]} + \|Z^n\|_{p,[0,T]}$, and $\mathbb{E}[\|N^n\|^2_{p,[0,T]}] \lesssim \mathbb{E}[\langle N^n\rangle_T] \lesssim 1$ by Theorem A.2 together with (4.2), and we already know that $\|Z^n\|_{p,[0,T]}$ is tight. From (4.5), we get

$$\|\mathbb{X}^{\eta,n}\|_{p/2,[0,T]} \leq \left\| \left(\int_s^t X^{\eta,n}_{r,s-} \otimes \mathrm{d}N^n_r \right)_{0\leq s\leq t\leq T} \right\|_{p/2,[0,T]} + \left\| \left(\int_s^t N^n_{r,s} \otimes \mathrm{d}Z^n_r \right)_{0\leq s\leq t\leq T} \right\|_{p/2,[0,T]} + \|\mathbb{Z}^n\|_{p/2,[0,T]}.$$

We apply Proposition 3.8 for the first term on the right-hand side and obtain

$$\mathbb{E}\left[\left\| \left(\int_s^t X^{\eta,n}_{r,s-} \otimes \mathrm{d}N^n_r \right)_{0\leq s\leq t\leq T} \right\|^{1-\varepsilon}_{p/2,[0,T]} \right] \lesssim \left(1 + \mathbb{E}[\|X^{\eta,n}\|^2_{p',[0,T]}]^{1/2}\right) \times \left(1 + \mathbb{E}[\langle N^n\rangle_T]^{1/2}\right) \lesssim 1,$$

where $p' \in (2, p)$. The second term on the right-hand side can be controlled via integration by parts and a similar application of Proposition 3.8. And we already know that $(\|\mathbb{Z}^n\|_{p/2,[0,T]})_n$ is tight. Hence we get the tightness of $(\|\mathbb{X}^{\eta,n}\|_{p/2,[0,T]})_n$, and this concludes the proof. □

REMARK 4.4. We did not really use that the conductances are i.i.d., and the same proof works if they are only ergodic with respect to the shifts on $\mathbb{Z}^d$. In that case, the correction Γ of Theorem 4.3 is given by

$$\Gamma = \frac{1}{2}\langle B, B\rangle_1 - \mathrm{diag}\big(E_\pi[\eta(0, e_1)], \dots, E_\pi[\eta(0, e_d)]\big),$$

where diag$(\dots)$ is a diagonal matrix with the respective entries on the diagonal. In the i.i.d. setting and for $d > 2$, we expect that it is possible to get stronger results (Hölder topology instead of p-variation, speed of convergence, convergence under the quenched measure) by using the spectral gap result of [13].

4.1.2. *Scaling limit for the Stratonovich rough path.* In our discrete setting of the random walk in random environment, it seems natural to consider the Itô iterated integrals $\int_s^t X^{\eta,n}_{s,r-} \otimes \mathrm{d}X^{\eta,n}_r$. But of course this is not the only option, and we might also turn X^η into a continuous path by connecting the jumps piecewise linearly, as it is often done for Donsker's invariance principle. More precisely, if σ^n_k, $k = 1, 2, \dots$ are the jump times of the process $X^{\eta,n}$, then we set

$$\bar{X}^{\eta,n}_t := X^{\eta,n}_{\sigma^n_k} + \frac{t - \sigma^n_k}{\sigma^n_{k+1} - \sigma^n_k} X^{\eta,n}_{\sigma^n_k,\sigma^n_{k+1}}$$

for $\sigma^n_k \leq t < \sigma^n_{k+1}$. We then define $\bar{\mathbb{X}}^{\eta,n}_{s,t} = \int_s^t \bar{X}^{\eta,n}_{s,r} \otimes \mathrm{d}\bar{X}^{\eta,n}_r$, and as usual $\bar{\mathbb{X}}^{\eta,n}_t = \bar{\mathbb{X}}^{\eta,n}_{0,t}$. Note that $\sup_{t\geq 0} |\bar{X}^{\eta,n}_t - X^{\eta,n}_t| \leq n^{-1/2}$ and, therefore, $\bar{X}^{\eta,n}$ converges to the same Brownian motion as $X^{\eta,n}$. The difference arises only on the level of the iterated integrals: We have

$$\begin{aligned}\int_{\sigma^n_k}^{\sigma^n_{k+1}} \bar{X}^{\eta,n}_{0,r} \otimes \mathrm{d}\bar{X}^{\eta,n}_r &= \frac{1}{2}(\sigma^n_{k+1} - \sigma^n_k) X^{\eta,n}_{\sigma^n_k,\sigma^n_{k+1}} \otimes \frac{X^{\eta,n}_{\sigma^n_k,\sigma^n_{k+1}}}{\sigma^n_{k+1} - \sigma^n_k} + X^{\eta,n}_{\sigma^n_k} \otimes X^{\eta,n}_{\sigma^n_k,\sigma^n_{k+1}} \\ &= \frac{1}{2}\big(X^{\eta,n}_{\sigma^n_k} + X^{\eta,n}_{\sigma^n_{k+1}}\big) \otimes X^{\eta,n}_{\sigma^n_k,\sigma^n_{k+1}}.\end{aligned}$$

Therefore,

$$\begin{aligned}\bar{\mathbb{X}}^{\eta,n}_{\sigma^n_k} - \mathbb{X}^{\eta,n}_{\sigma^n_k} &= \sum_{j=0}^{k-1}\Big(\int_{\sigma^n_j}^{\sigma^n_{j+1}} \bar{X}^{\eta,n}_{0,r} \otimes \mathrm{d}\bar{X}^{\eta,n}_r - X^{\eta,n}_{\sigma^n_j} \otimes X^{\eta,n}_{\sigma^n_j,\sigma^n_{j+1}}\Big)\\ &= \frac{1}{2}\sum_{j=0}^{k-1}(X^{\eta,n}_{\sigma^n_j,\sigma^n_{j+1}})^{\otimes 2}.\end{aligned}$$

Note that the sequence $(\eta^{(k)} := \{\eta(s) : s \in [k, k+1]\})_{k\geq 0}$ is stationary and ergodic with respect to the shift map $\theta(\eta^{(k)})_{k\geq 0} = (\eta^{(k+1)})_{k\geq 0}$. Using the ergodic theorem and the fact that $\sigma^n_k = n\sigma^1_k$, for any fixed $t > 0$, we have

$$\begin{aligned}\frac{1}{2}\sum_{k:\sigma^n_k\leq t}(X^{\eta,n}_{\sigma^n_k,\sigma^n_{k+1}})^{\otimes 2} = \frac{1}{2n}\sum_{0<s\leq nt}(X^\eta_{s-,s})^{\otimes 2}\\ &= \frac{\lfloor nt\rfloor}{2n}\frac{1}{\lfloor nt\rfloor}\sum_{k=0}^{\lfloor nt\rfloor-1}\Big(\sum_{k\leq s<k+1}\Big(\sum_{|y|=1} y1_{\eta(s)=\tau_y\eta(s-)}\Big)^{\otimes 2}\Big)\\ &\quad + \frac{1}{2n}\sum_{\lfloor nt\rfloor\leq s<nt}(X^\eta_{s-,s})^{\otimes 2}\\ &= \frac{\lfloor nt\rfloor}{2n}\frac{1}{\lfloor nt\rfloor}\sum_{k=0}^{\lfloor nt\rfloor-1}\Psi(\theta^k((\eta^{(j)})_{j\geq 0})) + O\Big(\frac{1}{n}\Big)\\ &\to \frac{t}{2}\mathbb{E}_\pi[\Psi((\eta^{(k)})_{k\geq 0})]\\ &= \frac{t}{2}\mathbb{E}_\pi\Big[\sum_{0<j:\sigma^1_j\leq 1}(X^\eta_{\sigma^1_{j-1},\sigma^1_j})^{\otimes 2}\Big],\end{aligned}$$

where the convergence as $n\to\infty$ is in $L^1(\mathbb{P}_\pi)$ (easy to see) and $\mathbb{P}_\pi$ almost surely (to be justified below), and where $\theta^0 = \text{i.d.}$, $\theta^k := \theta^{k-1}\theta$ for $k > 1$, and

$$\Psi((\eta^{(k)})_{k\geq 0}) = \Psi(\eta^{(0)}) := \sum_{0<s\leq 1}\Big(\sum_{|y|=1} y\mathbb{1}_{\eta(s)=\tau_y\eta(s-)}\Big)^{\otimes 2}. \tag{4.6}$$

