

MULTILEVEL MONTE CARLO FOR LÉVY-DRIVEN SDEs: CENTRAL LIMIT THEOREMS FOR ADAPTIVE EULER SCHEMES

BY STEFFEN DEREICH AND SANGMENG LI¹

Westfälische Wilhelms-Universität Münster

In this article, we consider multilevel Monte Carlo for the numerical computation of expectations for stochastic differential equations driven by Lévy processes. The underlying numerical schemes are based on jump-adapted Euler schemes. We prove stable convergence of an idealised scheme. Further, we deduce limit theorems for certain classes of functionals depending on the whole trajectory of the process. In particular, we allow dependence on marginals, integral averages and the supremum of the process. The idealised scheme is related to two practically implementable schemes and corresponding central limit theorems are given. In all cases, we obtain errors of order $N^{-1/2}(\log N)^{1/2}$ in the computational time N which is the same order as obtained in the classical set-up analysed by Giles [*Oper. Res.* **56** (2008) 607–617]. Finally, we use the central limit theorems to optimise the parameters of the multilevel scheme.

1. Introduction. The numerical computation of expectations $\mathbb{E}[F(X)]$ for solutions $(X_t)_{t \in [0, T]}$ of stochastic differential equations (SDEs) is a classical problem in stochastic analysis and numerous numerical schemes were developed and analysed within the last twenty years; see, for instance, the textbooks by Kloeden and Platen [21] and Glasserman [13]. Recently, a new very efficient class of Monte Carlo algorithms was introduced by Giles [12]; see also Heinrich [14] for an earlier variant of the computational concept. Central to these *multilevel* Monte Carlo algorithms is the use of whole hierarchies of approximations in numerical simulations. For SDEs, multilevel algorithms often achieve errors of order $N^{-1/2+o(1)}$ in the computational time N (see [10, 12]) despite the infinite-dimensional nature of the stochastic differential equation. Further, the algorithms are in many cases optimal in a worst case sense [7]. So far, the main focus of research was concerned with asymptotic error estimates, whereas central limit theorems have only found minor attention yet. Beyond the central limit theorem, developed by Ben Alaya and Kebaier [4] for the Euler scheme for diffusions no further results are available yet. In general, central limit theorems illustrate how the choice of parameters affects the efficiency of the scheme and they are a central tool for tuning the parameters.

Received March 2014; revised November 2014.

¹Supported by the Deutsche Forschungsgemeinschaft (DFG) within the Priority Programme 1324 in the project “Constructive Quantization and Multilevel Algorithms for Quadrature of SDEs.”

MSC2010 subject classifications. Primary 65C05; secondary 60G51, 60F05.

Key words and phrases. Multilevel Monte Carlo, central limit theorem, Lévy-driven stochastic differential equation, Euler scheme, jump-adapted scheme, stable convergence.

In this article, we focus on central limit theorems for Lévy-driven stochastic differential equations. We prove stable convergence of the error process of an idealised jump-adapted Euler schemes. Based on this result, we derive central limit theorems for multilevel schemes for the approximate computation of expectations of functionals depending on marginals, integral averages and the supremum of the SDE. We then introduce implementable jump-adapted Euler schemes that inherit the properties of the idealised schemes so that the main results prevail. Finally, we use our new results to optimise over the parameters of the scheme, and thereby complement the research conducted in [12]. In the Parameter Optimisation 1.13 below, we find that often it is preferable to increase the number of Euler steps from level to level by a factor of 6. For ease of presentation, we restrict attention to the one-dimensional setting although a generalisation to finite-dimensional stochastic differential equations is canonical.