To see that the $O(\frac{1}{n})$ term converges $\mathbb{P}_\pi$ almost surely to zero, note that by stationarity and since the norm of $t\mapsto\sum_{0<s\leq t}(\sum_{|y|=1} y\mathbb{1}_{\eta(s)=\tau_y\eta(s-)})^{\otimes 2}\in\mathbb{R}^{d\times d}$ is increasing in t on the diagonal terms it follows that

$$\mathbb{E}_\pi\Big[\Big|\frac{1}{2n}\sum_{\lfloor nt\rfloor\leq s<nt}(X^\eta_{s-,s})^{\otimes 2}\Big|^2\Big] \lesssim \frac{1}{n^2}\mathbb{E}_\pi\Big[\Big|\sum_{0\leq s<1}(X^\eta_{s-,s})^{\otimes 2}\Big|^2\Big].$$

Our jump rates are uniformly bounded and the size of each jump is bounded by 1 and, therefore, the expectation on the right-hand side is finite. Consequently,

$$\mathbb{E}_\pi\Big[\sum_n\Big|\frac{1}{2n}\sum_{\lfloor nt\rfloor\leq s<nt}(X^\eta_{s-,s})^{\otimes 2}\Big|^2\Big] < \infty,$$

and thus the summands converge almost surely to zero.

By Lemma A.3 in the Appendix, the $L^1(\pi)$-convergence holds even uniformly on compact time intervals. Let us compute the limit: Since the additive functional in the decomposition of X^η in (4.1) does not jump we have $X^\eta_{t-,t} = N_{t-,t}$ for all $t > 0$ and, therefore,

$$\mathbb{E}_\pi\Big[\sum_{0<j:\sigma^1_j\le 1} \big(X^\eta_{\sigma^1_{j-1},\sigma^1_j}\big)^{\otimes 2}\Big] = \mathbb{E}_\pi\Big[\sum_{0<t\le 1} \big(X^\eta_{t-,t}\big)^{\otimes 2}\Big] = E\Big[\sum_{0<t\le 1} (N_{t-,t})^{\otimes 2}\Big]$$
$$= \mathbb{E}_\pi\big[[N,N]_1\big] = \mathbb{E}_\pi\big[\langle N,N\rangle_1\big] = \langle B^N, B^N\rangle_1.$$

Therefore,

$$(\bar{\mathbb{X}}^{\eta,n}_t - \mathbb{X}^{\eta,n}_t)_{t\in[0,T]} \to \frac{1}{2}\big(\langle B^N, B^N\rangle_t\big)_{t\in[0,T]} \quad \text{as } n\to\infty, \tag{4.7}$$

and since the left-hand side is increasing in the sense of positive definite matrices, and thus in norm and the convergence is uniform in L^1, it holds even for the 1-variation norm. In the proof of Theorem 4.3, we saw that

$$\frac{1}{2}\langle B^N, B^N\rangle_t = \frac{1}{2}\langle B, B\rangle_t - \Gamma t,$$

so together with (4.7) and the fact that the Stratonovich integral equals $\int_0^t B_s \otimes \circ\, \mathrm{d}B_s = \int_0^t B_s \otimes \mathrm{d}B_s + \frac{1}{2}\langle B, B\rangle_t$, we deduce for the case of linear interpolations that the limit is the Stratonovich Brownian rough path, with no correction.

COROLLARY 4.5. *Let $(\bar{X}^{\eta,n}, \bar{\mathbb{X}}^{\eta,n})$ be the linear interpolation of the path $X^{\eta,n}$ and its corresponding iterated integral defined above. Then $(\bar{X}^{\eta,n}, \bar{\mathbb{X}}^{\eta,n})$ converges in distribution in the p-variation topology to $(B, \int_0^\cdot B_s \otimes \circ\, \mathrm{d}B_s)$, where B is the same Brownian motion as in Theorem 4.3, and $\circ$ denotes Stratonovich integration.*

4.2. *Additive functional of Ornstein–Uhlenbeck process with divergence-free drift.* In this section, we give a simple example of an additive functional of an Ornstein–Uhlenbeck process with divergence free drift with a nonvanishing area anomaly, that is, so that the correction Γ from (3.2) is nonzero. Let $U : \mathbb{R}^2 \to \mathbb{R}$ be given by $U(x) = \frac{1}{2}|x|^2 - \log 2\pi$ and let

$$b(x) = \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix} e^{-U(x)} = Axe^{-U(x)} \quad \text{where } A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Note that b is divergence-free. We define the operator

$$\mathcal{L}f = \Delta f - \nabla U \cdot \nabla f - be^U \cdot \nabla f = e^U \nabla \cdot \big(e^{-U}\nabla f\big) - be^U \cdot \nabla f,$$

which is the generator of the Ornstein–Uhlenbeck process

$$\mathrm{d}X_t = -\nabla U(X_t)\,\mathrm{d}t - b(X_t)e^{U(X_t)}\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t = -(I + A)X_t\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t.$$

One can check that $\pi(\mathrm{d}x) = e^{-U(x)}\,\mathrm{d}x$ is invariant for X. Indeed, if $f \in C^2_b(\mathbb{R}^d)$, then integration by parts yields

$$\int \mathcal{L}f e^{-U}\,\mathrm{d}x = \int \big(\nabla\cdot\big(e^{-U}\nabla f\big) - b\cdot\nabla f\big)\,\mathrm{d}x$$
$$= \int \big(\nabla 1\cdot\big(e^{-U}\nabla f\big) + f\nabla\cdot b\big)\,\mathrm{d}x = 0,$$

because $\nabla\cdot b = 0$. We consider X started in the invariant measure and we are interested in the rough path limit of

$$Z^n_t = \frac{1}{\sqrt{n}}\int_0^{nt} X_s\,\mathrm{d}s.$$

For that purpose, let $F(x) = x$. Since X has a spectral gap, it converges exponentially fast to its invariant measure and we can directly solve the Poisson equation $-\mathcal{L}\Phi = F$, that is, there is no need to first consider the resolvent equation $(\lambda - \mathcal{L})\Phi_\lambda = F$ and then send $\lambda \to 0$. To compute the explicit solution to the Poisson equation, we use the standard ansatz

$$\Phi(x) = Cx = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} C_{11}x_1 + C_{12}x_2 \\ C_{21}x_1 + C_{22}x_2 \end{pmatrix}.$$

Write $\Phi(x) = \left(\begin{smallmatrix} \Phi_1(x) \\ \Phi_2(x) \end{smallmatrix}\right)$, then $\nabla\Phi(x) = \left(\begin{smallmatrix} \nabla\Phi_1(x) \\ \nabla\Phi_2(x) \end{smallmatrix}\right) = \left(\begin{smallmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{smallmatrix}\right) = C$ for $\Phi(x) = Cx$. Hence, for $j = 1, 2$,

$$-\mathcal{L}\Phi_j(x) = \big((I + A)x - \nabla\big) \cdot \nabla\Phi_j(x) = (I + A)x \cdot C_{j,\cdot},$$

or more compactly

$$-\mathcal{L}\Phi(x) = C(I + A)x.$$

The equation $-\mathcal{L}\Phi(x) = F(x) = x$ then yields

$$C(I + A) = I.$$

Since $A^2 = -I$, this implies

$$C = \frac{1}{2}(I - A) = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

The Ornstein–Uhlenbeck operator has a spectral gap, so condition (3.1) is satisfied; see [19], Theorem 2.18. By Theorem 3.3, we get the convergence

$$(Z^n, \mathbb{Z}^n) \longrightarrow \left(B, \left(\int_0^t B_s \circ \mathrm{d}B_s + \Gamma t\right)_{t \in [0,T]}\right),$$

in distribution in p-variation, where B is a Brownian motion with covariance

$$\langle B^i, B^j \rangle_t = 2t E_\pi\big[\Phi^i(-\mathcal{L}_S)\Phi^j\big] = 2t E_\pi\big[(Cx)^i(Cx)^j\big] = 2t(C_{i1}C_{j1} + C_{i2}C_{j2}) = tI,$$

where we used that under π the coordinates (x^1, x^2) are independent standard Gaussians, and where

$$\begin{aligned}
\Gamma_{ij} &= E_\pi\big[\Phi^i \mathcal{L}_A \Phi^j\big] = E_\pi\big[(Cx)^i(-CAx)^j\big] = (C_{i1} \quad C_{i2})\, E_\pi\big[x(Ax)^T\big] \begin{pmatrix} C_{j1} \\ C_{j2} \end{pmatrix} \\
&= (C_{i1} \quad C_{i2})\, E_\pi\left[\begin{pmatrix} -x_1x_2 & x_1^2 \\ -x_2^2 & x_1x_2 \end{pmatrix}\right] \begin{pmatrix} C_{j1} \\ C_{j2} \end{pmatrix} = -C_{i2}C_{j1} + C_{i1}C_{j2} \\
&= \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}_{ij}.
\end{aligned}$$

In particular, we see a nontrivial correction to the iterated integrals of B.

4.3. *Diffusions with periodic coefficients.* Consider a smooth $\mathbb{Z}^d$-periodic function $a : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ and $L = \nabla \cdot (a\nabla)$, that is,

$$Lf(x) = \sum_{i,j=1}^d \big(a_{ij}(x)\partial_i\partial_j f(x) + \partial_i a_{ij}(x)\partial_j f(x)\big).$$

We assume that the symmetric part of a is uniformly elliptic (a itself is not necessarily symmetric). Then there is a unique diffusion process associated with L, with coefficients

$$\mathrm{d}X_t^j = \sum_{i=1}^d \partial_i a_{ij}(X_t)\,\mathrm{d}t + \sqrt{2}\sum_{i=1}^d \sigma_{ji}(X_t)\,\mathrm{d}W_t^i,$$

where

$$\sigma = \sqrt{a^S}, \qquad a^S = \frac{1}{2}(a + a^*), \qquad a^A = \frac{1}{2}(a - a^*).$$

To simplify notation, we write

$$b_j = \sum_{i=1}^d \partial_i a_{ij} = \nabla \cdot a_{\cdot j}$$

so

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sqrt{2}\sigma(X_t)\,\mathrm{d}W_t.$$

We assume that X_0 is uniformly distributed on $[-\frac{1}{2}, \frac{1}{2}]^d$ (just so that the Markov process Y below is stationary) and we want to understand the large scale behavior of X in rough path topology, for which we will derive the following result.

THEOREM 4.6. *Let*

$$X_t^n = n^{-1/2} X_{nt}, \quad t \in [0, T].$$

Then the following convergence holds in p-variation rough path topology:

$$\left(X_t^n, \int_0^t X_s^n \otimes \circ\,\mathrm{d}X_s^n\right) \to \left(B_t, \int_0^t B_s \otimes \circ\,\mathrm{d}B_s + t\Gamma\right),$$

where $\Gamma := \int (\nabla\Phi^i \cdot (-a^A)\nabla\Phi^j)_{i,j=1,\dots,d}\,\mathrm{d}x$, Φ *solves the Poisson equation*

$$-\nabla \cdot (a\nabla\Phi) = b$$

and B is a Brownian motion with quadratic variation

$$\langle B^i, B^j\rangle_t = 2t \int (\nabla\Phi^i + e_i) \cdot a^S (\nabla\Phi^j + e_j)\,\mathrm{d}x,$$

for the standard basis $(e_1, \dots, e_d)$ of $\mathbb{R}^d$. For the Itô rough path, we see an additional correction:

$$\left(X_t^n, \int_0^t X_s^n \otimes \mathrm{d}X_s^n\right) \to \left(B_t, \int_0^t B_s \otimes \mathrm{d}B_s + \frac{1}{2}\langle B, B\rangle_t - t\int a^S\,\mathrm{d}x + t\Gamma\right).$$

REMARK 4.7. The convergence of the Stratonovich rough path was previously shown by Lejay and Lyons [21], Proposition 6. Their proof uses the fact that we control all moments of X, from where the required tightness in Hölder topology (which is stronger than p-variation) easily follows via a Kolmogorov continuity criterion for rough paths, and there is no need to invoke a result like Proposition 3.8. Our general approach has the advantage that it applies to a much wider class of models and that we can apply it without having to do additional estimations, but in this special case it gives a weaker result.

We now sketch the proof of the claimed convergence. Let us rewrite

$$Y_t = X_t \bmod \mathbb{Z}^d.$$

Since the coefficients of X are periodic, Y is a Markov process with values in $\mathbb{T}^d = (\mathbb{R}/\mathbb{Z})^d$, with generator $\mathcal{L}$ given by the same expression as L, except now it acts on $C^2(\mathbb{T}^d)$ rather than on $C^2(\mathbb{R}^d)$. The Lebesgue measure on $\mathbb{T}^d$ is invariant for Y, and we have

$$\int_{\mathbb{T}^d} b_j(x)\,\mathrm{d}x = \int_{\mathbb{T}^d} \sum_{i=1}^d \partial_i a_{ij}(x)\,\mathrm{d}x = 0$$

by the periodic boundary conditions.

Therefore, we can write (slightly abusing notation by also considering b, σ etc. as functions on $\mathbb{T}^d$)

$$n^{-1/2}X_{nt} = n^{-1/2}\int_0^{nt} b(Y_s)\,\mathrm{d}s + n^{-1/2}\int_0^{nt} \sqrt{2}\sigma(Y_s)\,\mathrm{d}W_s = Z_t^n + M_t^n,$$

where M^n is a martingale with quadratic variation

$$\langle M^n\rangle_t^{ij} = \frac{2}{n}\int_0^{nt} (\sigma\sigma^*)_{ij}(X_s)\,\mathrm{d}s = \frac{2}{n}\int_0^{nt} a_{ij}^S(X_s)\,\mathrm{d}s = \frac{2}{n}\int_0^{nt} a_{ij}^S(Y_s)\,\mathrm{d}s,$$

and where Z^n is a functional that we can control with our tools from Section 3. By the uniform ellipticity of a^S together with the Poincaré inequality, we have for all f with $\int f\,\mathrm{d}x = 0$,

$$\int f(-\mathcal{L})f\,\mathrm{d}x = \int \nabla f\cdot a\nabla f\,\mathrm{d}x = \int \nabla f\cdot a^S\nabla f\,\mathrm{d}x \gtrsim \int |\nabla f|^2\,\mathrm{d}x \gtrsim \int f^2\,\mathrm{d}x,$$

that is, $\mathcal{L}$ has a spectral gap and Y is exponentially ergodic. Thus $\langle M^n\rangle_t \to 2t\int a^S(x)\,\mathrm{d}x$, from where we can show with some more work that $M^n \to B^M$ for a d-dimensional Brownian motion with covariance

$$\langle B^M, B^M\rangle_t = 2t\int a^S(x)\,\mathrm{d}x.$$

Since M^n satisfies the UCV condition, the convergence of the lifted path also holds in the p-variation rough path topology for every $p > 2$ by Proposition 3.8.

$$-\mathcal{L}\Phi = -\nabla\cdot(a\nabla\Phi) = b,$$

which is given by $\Phi = \int_0^\infty P_t b\,\mathrm{d}t$, where (P_t) is the semigroup of Y. The time integral converges because $P_t b$ converges exponentially fast to $\int b\,\mathrm{d}x = 0$. Since Y has a spectral gap, the conditions of Theorem 3.3 are satisfied (see [19], Theorem 2.18) and, therefore, $(Z^n, \int_0^\cdot Z_s^n\otimes\mathrm{d}Z_s^n)$ converges to the corrected Stratonovich rough path

$$\Bigl(B^Z, \int_0^\cdot B_s^Z\otimes\circ\,\mathrm{d}B_s^Z + t\int \Phi\otimes\mathcal{L}_A\Phi\,\mathrm{d}x\Bigr),$$

where B^Z is a Brownian motion with covariance

$$\langle B^Z, B^Z\rangle_t^{ij} = 2\int \Phi^i(-\mathcal{L}_S)\Phi^j\,\mathrm{d}x = 2\int \nabla\Phi^i\cdot a^S\nabla\Phi^j\,\mathrm{d}x,$$

and where

$$\Bigl(\int \Phi\otimes\mathcal{L}_A\Phi\,\mathrm{d}x\Bigr)_{ij} = -\int \nabla\Phi^i\cdot\bigl(a^A\nabla\Phi^j\bigr)\,\mathrm{d}x.$$