In the following, $(\Omega, \mathcal{F}, \mathbb{P})$ denotes a probability space that is sufficiently rich to ensure existence of all random variables used in the exposition. We let $Y = (Y_t)_{t \in [0, T]}$ be a square integrable Lévy-process and note that there exist $b \in \mathbb{R}$ (drift), $\sigma^2 \in [0, \infty)$ (diffusion coefficient) and a measure ν on $\mathbb{R} \setminus \{0\}$ with $\int x^2 \nu(dx) < \infty$ (Lévy measure) such that

$$\mathbb{E}[e^{izY_t}] = \exp \left\{ t \left(ibz - \frac{1}{2} \sigma^2 z + \int (e^{izx} - 1 - izx) \nu(dx) \right) \right\}$$

for $t \in [0, T]$ and $z \in \mathbb{R}$. We call the unique triplet (b, σ^2, ν) the Lévy triplet of Y . We refer the reader to the textbooks by Applebaum [2], Bertoin [5] and Sato [31] for a concise treatment of Lévy processes. The process $X = (X_t)_{t \in [0, T]}$ denotes the solution to the stochastic integral equation

$$(1.1) \quad X_t = x_0 + \int_0^t a(X_{s-}) dY_s, \quad t \in [0, T],$$

where $a: \mathbb{R} \rightarrow \mathbb{R}$ is a continuously differentiable Lipschitz function and $x_0 \in \mathbb{R}$. Both processes Y and X attain values in the space of càdlàg functions on $[0, T]$ which we will denote by $\mathbb{D}(\mathbb{R})$ and endow with the Skorokhod topology. We will analyse multilevel algorithms for the computation of expectations $\mathbb{E}[F(X)]$, where $F: \mathbb{D}(\mathbb{R}) \rightarrow \mathbb{R}$ is a measurable functional such that $F(x)$ depends on the marginals, integrals and/or supremum of the path $x \in \mathbb{D}(\mathbb{R})$. Before we state the results, we introduce the underlying numerical schemes.

1.1. Jump-adapted Euler scheme. In the context of Lévy-driven stochastic differential equations, there are various Euler-type schemes analysed in the literature. We consider jump-adapted Euler schemes. For finite Lévy measures, these were introduced by Platen [27] and analysed by various authors; see, for example, [6, 25]. For infinite Lévy measures, an error analysis is conducted in [10] and [8] for two multilevel Monte Carlo schemes. Further, weak approximation is analysed in [22] and [26]. In general, the simulation of increments of the Lévy-process is delicate.

One can use truncated shot noise representations as in [30]. These perform well for Blumenthal–Gettoor indices smaller than one, but are less efficient when the BG-index gets larger than one [10], even when combined with a Gaussian compensation in the spirit of [3]; see [9]. A faster simulation technique is to do an inversion of the characteristic function of the Lévy process and to establish direct simulation routines in a precomputation. Certainly, this approach is more involved and its realisation imposes severe restrictions on the dimension of the Lévy process; see [11].

In this article, we analyse one prototype of adaptive approximations that is intimately related to implementable adaptive schemes and we thus believe that our results have a universal appeal. The approximations depend on two positive parameters:

- h , the threshold for the size of the jumps being considered large and causing immediate updates, and
- ε with $T \in \varepsilon\mathbb{N}$, the length of the regular update intervals.

For the definition of the approximations, we use the simple Poisson point process Π on the Borel sets of $(0, T] \times (\mathbb{R} \setminus \{0\})$ associated to Y , that is,

$$\Pi = \sum_{s \in (0, T]: \Delta Y_s \neq 0} \delta_{(s, \Delta Y_s)},$$

where we use the notation $\Delta x_t = x_t - x_{t-}$ for $x \in \mathbb{D}(\mathbb{R})$ and $t \in (0, T]$. It has intensity $\ell_{(0, T]} \otimes \nu$, where $\ell_{(0, T]}$ denotes Lebesgue measure on $(0, T]$. Further, let $\bar{\Pi}$ be the compensated variant of Π that is the random signed measure on $(0, T] \times (\mathbb{R} \setminus \{0\})$ given by

$$\bar{\Pi} = \Pi - \ell_{(0, T]} \otimes \nu.$$

The process $(Y_t)_{t \in [0, T]}$ admits the representation

$$(1.2) \quad Y_t = bt + \sigma W_t + \lim_{\delta \downarrow 0} \int_{(0, t] \times B(0, \delta)^c} x \, d\bar{\Pi}(s, x),$$

where $(W_t)_{t \in [0, T]}$ is an appropriate Brownian motion that is independent of Π and the limit is to be understood uniformly in \mathbb{L}^2 . We enumerate the random set

$$(\varepsilon\mathbb{Z} \cap [0, T]) \cup \{t \in (0, T] : |\Delta Y_t| \geq h\} = \{T_0, T_1, \dots\}$$

in increasing order and define the approximation $X^{h, \varepsilon} = (X_t^{h, \varepsilon})_{t \in [0, T]}$ by $X_0^{h, \varepsilon} = x_0$ and, for $n = 1, 2, \dots$ and $t \in (T_{n-1}, T_n]$

$$(1.3) \quad X_t^{h, \varepsilon} = X_{T_{n-1}}^{h, \varepsilon} + a(X_{T_{n-1}}^{h, \varepsilon})(Y_t - Y_{T_{n-1}}).$$

1.2. *Multilevel Monte Carlo.* In general, multilevel schemes make use of whole hierarchies of approximate solutions and we choose decreasing sequences $(\varepsilon_k)_{k \in \mathbb{N}}$ and $(h_k)_{k \in \mathbb{N}}$ with:

(ML1) $\varepsilon_k = M^{-k}T$, where $M \in \{2, 3, \dots\}$ is fixed,

(ML2) $\lim_{k \rightarrow \infty} \nu(B(0, h_k)^c) \varepsilon_k = \theta$ for a $\theta \in [0, \infty)$ and $\lim_{k \rightarrow \infty} h_k / \sqrt{\varepsilon_k} = 0$.

We remark that whenever θ in (ML2) is strictly positive, then one automatically has that $h_k = o(\sqrt{\varepsilon_k})$; see Lemma A.10.

For every $k \in \mathbb{N}$, we denote by $X^k := X^{h_k, \varepsilon_k}$ the corresponding adaptive Euler approximation with update rule (1.3). Once this hierarchy of approximations has been fixed, a multilevel scheme \widehat{S} is parameterised by a \mathbb{N} -valued vector (n_1, \dots, n_L) of arbitrary finite length L : for a measurable function $F : \mathbb{D}(\mathbb{R}) \rightarrow \mathbb{R}$ we approximate $\mathbb{E}[F(X)]$ by

$$\mathbb{E}[F(X^1)] + \mathbb{E}[F(X^2) - F(X^1)] + \dots + \mathbb{E}[F(X^L) - F(X^{L-1})]$$

and denote by $\widehat{S}(F)$ the random output that is obtained when estimating the individual expectations $\mathbb{E}[F(X^1)]$, $\mathbb{E}[F(X^2) - F(X^1)]$, \dots , $\mathbb{E}[F(X^L) - F(X^{L-1})]$ independently by classical Monte Carlo with n_1, \dots, n_L iterations and summing up the individual estimates. More explicitly, a multilevel scheme \widehat{S} associates to each measurable F a random variable

$$(1.4) \quad \widehat{S}(F) = \frac{1}{n_1} \sum_{i=1}^{n_1} F(X^{1,i}) + \sum_{k=2}^L \frac{1}{n_k} \sum_{i=1}^{n_k} (F(X^{k,i,f}) - F(X^{k-1,i,c})),$$

where the pairs of random variables $(X^{k,i,f}, X^{k-1,i,c})$, respectively, the random variables $X^{1,i}$, appearing in the sums are all independent with identical distribution as (X^k, X^{k-1}) , respectively, X^1 . Note that the upper indices f and c refer to fine and coarse and that the entries of each pair are *not* independent.

1.3. *Implementable schemes.* We give two implementable schemes. The first one relies on precomputation for direct simulation of Lévy increments. The second one ignores jumps of size smaller than a threshold which leads to schemes of optimal order only in the case where—roughly speaking—the Blumenthal–Gettoor index is smaller than one.