It remains to understand the quadratic covariation of B^M and B^Z, as well as the cross-integrals $\int Z^n\otimes\mathrm{d}M^n$ and $\int M^n\otimes\mathrm{d}Z^n$. To derive the covariation, note that we get with the solution to the Poisson equation Φ,

$$Z_t^n = \underbrace{\frac{1}{\sqrt{n}}\bigl(\Phi(Y_0)-\Phi(Y_{nt})\bigr)}_{=:R_t^n} + \underbrace{\frac{1}{\sqrt{n}}\int_0^{nt}\sqrt{2}\sum_{j,i=1}^d \partial_j\Phi(X_s)\sigma_{ji}(X_s)\,\mathrm{d}W_s^i}_{=:N_t^n},$$

and the covariation of N^n and M^n is thus given by

$$\langle N^n, M^n\rangle_t^{ij} = \frac{2}{n}\int_0^{nt}\sum_{k,\ell=1}^d \partial_k\Phi^i(X_s)\sigma_{k\ell}(X_s)\sigma_{j\ell}(X_s)\,\mathrm{d}s$$

$$
\begin{aligned}
&= \frac{2}{n}\int_0^{nt} \sum_{k=1}^{d} \partial_k \Phi^i a^S_{kj}(X_s)\,\mathrm{d}s \\
&\to 2t \int \sum_{k=1}^{d} \partial_k \Phi^i a^S_{kj}\,\mathrm{d}x,
\end{aligned}
$$

so that $B = B^Z + B^M$ is a Brownian motion with covariance

$$
\begin{aligned}
\langle B, B\rangle_t &= \langle B^M, B^M\rangle_t + \langle B^Z, B^Z\rangle_t + 2\langle B^Z, B^M\rangle_t \\
&= 2t \int \left(a^S_{ij} + \nabla\Phi^i \cdot a^S \nabla\Phi^j + 2\sum_{k=1}^{d} \partial_k \Phi^i a^S_{kj} \right)_{ij} \mathrm{d}x \\
&= 2t \int \big((\nabla\Phi^i + e_i)\cdot a^S(\nabla\Phi^j + e_j)\big)_{ij}\,\mathrm{d}x.
\end{aligned}
$$

The cross-iterated integrals satisfy according to Theorem 3.15 and Corollary 3.16,

$$
\begin{aligned}
&\int_0^{\cdot} Z^n_s \otimes \mathrm{d}M^n_s \to \int_0^{\cdot} B^Z_s \otimes \mathrm{d}B^M_s, \\
&\int_0^{\cdot} M^n_s \otimes \mathrm{d}Z^n_s \to \int_0^{\cdot} B^M_s \otimes \mathrm{d}B^Z_s + \langle B^M, B^Z\rangle - 0,
\end{aligned}
$$

so that overall

$$
\begin{aligned}
&\left(X^n_t, \int_0^t X^n_s \otimes \mathrm{d}X^n_s \right) \\
&\quad \to \left(B_t, \int_0^t B_s \otimes \mathrm{d}B_s + \frac{1}{2}\langle B^Z, B^Z\rangle_t + \langle B^M, B^Z\rangle_t + t\int \Phi(x)\otimes \mathcal{L}_A\Phi(x)\,\mathrm{d}x \right),
\end{aligned}
$$

and the first part of the correction can be further simplified to

$$
\frac{1}{2}\langle B^Z, B^Z\rangle_t + \langle B^M, B^Z\rangle_t = \frac{1}{2}\big(\langle B, B^Z\rangle_t + \langle B^M, B^Z\rangle_t\big) = \frac{1}{2}\langle B, B\rangle_t - \frac{1}{2}\langle B^M, B^M\rangle_t,
$$

which finally yields the limit

$$
\int_0^t X^n_s \otimes \mathrm{d}X^n_s \to \int_0^t B_s \otimes \mathrm{d}B_s + \frac{1}{2}\langle B, B\rangle_t - t\int a^S\,\mathrm{d}x - t\int \big(\nabla\Phi^i \cdot a^A \nabla\Phi^j\big)_{ij}\,\mathrm{d}x.
$$

Tightness in p-variation follows as in the example of the random conductance model. This proves the first claim of Theorem 4.6, about the limit of the Itô rough path.

To identify the limit of the Stratonovich rough path, we use that $\langle X^n, X^n\rangle = \langle M^n, M^n\rangle$, and thus

$$
\begin{aligned}
\int_0^t X^n_s \otimes \circ\,\mathrm{d}X^n_s &= \int_0^t X^n_s \otimes \mathrm{d}X^n_s + \frac{1}{2}\langle X^n, X^n\rangle_t \\
&= \int_0^t X^n_s \otimes \mathrm{d}X^n_s + \frac{1}{2}\langle M^n, M^n\rangle_t \\
&\to \int_0^t B_s \otimes \mathrm{d}B_s + \frac{1}{2}\langle B, B\rangle_t - \frac{1}{2}\langle B^M, B^M\rangle_t \\
&\qquad + t\int \Phi(x)\otimes \mathcal{L}_A\Phi(x)\,\mathrm{d}x + \frac{1}{2}\langle B^M, B^M\rangle_t \\
&= \int_0^t B_s \otimes \mathrm{d}B_s + \frac{1}{2}\langle B, B\rangle_t + t\int \Phi(x)\otimes \mathcal{L}_A\Phi(x)\,\mathrm{d}x \\
&= \int_0^t B_s \otimes \circ\,\mathrm{d}B_s - t\int \nabla\Phi^i \cdot \big(a^A \nabla\Phi^j\big)\,\mathrm{d}x.
\end{aligned}
$$

REMARK 4.8. An alternative formula for Γ from Theorem 4.6 is given by

$$\Gamma_{ij} = \frac{1}{2}\int_0^\infty \mathbb{E}_U\big[b_j(Y_s)b_i(Y_0) - b_i(Y_s)b_j(Y_0)\big]\,\mathrm{d}s,$$

where $\mathbb{E}_U$ is the expectation corresponding to the law process X so that $X_0 \sim U$, where U is the uniform distribution on $[-\frac{1}{2}, \frac{1}{2}]^d$.

PROOF. Recall the formula for Γ in Theorem 4.6. Integration by parts and the definition of b imply

$$\Gamma_{ij} = \frac{1}{2}\int_{\mathbb{T}^d}\big(\Phi^j(x)b_i(x) - \Phi^i(x)b_j(x)\big)\,\mathrm{d}x.$$

By the presentation $\Phi(x) = \int_0^\infty \mathbb{E}_x[b(X_s)]\,\mathrm{d}s$, the periodicity of b and Fubini's theorem

$$\begin{aligned}\Gamma_{ij} &= \frac{1}{2}\int_{\mathbb{T}^d}\int_0^\infty \mathbb{E}_x\big[b_j(X_s)b_i(X_0) - b_i(X_s)b_j(X_0)\big]\,\mathrm{d}s\,\mathrm{d}x \\ &= \frac{1}{2}\int_{\mathbb{T}^d}\int_0^\infty \mathbb{E}_x\big[b_j(Y_s)b_i(Y_0) - b_i(Y_s)b_j(Y_0)\big]\,\mathrm{d}s\,\mathrm{d}x \\ &= \frac{1}{2}\int_0^\infty \mathbb{E}_U\big[b_j(Y_s)b_i(Y_0) - b_i(Y_s)b_j(Y_0)\big]\,\mathrm{d}s.\end{aligned}$$

□

5. Proof of Proposition 3.8. We write $\|f\|_{p,[s,t]}$ for the p-variation of f restricted to the interval $[s,t]$.

DEFINITION 5.1. A *control function* is a map $c : \Delta_T \to [0,\infty)$ with $c(t,t) = 0$ for all $t \in [0,T]$ and such that $c(s,u) + c(u,t) \le c(s,t)$ for all $0 \le s \le u \le t \le T$.