Schemes with direct simulation of small jumps. For $h > 0$, we let $Y^h = (Y_t^h)_{t \in [0, T]}$ denote the Lévy process given by

$$(1.5) \quad Y_t^h = bt + \sigma W_t + \int_{(0, t] \times B(0, h)^c} x \, d\overline{\Pi}(s, x).$$

Using the shot noise representation (see [3]), we can simulate Y^h on arbitrary (random) time sets. The remainder $M^h = (M_t^h)_{t \in [0, T]}$, that is,

$$M_t^h = \lim_{\delta \downarrow 0} \int_{(0, t] \times (B(0, h) \setminus B(0, \delta))} x \, d\overline{\Pi}(s, x) = Y - Y^h,$$

can be simulated on a fixed time grid $\varepsilon'\mathbb{Z} \cap [0, T]$ with $\varepsilon' \in \varepsilon\mathbb{N}$ denoting an additional parameter of the scheme. A corresponding approximation is given by $\widehat{X}^{h,\varepsilon,\varepsilon'} = (\widehat{X}_t^{h,\varepsilon,\varepsilon'})_{t \in [0, T]}$ via $\widehat{X}_0^{h,\varepsilon,\varepsilon'} = x_0$ and, for $n = 1, 2, \dots$ and $t \in (T_{n-1}, T_n]$,

$$(1.6) \quad \begin{aligned} \widehat{X}_t^{h,\varepsilon,\varepsilon'} &= \widehat{X}_{T_{n-1}}^{h,\varepsilon,\varepsilon'} + a(\widehat{X}_{T_{n-1}}^{h,\varepsilon,\varepsilon'})(Y_t^h - Y_{T_{n-1}}^h) \\ &\quad + \mathbb{1}_{\varepsilon'\mathbb{Z}}(t)a(\widehat{X}_{t-\varepsilon'}^{h,\varepsilon,\varepsilon'})(M_t^h - M_{t-\varepsilon'}^h). \end{aligned}$$

We call $\widehat{X}^{h,\varepsilon,\varepsilon'}$ the *continuous approximation* with parameters $h, \varepsilon, \varepsilon'$. Further, we define the *piecewise constant approximation* $\overline{X}^{h,\varepsilon,\varepsilon'} = (\overline{X}_t^{h,\varepsilon,\varepsilon'})_{t \in [0, T]}$ via demanding that for $n = 1, 2, \dots$ and $t \in [T_{n-1}, T_n]$,

$$(1.7) \quad \overline{X}_t^{h,\varepsilon,\varepsilon'} = \widehat{X}_{T_{n-1}}^{h,\varepsilon,\varepsilon'}$$

and $\overline{X}_T^{h,\varepsilon,\varepsilon'} = \widehat{X}_T^{h,\varepsilon,\varepsilon'}$.

In corresponding multilevel schemes, we choose $(\varepsilon_k)_{k \in \mathbb{N}}$ and $(h_k)_{k \in \mathbb{N}}$ as before. Further, we choose monotonically decreasing parameters $(\varepsilon'_k)_{k \in \mathbb{N}}$ with $\varepsilon'_k \in \varepsilon_k\mathbb{N}$ and:

$$(ML3a) \quad \varepsilon'_k \int_{B(0, h_k)} x^2 \nu(dx) \log^2(1 + 1/\varepsilon'_k) = o(\varepsilon_k),$$

$$(ML3b) \quad h_k^2 \log^2(1 + 1/\varepsilon'_k) = o(\varepsilon_k).$$

REMARK 1.1. If

$$(1.8) \quad \int x^2 \log^2\left(1 + \frac{1}{x}\right) \nu(dx) < \infty,$$

there exist appropriate parameters $(h_k, \varepsilon_k, \varepsilon'_k)_{k \in \mathbb{N}}$ satisfying (ML1), (ML2), (ML3a) and (ML3b). More precisely, in the case where ν is infinite, appropriate parameters are obtained by choosing $\varepsilon'_k = \varepsilon_k$ and (h_k) with $\lim_{k \rightarrow \infty} \varepsilon_k \nu(B(0, h_k)^c) = \theta > 0$; see Lemma A.10.

In analogy to before, we denote by $(\widehat{X}^k : k \in \mathbb{N})$ and $(\overline{X}^k : k \in \mathbb{N})$ the corresponding approximate continuous and piecewise constant solutions. We state a result of [24] which implies that in most cases the central limit theorems to be provided later are also valid for the continuous approximations.

LEMMA 1.2. *If assumptions (ML1), (ML3a) and (ML3b) are satisfied, then*

$$\lim_{k \rightarrow \infty} \varepsilon_k^{-1} \mathbb{E} \left[\sup_{t \in [0, T]} |X_t^k - \widehat{X}_t^k|^2 \right] = 0.$$

Practical issues of numerical schemes with direct simulation of increments are discussed in [11].

Truncated shot noise scheme. The truncated shot noise scheme is parameterised by two positive parameters h, ε as above. The *continuous approximations* $\widehat{X}^{h,\varepsilon} = (\widehat{X}_t^{h,\varepsilon})_{t \in [0, T]}$ are defined via $\widehat{X}_0^{h,\varepsilon} = x_0$ and, for $n = 1, 2, \dots$ and $t \in (T_{n-1}, T_n]$,

$$(1.9) \quad \widehat{X}_t^{h,\varepsilon} = \widehat{X}_{T_{n-1}}^{h,\varepsilon} + a(\widehat{X}_{T_{n-1}}^{h,\varepsilon})(Y_t^h - Y_{T_{n-1}}^h)$$

and the *piecewise constant approximations* $\overline{X}^{h,\varepsilon} = (\overline{X}_t^{h,\varepsilon})_{t \in [0, T]}$ are defined as before by demanding that, for $n = 1, 2, \dots$ and $t \in [T_{n-1}, T_n]$,

$$(1.10) \quad \overline{X}_t^{h,\varepsilon} = \widehat{X}_{T_{n-1}}^{h,\varepsilon}$$

and $\overline{X}_T^{h,\varepsilon} = \widehat{X}_T^{h,\varepsilon}$. Again we will use decreasing sequences (ε_k) and (h_k) as before to specify sequences of approximations (\widehat{X}^k) and (\overline{X}^k) . In the context of truncated shot noise schemes, we will impose as additional assumption:

$$(ML4) \quad \int_{B(0, h_k)} x^2 \nu(dx) = o(\varepsilon_k).$$

REMARK 1.3. If $\int |x| \nu(dx) < \infty$, then (ML1), (ML2) and (ML4) are satisfied for appropriate parameters.

The following result is a minor modification of [10], Proposition 1; see also [24].

LEMMA 1.4. *If assumptions (ML1) and (ML4) are satisfied, then*

$$\lim_{k \rightarrow \infty} \varepsilon_k^{-1} \mathbb{E} \left[\sup_{t \in [0, T]} |X_t^k - \widehat{X}_t^k|^2 \right] = 0.$$

1.4. *Main results.* In the following, we will always assume that $Y = (Y_t)_{t \in [0, T]}$ is a square integrable Lévy process with Lévy triplet (b, σ^2, ν) satisfying $\sigma^2 > 0$ and that $X = (X_t)_{t \in [0, T]}$ solves the SDE

$$dX_t = a(X_{t-}) dY_t$$

with $X_0 = x_0$, where $a: \mathbb{R} \rightarrow \mathbb{R}$ is a continuously differentiable Lipschitz function. Further, for each $k \in \mathbb{N}$, X^k denotes the jump-adapted Euler scheme with updates at all times in

$$(\varepsilon_k \mathbb{N} \cap [0, T]) \cup \{t \in (0, T) : |\Delta Y_t| \geq h_k\};$$

see (1.3). The decreasing sequences of parameters (ε_k) and (h_k) are assumed to satisfy (ML1) and (ML2) from Section 1.2.