Observe that if $f : [0,T] \to \mathbb{R}^d$ satisfies $|f_{s,t}|^p \le c(s,t)$ for all $(s,t) \in \Delta_T$, then the p-variation of f is bounded from above by $c(0,T)^{1/p}$. Indeed, we have for any partition π of $[0,T]$,

$$\Big(\sum_{[s,t]\in\pi} |f_{s,t}|^p\Big)^{1/p} \le \Big(\sum_{[s,t]\in\pi} c(s,t)\Big)^{1/p} \le c(0,T)^{1/p}.$$

Conversely, if f is of finite p-variation, then $c(s,t) := \|f\|_{p,[s,t]}^p$ defines a control function because $c(t,t) = |f_{t,t}| = 0$ and

$$\begin{aligned}c(s,u) + c(u,t) &= \sup_{\pi \text{ Part. of } [s,u]} \sum_{[r,v]\in\pi} |f_{r,v}|^p + \sup_{\pi \text{ Part. of } [u,t]} \sum_{[r,v]\in\pi} |f_{r,v}|^p \\ &= \sup_{\substack{\pi \text{ Part. of } [s,t] \\ \text{s.t. } u \in \pi}} \sum_{[r,v]\in\pi} |f_{r,v}|^p \le \sup_{\pi \text{ Part. of } [s,t]} \sum_{[r,v]\in\pi} |f_{r,v}|^p \\ &= c(s,t).\end{aligned}$$

Note also that the sum of two control functions is a control function. Proposition 3.8 directly follows from the next result.

PROPOSITION 5.2. *Let $(Y_t)_{t\in[0,T]}$ be a càdlàg adapted process such that $\|Y\|_{p,[0,T]} < \infty$ and let N be a square-integrable martingale. Set $A_{s,t} := \int_s^t Y_{r-}\,\mathrm{d}N_r - Y_s N_{s,t}$. Then we have for all $p, q > 2$ and all $r > (\frac{1}{p} + \frac{1}{q})^{-1}$:*

$$\|A\|_{r,[0,T]} \lesssim \big(1 + \big|\log \|Y\|_{p,[0,T]}\big|\big)\|Y\|_{p,[0,T]}\big(K^{\frac{1}{q}} + \|N\|_{q,[0,T]}\big), \tag{5.1}$$

where K is a random variable with $\mathbb{E}[K^{\frac{2}{q}}] \lesssim \mathbb{E}[\langle N\rangle_T]$.

REMARK 5.3. For $p < 2$, it follows directly from Young integration estimates that $\|A\|_{1/(1/p+1/r),[0,T]} \lesssim \|Y\|_{p,[0,T]}\|N\|_{r,[0,T]}$ whenever $r > 2$ is such that $1/p + 1/r > 1$.

PROOF. Define the stopping times $\tau_0^n := 0$ and $\tau_{k+1}^n := \inf\{t \geq \tau_k^n : |Y_{\tau_k^n,t}| \geq 2^{-n}\}$ and set

$$Y_t^n := \sum_{k=0}^{\infty} \mathbb{1}_{(\tau_k^n,\tau_{k+1}^n]}(t) Y_{\tau_k^n},$$

such that $\sup_{t\in[0,T]} |Y_{t-} - Y_t^n| \leq 2^{-n}$, where $Y_{t-} := \lim_{s\uparrow t} Y_s$ and $Y_{0-} := Y_0$ and we also write $(Y_-)_t := Y_{t-}$. We have

$$|A_{s,t}| \leq \left|\int_s^t (Y_{r-} - Y_r^n)\,\mathrm{d}N_r\right| + \left|\int_s^t Y_r^n\,\mathrm{d}N_r - Y_s N_{s,t}\right|. \tag{5.2}$$

The first term on the right-hand side is bounded for $q > 2$ and $n \in \mathbb{Z} \setminus \{0\}$ by

$$\left|\int_s^t (Y_{r-} - Y_r^n)\,\mathrm{d}N_r\right| \leq |n| 2^{-n} c(s,t)^{\frac{1}{q}} K^{\frac{1}{q}}, \tag{5.3}$$

where we define

$$K := \sum_{m\in\mathbb{Z}\setminus\{0\}} |m|^{-q} 2^{mq} \left\|\int_0^{\cdot} (Y_{r-} - Y_r^m)\,\mathrm{d}N_r\right\|_{q,[0,T]}^q \tag{5.4}$$

and

$$c(s,t) := \frac{\sum_{m\in\mathbb{Z}\setminus\{0\}} |m|^{-q} 2^{mq} \| \int_0^{\cdot} (Y_{r-} - Y_r^m)\,\mathrm{d}N_r\|_{q,[s,t]}^q}{\sum_{m\in\mathbb{Z}\setminus\{0\}} |m|^{-q} 2^{mq} \| \int_0^{\cdot} (Y_{r-} - Y_r^m)\,\mathrm{d}N_r\|_{q,[0,T]}^q} + \frac{\|Y_-\|_{p,[s,t]}^p + \|Y\|_{p,[s,t]}^p}{\|Y\|_{p,[0,T]}^p} + \frac{\|N\|_{q,[s,t]}^q}{\|N\|_{q,[0,T]}^q}.$$

Note that $\|Y\|_{p,[0,T]}^p = \|Y_-\|_{p,[0,T]}^p$ and, therefore, $c(s,t) \leq c(0,T) = 4$.

To bound the second term in (5.2), let $t_0 := \min\{\tau_k^n : \tau_k^n \in (s,t)\} \wedge t$. If $t_0 = \tau_{k_0}^n < t$, we let $\tau_{k_0+m-1}^n$ be the maximal $\tau_k^n \in (s,t)$, for $m \geq 1$, and we write $t_k := \tau_{k_0+k}^n$ for $k = 1, \ldots, m-1$, while $t_m := t$. Otherwise, we set $m := 0$. Then

$$\begin{aligned} \left|\int_s^t Y_r^n\,\mathrm{d}N_r - Y_s N_{s,t}\right| &\leq \left|\int_s^{t_0} Y_r^n\,\mathrm{d}N_r - Y_s N_{s,t_0}\right| + \left|\int_{t_0}^{t_m} (Y_r^n - Y_{t_0}^n)\,\mathrm{d}N_r\right| \\ &\quad + \left|(Y_{t_0}^n - Y_{t_0}) N_{t_0,t_m}\right| + \left|(Y_{t_0} - Y_s) N_{t_0,t_m}\right|, \end{aligned} \tag{5.5}$$

The first and third term on the right-hand side are bounded by

$$\begin{aligned} &\left|\int_s^{t_0} Y_r^n\,\mathrm{d}N_r - Y_s N_{s,t_0}\right| + \left|(Y_{t_0}^n - Y_{t_0}) N_{t_0,t_m}\right| \\ &\quad = \left|Y_s^n N_{s,t_0} - Y_s N_{s,t_0}\right| + \left|(Y_{t_0}^n - Y_{t_0}) N_{t_0,t_m}\right| \\ &\quad \leq 2 \times 2^{-n} c(s,t)^{\frac{1}{q}} \|N\|_{q,[0,T]}, \end{aligned} \tag{5.6}$$

and the last term is controlled by

$$\left|(Y_{t_0} - Y_s) N_{t_0,t_m}\right| \leq c(s,t)^{\frac{1}{p}+\frac{1}{q}} \|Y\|_{p,[0,T]} \|N\|_{q,[0,T]}.$$