Convergence of the error process. We consider the normalised sequence of error processes associated to the multilevel scheme that is the sequence $(\varepsilon_k^{-1/2}(X^{k+1} - X^k) : k \in \mathbb{N})$. Let us introduce the process appearing as a limit. We equip the points of the associated point process Π with independent marks and denote for a point $(s, x) \in \Pi$:

- by ξ_s , a standard normal random variable,
- by \mathcal{U}_s , an independent uniform random variable on $[0, 1]$, and
- by \mathcal{E}_s^θ and $\mathcal{E}_s^{(M-1)\theta}$ independent $\text{Exp}(\theta)$ and $\text{Exp}((M-1)\theta)$ -distributed random variables, respectively.

Further, we denote by $B = (B_t)_{t \in [0, T]}$ an independent standard Brownian motion.

The idealised error process $U = (U_t)_{t \in [0, T]}$ is defined as the solution of the integral equation

$$(1.11) \quad \begin{aligned} U_t = & \int_0^t a'(X_{s-})U_{s-} dY_s + \sigma^2 \Upsilon \int_0^t (aa')(X_{s-}) dB_s \\ & + \sum_{s \in (0, t] : \Delta Y_s \neq 0} \sigma_s \xi_s (aa')(X_{s-}) \Delta Y_s, \end{aligned}$$

where $\Upsilon^2 = \frac{e^{-\theta} - 1 + \theta}{\theta^2} (1 - \frac{1}{M})$, if $\theta > 0$, and $\Upsilon^2 = \frac{1}{2} (1 - \frac{1}{M})$, if $\theta = 0$, and the positive marks (σ_s) are defined by

$$\begin{aligned} \sigma_s^2 = \sigma^2 \sum_{1 \leq m \leq M} \mathbb{1}_{\{(m-1)/M \leq \mathcal{U}_s < m/M\}} & \left[\min(\mathcal{E}_s^\theta, \mathcal{U}_s) \right. \\ & \left. - \min\left(\mathcal{E}_s^\theta, \mathcal{E}_s^{(M-1)\theta}, \mathcal{U}_s - \frac{m-1}{M}\right) \right]. \end{aligned}$$

Note that the above infinite sum has to be understood as an appropriate martingale limit. More explicitly, denoting by $L = (L_t)_{t \in [0, T]}$ the Lévy process

$$L_t = \sigma^2 \Upsilon B_t + \lim_{\delta \downarrow 0} \sum_{s \in (0, t] : |\Delta Y_s| \geq \delta} \sigma_s \xi_s \Delta Y_s$$

we can rewrite (1.11) as

$$U_t = \int_0^t a'(X_{s-})U_{s-} dY_s + \int_0^t (aa')(X_{s-}) dL_s.$$

Strong uniqueness and existence of the solution follow from Jacod and Memin [16], Theorem 4.5.

THEOREM 1.5. *Under the above assumptions, we have weak convergence*

$$(1.12) \quad (Y, \varepsilon_n^{-1/2}(X^{n+1} - X^n)) \Rightarrow (Y, U) \quad \text{in } \mathbb{D}(\mathbb{R}^2).$$

Central limit theorem for linear functionals. We consider functionals $F : \mathbb{D}(\mathbb{R}) \rightarrow \mathbb{R}$ of the form

$$F(x) = f(Ax)$$

with $f : \mathbb{R}^d \rightarrow \mathbb{R}$ and $A : \mathbb{D}(\mathbb{R}) \rightarrow \mathbb{R}^d$ being linear and measurable. We set

$$D_f := \{z \in \mathbb{R}^d : f \text{ is differentiable in } z\}.$$

THEOREM 1.6. *Suppose that f is Lipschitz continuous and that A is Lipschitz continuous with respect to supremum norm and continuous with respect to the Skorokhod topology in \mathbb{P}_U -almost every path. Further suppose that $AX \in D_f$, almost surely, and that $\alpha \geq \frac{1}{2}$ is such that the limit*

$$\lim_{n \rightarrow \infty} \varepsilon_n^{-\alpha} \mathbb{E}[F(X^n) - F(X)] =: \kappa$$

exists. We denote for $\delta \in (0, 1)$ by \widehat{S}_δ the multilevel Monte Carlo scheme with parameters $(n_1^{(\delta)}, n_2^{(\delta)}, \dots, n_{L(\delta)}^{(\delta)})$, where