To bound the second term in (5.5), we use an idea from [28], Theorem 4.12: We apply Young's maximal inequality despite the fact that Y_- and N are not sufficiently regular for the construction of the Young integral. This will give us a divergent factor in n, but on the other hand it gives us a large power of $c(s,t)$. Then we balance this term with the other terms in the upper bound for $|A_{s,t}|$ (which all contain a factor 2^{-n}) by choosing the right n. Young's idea is to successively delete points from the partition $t_0 < \cdots < t_m$ in order to pass from $\sum_{k=0}^{m-1} Y_{t_k} N_{t_k,t_{k+1}}$ to $Y_{t_0} N_{t_0,t_m}$. We want to delete the points in an optimal way, and to express what optimal means we first renormalize Y and N:

$$\int_{t_0}^{t_m} (Y_r^n - Y_{t_0}^n)\,\mathrm{d}N_r = \sum_{k=0}^{m-1} (Y_{t_k} - Y_{t_0}) N_{t_k,t_{k+1}}$$
$$= \sum_{k=0}^{m-1} (\tilde{Y}_{t_k} - \tilde{Y}_{t_0}) \tilde{N}_{t_k,t_{k+1}} \|Y\|_{p,[0,T]} \|N\|_{q,[0,T]},$$

where $\tilde{Y} = \frac{Y}{\|Y\|_{p,[0,T]}}$ and $\tilde{N} = \frac{N}{\|N\|_{q,[0,T]}}$. Then c controls $\tilde{Y}$ and $\tilde{N}$, and by the superadditivity of c there exists $\ell \in \{1,\dots,m-1\}$ with $c(t_{\ell-1}, t_{\ell+1}) \le \frac{2}{m-1} c(s,t)$ whenever $m > 1$ (for $m = 1$ the integral vanishes). By deleting the point t_ℓ from the partition and subtracting the resulting expression, we get

$$|\tilde{Y}_{t_{\ell-1}} \tilde{N}_{t_{\ell-1},t_\ell} + \tilde{Y}_{t_\ell} \tilde{N}_{t_\ell,t_{\ell+1}} - \tilde{Y}_{t_{\ell-1}} \tilde{N}_{t_{\ell-1},t_{\ell+1}}| = |\tilde{Y}_{t_{\ell-1},t_\ell} \tilde{N}_{t_\ell,t_{\ell+1}}|$$
$$\le c(t_{\ell-1}, t_{\ell+1})^{\frac{1}{p}+\frac{1}{q}}$$
$$\le \left(\frac{2}{m-1} c(s,t)\right)^{\frac{1}{p}+\frac{1}{q}}.$$

We proceed by successively deleting all points except t_0 and t_m from the partition, each time in such an "optimal" way, and obtain

$$\left| \sum_{k=0}^{m-1} (\tilde{Y}_{t_k} - \tilde{Y}_{t_0}) \tilde{N}_{t_k,t_{k+1}} \right| \le \sum_{k=1}^{m-1} \left(\frac{2}{k} c(s,t)\right)^{\frac{1}{p}+\frac{1}{q}} \lesssim (m-1)^{1-\frac{1}{p}-\frac{1}{q}} c(s,t)^{\frac{1}{p}+\frac{1}{q}}.$$

Moreover,

$$m - 1 = \#\{k : \tau_k^n \in (\tau_{k_0}^n, t)\} \le 2^{np} \|Y\|_{p,[s,t]}^p \le 2^{np} c(s,t) \|Y\|_{p,[0,T]}^p.$$

So overall

$$\left| \int_{t_0}^{t_m} (Y_r^n - Y_{t_0}^n)\,\mathrm{d}N_r \right| \lesssim 2^{np(1-\frac{1}{p}-\frac{1}{q})} c(s,t) \|Y\|_{p,[0,T]}^{p(1-\frac{1}{p}-\frac{1}{q})+1} \|N\|_{q,[0,T]}. \tag{5.7}$$

We combine (5.2), (5.3), (5.5), (5.6), (5.7) and obtain the key bound

$$|A_{s,t}| \lesssim |n| 2^{-n} c(s,t)^{\frac{1}{q}} \big(K^{\frac{1}{q}} + \|N\|_{q,[0,T]}\big)$$
$$+ 2^{np(1-\frac{1}{p}-\frac{1}{q})} c(s,t) \|Y\|_{p,[0,T]}^{p(1-\frac{1}{p}-\frac{1}{q})+1} \|N\|_{q,[0,T]}$$
$$+ c(s,t)^{\frac{1}{p}+\frac{1}{q}} \|Y\|_{p,[0,T]} \|N\|_{q,[0,T]}.$$

To balance the first and second term, choose $n \in \mathbb{Z} \setminus \{0\}$ so that $\frac{1}{2} < 2^n c(s,t)^{\frac{1}{p}} \|Y\|_{p,[0,T]} \le 2$. Then

$$|A_{s,t}| \lesssim |n| c(s,t)^{\frac{1}{p}+\frac{1}{q}} \|Y\|_{p,[0,T]} \big(K^{\frac{1}{q}} + \|N\|_{q,[0,T]}\big)$$
$$\simeq \big|\log\big(c(s,t)^{\frac{1}{p}} \|Y\|_{p,[0,T]}\big)\big| c(s,t)^{\frac{1}{p}+\frac{1}{q}} \|Y\|_{p,[0,T]} \big(K^{\frac{1}{q}} + \|N\|_{q,[0,T]}\big)$$
$$\le \big(\big|\log\big(c(s,t)^{\frac{1}{p}}\big)\big| + \big|\log \|Y\|_{p,[0,T]}\big|\big) c(s,t)^{\frac{1}{p}+\frac{1}{q}} \|Y\|_{p,[0,T]} \big(K^{\frac{1}{q}} + \|N\|_{q,[0,T]}\big)$$

and since $c(s,t)^{\frac{1}{p}} \le 4$, for $\delta > 0$ we have $|\log(c(s,t)^{\frac{1}{p}})| \lesssim c(s,t)^{-\delta}$. Thus, we get (5.1) whenever $r > (1/p + 1/q)^{-1}$.

To estimate K, note first that since $2/q < 1$ we have $(\sum_m a_m)^{2/q} \le \sum_m a_m^{2/q}$. Applying Theorem A.2 (Lépingle's q-variation Burkholder–Davis–Gundy inequality), we get

$$\begin{aligned}
\mathbb{E}\big[K^{2/q}\big] &\le \sum_{m\in\mathbb{Z}\setminus\{0\}} |m|^{-2} 2^{2m} \mathbb{E}\bigg[\bigg\| \int_0^\cdot (Y_{r-} - Y_r^m)\,\mathrm{d}N_r \bigg\|_{q,[0,T]}^2\bigg] \\
&\lesssim \sum_{m\in\mathbb{Z}\setminus\{0\}} |m|^{-2} 2^{2m} \mathbb{E}\bigg[\bigg[\int_0^\cdot (Y_{r-} - Y_r^m)\,\mathrm{d}N_r \bigg]_T\bigg] \\
&= \sum_{m\in\mathbb{Z}\setminus\{0\}} |m|^{-2} 2^{2m} \mathbb{E}\bigg[\bigg\langle \int_0^\cdot (Y_{r-} - Y_r^m)\,\mathrm{d}N_r \bigg\rangle_T\bigg] \\
&= \sum_{m\in\mathbb{Z}\setminus\{0\}} |m|^{-2} 2^{2m} \mathbb{E}\bigg[\int_0^T (Y_{r-} - Y_r^m)^2\,\mathrm{d}\langle N\rangle_r \bigg] \\
&\lesssim \mathbb{E}\big[\langle N\rangle_T\big],
\end{aligned}$$

as required. □

PROOF OF PROPOSITION 3.8. Using Proposition 5.2 we have the following estimate for any $\tilde{q} > 2$, $r > (\frac{1}{p} + \frac{1}{\tilde{q}})^{-1}$ and $0 < \varepsilon < 1/2$:

$$\mathbb{E}\big[\|A\|_{r,[0,T]}^{1-\varepsilon}\big] \lesssim \mathbb{E}\big[\big(\big(1 + \big|\log \|Y\|_{p,[0,T]}\big|\big)\|Y\|_{p,[0,T]}\big)^{2-2\varepsilon}\big]^{\frac{1}{2}} \mathbb{E}\big[K^{2/\tilde{q}} + \|N\|_{\tilde{q},[0,T]}^2\big]^{\frac{1-\varepsilon}{2}},$$

where $\mathbb{E}[K^{2/\tilde{q}}] \lesssim \mathbb{E}[\langle N\rangle_T]$. Applying Theorem A.2 (Lépingle's $\tilde{q}$-variation Burkholder–Davis–Gundy inequality), we get

$$\mathbb{E}\big[\|N\|_{\tilde{q},[0,T]}^2\big] \lesssim \mathbb{E}\big[\langle N\rangle_T\big].$$

The remaining expectation is bounded by

$$\mathbb{E}\big[\big(\big(1 + \big|\log \|Y\|_{p,[0,T]}\big|\big)\|Y\|_{p,[0,T]}\big)^{2-2\varepsilon}\big] \lesssim 1 + \mathbb{E}\big[\|Y\|_{p,[0,T]}^2\big].$$

To end, note that we can write $\frac{q}{2} > (\frac{1}{p} + \frac{1}{\tilde{q}})^{-1}$ for some $\tilde{q} > 2$ close enough to 2 since $\frac{q}{2} > \frac{2p}{p+2} = (\frac{1}{p} + \frac{1}{2})^{-1}$. The required statement follows by taking $r = q/2$ together with the last $\tilde{q}$. □

APPENDIX: AUXILIARY ESTIMATES

LEMMA A.1 (Iterated Kipnis–Varadhan estimate). *Let $H \in \mathcal{H}^{-1} \cap L^2(\pi)$ and let A be a continuous adapted process of finite variation. Then*

$$\mathbb{E}\bigg[\sup_{t\le T}\bigg|\int_0^t A_s H(X_s)\,\mathrm{d}s\bigg|\bigg] \lesssim \mathbb{E}\Big[\sup_{t\le T}|A_t|^2\Big]^{1/2} T^{1/2} \|H\|_{\mathcal{H}^{-1}},$$

so in particular we get for $A_t = \int_0^t G(X_s)\,\mathrm{d}s$ with $G \in \mathcal{H}^{-1} \cap L^2(\pi)$,

$$\mathbb{E}\bigg[\sup_{t\le T}\bigg|\int_0^t \int_0^s G(X_r)\,\mathrm{d}r\, H(X_s)\,\mathrm{d}s\bigg|\bigg] \lesssim T\|G\|_{\mathcal{H}^{-1}}\|H\|_{\mathcal{H}^{-1}}.$$

PROOF. The second inequality follows from the first one together with the usual Kipnis–Varadhan estimate from Lemma 3.7. To show the first inequality, let $\Psi \in \mathcal{C}$ and apply Lemma 3.6:

$$\int_0^t A_s H(X_s)\,\mathrm{d}s = \frac{1}{2}\int_0^t A_s\,\mathrm{d}M_s^\Psi - \frac{1}{2}\int_{T-t}^T (A_T - A_{T-s})\,\mathrm{d}\hat{M}_s^\Psi + \frac{1}{2}A_T(\hat{M}_T^\Psi - \hat{M}_{T-t}^\Psi) + \int_0^t A_s(H(X_s) - \mathcal{L}_S\Psi(X_s))\,\mathrm{d}s,$$

where we need that A is continuous and of finite variation in order to interpret the integrals against $\hat{M}^\Psi$ in a pathwise sense and without having to worry about the difference of forward and backward integral. Now we get from the Burkholder–Davis–Gundy and Cauchy–Schwarz inequalities together with Lemma 3.7,

$$\begin{aligned}\mathbb{E}\Big[\sup_{t\le T}\Big|\int_0^t A_s H(X_s)\,\mathrm{d}s\Big|\Big] &\lesssim \mathbb{E}\Big[\sup_{t\le T}|A_t|^2\Big]^{1/2} T^{1/2}\|\Psi\|_{\mathcal{H}^1} \\ &\quad + \mathbb{E}\Big[\sup_{t\le T}\Big|\int_0^t A_s(H(X_s) - \mathcal{L}_S\Psi(X_s))\,\mathrm{d}s\Big|\Big] \\ &\lesssim \mathbb{E}\Big[\sup_{t\le T}|A_t|^2\Big]^{1/2}(T^{1/2}\|\Psi\|_{\mathcal{H}^1} + T\|H - \mathcal{L}_S\Psi\|_{L^2(\pi)}).\end{aligned}$$

By approximation, we can take $\Psi = \Phi_\lambda^H$ as the solution to the Poisson equation $(\lambda - \mathcal{L}_S)\Phi_\lambda^H = -H$ and as in the proof of Corollary 3.10 we use that $\|\Phi_\lambda^H\|_{\mathcal{H}^1} \le \|H\|_{\mathcal{H}^{-1}}$ for all $\lambda > 0$ and that $\|H - \mathcal{L}_S\Phi_\lambda^H\|_{L^2(\pi)} \to 0$ as $\lambda \to 0$ to deduce the claimed estimate. □

The following is the Lépingle p-variation inequality [22], Proposition 2, combined with the well-known Burkholder–Davis–Gundy inequality, cf. [10].

THEOREM A.2 (Lépingle p-variation Burkholder–Davis–Gundy inequality). *Let $(M_t)_{t\ge 0}$ be a local martingale with trajectories in $D(\mathbb{R}_+, \mathbb{R}^m)$. For every $T > 0$ and $p > 2$,*

$$c_p\mathbb{E}[[M]_T] \le \mathbb{E}[\|M\|_{p,[0,T]}^2] \le C_p\mathbb{E}[[M]_T],$$

where $c_p, C_p > 0$.

The next lemma is a strengthening of the ergodic theorem to give a path uniform convergence.

LEMMA A.3. *Let $(Y_t)_{t\ge 0}$ be a process with trajectories in $D(\mathbb{R}_+, \mathbb{R}^m)$ and with stationary increments and such that $\mathbb{E}[\sup_{t\in[0,T]}|Y_t|] \le CT$ for all $T > 0$ and such that $n^{-1}Y_n \to a$ for some $a \in \mathbb{R}^m$, both a.s. and in L^1. Assume also that $(\sup_{t\in[0,T]} n^{-1}|Y_{nt}|)_{n\in\mathbb{N}}$ is uniformly integrable for all $T > 0$. Then we have for all $T > 0$,*

$$\lim_{n\to\infty}\mathbb{E}\Big[\sup_{t\le T}|n^{-1}Y_{nt} - at|\Big] = 0.$$

PROOF. This follows from a minor adaptation of the proof of Theorem 2.29 in [19]: As in that proof, we decompose

$$|n^{-1}Y_{nt} - at| \le \sup_{s\in[0,1]}\frac{|Y_{\lfloor nt\rfloor+s} - Y_{\lfloor nt\rfloor}|}{n} + \frac{\lfloor nt\rfloor}{n}\left|\frac{Y_{\lfloor nt\rfloor}}{\lfloor nt\rfloor} - a\right| + |a|\left(t - \frac{\lfloor nt\rfloor}{n}\right).$$

The last term on the right-hand side is bounded by $|a|/n$. The first term on the right-hand side is bounded for all $t \in [0,T]$ by

$$\sup_{s\in[0,1]} \frac{|Y_{\lfloor nt\rfloor+s} - Y_{\lfloor nt\rfloor}|}{n} \le T \max_{k\le\lfloor nT\rfloor} \frac{\sup_{s\in[0,1]}|Y_{k+s}-Y_k|}{\lfloor nT\rfloor},$$

and by Lemma 2.30 in [19] the right-hand side vanishes as $n\to\infty$, both a.s. and in L^1 (here we need that Y has stationary increments). To handle the last remaining term, we decompose for $\delta \in (0,T)$:

$$\begin{aligned}\sup_{t\in[0,T]} \frac{\lfloor nt\rfloor}{n}\left|\frac{Y_{\lfloor nt\rfloor}}{\lfloor nt\rfloor} - a\right| &= \max_{k\le\lfloor nT\rfloor} \frac{k}{n}\left|\frac{Y_k}{k}-a\right| \\ &\le \max_{k\le\lfloor n\delta\rfloor}\frac{k}{n}\left|\frac{Y_k}{k}-a\right| + \max_{\lfloor n\delta\rfloor<k\le\lfloor nT\rfloor}\frac{k}{n}\left|\frac{Y_k}{k}-a\right|.\end{aligned}$$

The second term on the right-hand side converges almost surely to zero by assumption, and it is bounded from above by $\sup_{t\in[0,T]} n^{-1}|Y_{nt}|$. Since $(\sup_{t\in[0,T]} n^{-1}|Y_{nt}|)_n$ is uniformly integrable by assumption, this second term also converges to zero in L^1. The remaining part satisfies

$$\mathbb{E}\Big[\max_{k\le\lfloor n\delta\rfloor}\frac{k}{n}\left|\frac{Y_k}{k}-a\right|\Big] \le \frac{1}{n}\mathbb{E}\Big[\max_{k\le\lfloor n\delta\rfloor}|Y_k|\Big] + |a|\delta \le C\delta + |a|\delta,$$

where the last part follows by assumption on Y. The proof is then completed by sending $\delta\to 0$. □

Acknowledgments. We are grateful to the referees for their thorough read and numerous valuable remarks. In particular, the proof of the main result Theorem 3.3 was shortened due to a comment by one of the referees who observed a redundancy in the original argument.