$$(1.13) \quad L(\delta) = \left\lceil \frac{\log \delta^{-1}}{\alpha \log M} \right\rceil \quad \text{and} \quad n_k(\delta) = \lceil \delta^{-2} L(\delta) \varepsilon_{k-1} \rceil,$$

for $k = 1, 2, \dots, L(\delta)$. Then we have,

$$\delta^{-1} (\widehat{S}_\delta(F) - \mathbb{E}[F(X)]) \Rightarrow \mathcal{N}(\kappa, \rho^2) \quad \text{as } \delta \rightarrow 0,$$

where $\mathcal{N}(\kappa, \rho^2)$ is the normal distribution with mean κ and variance

$$\rho^2 = \text{Var}(\nabla f(AX) \cdot AU).$$

EXAMPLE 1.7. (a) For any finite signed measure μ , the integral $Ax = \int_0^T x_s d\mu(s)$ satisfies the assumptions of the theorem. Indeed, for every path $x \in \mathbb{D}(\mathbb{R})$ with

$$(1.14) \quad \mu(\{s \in [0, T] : \Delta x_s \neq 0\}) = 0$$

one has for $x^n \rightarrow x$ in the Skorokhod space that

$$Ax^n = \int_0^T x_s^n d\mu(s) \rightarrow \int_0^T x_s d\mu(s) = Ax$$

by dominated convergence and (1.14) is true for \mathbb{P}_U -almost all paths since μ has at most countably many atoms. Hence, the linear maps $Ax = x_t$ and $Ax = \int_0^T x_s ds$ are allowed choices in Theorem 1.6 since U is almost surely continuous in t .

(b) All combinations of admissible linear maps A_1, \dots, A_m satisfy again the assumptions of the theorem.

In view of implementable schemes, we state a further version of the theorem.

THEOREM 1.8. *Suppose that either $(\widehat{X}^k : k \in \mathbb{N})$ and $(\bar{X}^k : k \in \mathbb{N})$ denote the continuous and piecewise constant approximations of the scheme with direct simulation and that (ML1), (ML2) and (ML3) are fulfilled or that they are the approximations of the truncated shot noise scheme and that (ML1), (ML2) and (ML4) are fulfilled. Then Theorem 1.6 remains true when replacing the family $(X^k : k \in \mathbb{N})$ by $(\widehat{X}^k : k \in \mathbb{N})$. Further, if A is given by*

$$Ax = \left(x_T, \int_0^T x_s \, ds \right),$$

the statement of the central limit theorem remains true, when replacing the family $(X^k : k \in \mathbb{N})$ by $(\bar{X}^k : k \in \mathbb{N})$.

Central limit theorem for supremum-dependent functionals. In this section, we consider functionals $F : \mathbb{D}(\mathbb{R}) \rightarrow \mathbb{R}$ of the form

$$F(x) = f\left(\sup_{t \in [0, T]} x_t\right)$$

with $f : \mathbb{R} \rightarrow \mathbb{R}$ measurable.

THEOREM 1.9. *Suppose that $f : \mathbb{R} \rightarrow \mathbb{R}$ is Lipschitz continuous and that the coefficient a does not attain zero. Further, suppose that $\sup_{t \in [0, T]} X_t \in D_f$, almost surely, and that $\alpha \geq \frac{1}{2}$ is such that the limit*

$$\lim_{n \rightarrow \infty} \varepsilon_n^{-\alpha} \mathbb{E}[F(X^n) - F(X)] =: \kappa$$

exists. We denote for $\delta \in (0, 1)$ by \widehat{S}_δ the multilevel Monte Carlo scheme with parameters $(n_1^{(\delta)}, n_2^{(\delta)}, \dots, n_{L(\delta)}^{(\delta)})$, where

$$L(\delta) = \left\lceil \frac{\log \delta^{-1}}{\alpha \log M} \right\rceil \quad \text{and} \quad n_k(\delta) = \lceil \delta^{-2} L(\delta) \varepsilon_{k-1} \rceil,$$