We gratefully acknowledge financial support by the DFG via Research Unit FOR2402. The main part of the second author work was carried out while T.O. was employed at HU Berlin and TU Berlin. The main part of the third author work was carried out while N.P. was employed at MPI MIS Leipzig. N.P. gratefully acknowledges financial support by the DFG via the Heisenberg program.

REFERENCES

[1] BAILLEUL, I. and CATELLIER, R. (2017). Rough flows and homogenization in stochastic turbulence. *J. Differential Equations* **263** 4894–4928. MR3680942 https://doi.org/10.1016/j.jde.2017.06.006

[2] BRUNED, Y., CHANDRA, A., CHEVYREV, I. and HAIRER, M. Renormalising spdes in regularity structures. *J. Eur. Math. Soc.* (*JEMS*). To appear.

[3] BRUNED, Y., HAIRER, M. and ZAMBOTTI, L. (2019). Algebraic renormalisation of regularity structures. *Invent. Math.* **215** 1039–1156. MR3935036 https://doi.org/10.1007/s00222-018-0841-x

[4] CHANDRA, A. and HAIRER, M. (2016). An analytic BPHZ theorem for regularity structures. Preprint. Available at arXiv:1612.08138.

[5] CHEVYREV, I., FRIZ, P., KOREPANOV, A., MELBOURNE, I. and ZHANG, H. (2019). Deterministic homogenization for discrete-time fast-slow systems under optimal moment assumptions. Preprint. Available at arXiv:1903.10418.

[6] CHEVYREV, I. and FRIZ, P. K. (2019). Canonical RDEs and general semimartingales as rough paths. *Ann. Probab.* **47** 420–463. MR3909973 https://doi.org/10.1214/18-AOP1264

[7] CHEVYREV, I., FRIZ, P. K., KOREPANOV, A. and MELBOURNE, I. (2020). Superdiffusive limits for deterministic fast-slow dynamical systems. *Probab. Theory Related Fields* **178** 735–770. MR4168387 https://doi.org/10.1007/s00440-020-00988-5

[8] CHEVYREV, I., FRIZ, P. K., KOREPANOV, A., MELBOURNE, I. and ZHANG, H. (2019). Multiscale systems, homogenization, and rough paths. In *Probability and Analysis in Interacting Physical Systems*. *Springer Proc. Math. Stat.* **283** 17–48. Springer, Cham. MR3968507 https://doi.org/10.1007/978-3-030-15338-0_2

[9] FRIZ, P., GASSIAT, P. and LYONS, T. (2015). Physical Brownian motion in a magnetic field as a rough path. *Trans. Amer. Math. Soc.* **367** 7939–7955. MR3391905 https://doi.org/10.1090/S0002-9947-2015-06272-2

[10] FRIZ, P. and VICTOIR, N. (2008). The Burkholder–Davis–Gundy inequality for enhanced martingales. In *Séminaire de Probabilités XLI. Lecture Notes in Math.* **1934** 421–438. Springer, Berlin. MR2483743 https://doi.org/10.1007/978-3-540-77913-1_20

[11] FRIZ, P. and ZORIN-KRANICH, P. (2020). Rough semimartingales and p-variation estimates for martingale transforms. Preprint. Available at arXiv:2008.08897.

[12] FRIZ, P. K. and ZHANG, H. (2018). Differential equations driven by rough paths with jumps. *J. Differential Equations* **264** 6226–6301. MR3770049 https://doi.org/10.1016/j.jde.2018.01.031

[13] GLORIA, A., NEUKAMM, S. and OTTO, F. (2015). Quantification of ergodicity in stochastic homogenization: Optimal bounds via spectral gap on Glauber dynamics. *Invent. Math.* **199** 455–515. MR3302119 https://doi.org/10.1007/s00222-014-0518-z

[14] GUBINELLI, M. and PERKOWSKI, N. (2018). Energy solutions of KPZ are unique. *J. Amer. Math. Soc.* **31** 427–471. MR3758149 https://doi.org/10.1090/jams/889

[15] KELLY, D. (2016). Rough path recursions and diffusion approximations. *Ann. Appl. Probab.* **26** 425–461. MR3449323 https://doi.org/10.1214/15-AAP1096

[16] KELLY, D. and MELBOURNE, I. (2016). Smooth approximation of stochastic differential equations. *Ann. Probab.* **44** 479–520. MR3456344 https://doi.org/10.1214/14-AOP979

[17] KELLY, D. and MELBOURNE, I. (2017). Deterministic homogenization for fast-slow systems with chaotic noise. *J. Funct. Anal.* **272** 4063–4102. MR3626033 https://doi.org/10.1016/j.jfa.2017.01.015

[18] KIPNIS, C. and VARADHAN, S. R. S. (1986). Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Comm. Math. Phys.* **104** 1–19. MR0834478

[19] KOMOROWSKI, T., LANDIM, C. and OLLA, S. (2012). *Fluctuations in Markov Processes: Time Symmetry and Martingale Approximation. Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **345**. Springer, Heidelberg. MR2952852 https://doi.org/10.1007/978-3-642-29880-6

[20] KURTZ, T. G. and PROTTER, P. (1991). Weak limit theorems for stochastic integrals and stochastic differential equations. *Ann. Probab.* **19** 1035–1070. MR1112406

[21] LEJAY, A. and LYONS, T. (2005). On the importance of the Lévy area for studying the limits of functions of converging stochastic processes. Application to homogenization. In *Current Trends in Potential Theory. Theta Ser. Adv. Math.* **4** 63–84. Theta, Bucharest. MR2243956

[22] LÉPINGLE, D. (1975). Sur la variation d'ordre p des martingales locales. *C. R. Acad. Sci. Paris Sér. A-B* **281** A917–A919. MR0418225

[23] LOPUSANSCHI, O. and ORENSHTEIN, T. (2018). Ballistic random walks in random environment as rough paths: Convergence and area anomaly. Preprint. Available at arXiv:1812.01403.

[24] LOPUSANSCHI, O. and SIMON, D. (2018). Lévy area with a drift as a renormalization limit of Markov chains on periodic graphs. *Stochastic Process. Appl.* **128** 2404–2426. MR3804798 https://doi.org/10.1016/j.spa.2017.09.004

[25] MOURRAT, J.-C. (2012). A quantitative central limit theorem for the random walk among random conductances. *Electron. J. Probab.* **17** no. 97, 17. MR2994845 https://doi.org/10.1214/EJP.v17-2414

[26] ORENSHTEIN, T. (2021). Rough invariance principle for delayed regenerative processes. Preprint. Available at arXiv:2101.05222.

[27] PELIGRAD, M. (2020). A new CLT for additive functionals of Markov chains. *Stochastic Process. Appl.* **130** 5695–5708. MR4127343 https://doi.org/10.1016/j.spa.2020.04.004

[28] PERKOWSKI, N. and PRÖMEL, D. J. (2016). Pathwise stochastic integrals for model free finance. *Bernoulli* **22** 2486–2520. MR3498035 https://doi.org/10.3150/15-BEJ735