for $k = 1, 2, \dots, L(\delta)$. Then we have

$$\delta^{-1} (\widehat{S}_\delta(F) - \mathbb{E}[F(X)]) \Rightarrow \mathcal{N}(\kappa, \rho^2) \quad \text{as } \delta \rightarrow 0,$$

where $\mathcal{N}(\kappa, \rho^2)$ is the normal distribution with mean κ and variance

$$\rho^2 = \text{Var}\left(f'\left(\sup_{t \in [0, T]} X_t\right) U_S\right),$$

and S denotes the random time at which X attains its supremum.

THEOREM 1.10. *Theorem 1.9 remains true for the continuous approximations for the scheme with direct simulation of increments or the truncated shot noise scheme under the same assumptions as imposed in Theorem 1.8.*

Optimal parameters. We use the central limit theorems to adjust the parameters of the multilevel scheme. Here, we use the following result.

THEOREM 1.11. *Let F be as in Theorems 1.6 or 1.9 and assume that the assumptions of the respective theorem are fulfilled. Further assume in the first case that A is of integral type meaning that there exist finite signed measures μ_1, \dots, μ_d on $[0, T]$ such that $A = (A_1, \dots, A_d)$ with*

$$A_j x = \int_0^T x_s d\mu_j(s) \quad \text{for } x \in \mathbb{D}(\mathbb{R}) \text{ and } j = 1, \dots, d$$

and generally suppose that $a'(X_{s-})\Delta Y_s \neq -1$ for all $s \in [0, T]$, almost surely. Then there exists a constant κ depending on F and the underlying SDE, but not on M and θ such that the variance ρ^2 appearing as variance is of the form

$$\rho = \kappa \Upsilon,$$

where as before $\Upsilon^2 = \frac{e^{-\theta} - 1 + \theta}{\theta^2} (1 - \frac{1}{M})$, if $\theta > 0$, and $\Upsilon^2 = \frac{1}{2} (1 - \frac{1}{M})$, if $\theta = 0$.

REMARK 1.12. The assumption that $a'(X_{s-})\Delta Y_s \neq -1$ for all $s \in [0, T]$, almost surely, is automatically fulfilled if ν has no atoms. For every $s \in (0, T]$ with $a'(X_{s-})\Delta Y_s = -1$, the error process jumps to zero causing technical difficulties in our proofs. In general, the result remains true without this assumption, but for simplicity we only provide a proof under this technical assumption.

PARAMETER OPTIMISATION 1.13. We use Theorem 1.11 to optimise the parameters. We assume that θ of (ML2) and the bias κ are zero. Multilevel schemes are based on iterated sampling of $F(X^k) - F(X^{k-1})$, where (X^{k-1}, X^k) are coupled approximate solutions. Typically, one simulation causes cost (has runtime) of order

$$C_k = (1 + o(1)) \kappa_{\text{cost}} \varepsilon_{k-1}^{-1} (M + \beta),$$

where κ_{cost} is a constant that does not depend on M , and $\beta \in \mathbb{R}$ is an appropriate constant typically with values between zero and one: one coupled path simulation needs:

- to simulate $\varepsilon_{k-1}^{-1} T M$ increments of the Lévy process,
- to do $\varepsilon_{k-1}^{-1} T M$ Euler steps to gain the fine approximation,
- to concatenate $\varepsilon_{k-1}^{-1} T (M - 1)$ Lévy increments, and
- to do $\varepsilon_{k-1}^{-1} T$ Euler steps to gain the coarse approximation.

If every operation causes the same computational cost, one ends up with $\beta = 0$. If the concatenation procedure is significantly less expensive, the parameter β rises. Using that

$$\delta^{-1} (\widehat{S}_\delta(F) - \mathbb{E}[F(X)]) \Rightarrow \mathcal{N}(0, \kappa_{\text{err}}^2 (1 - 1/M)) \quad \text{as } \delta \downarrow 0,$$

we conclude that for $\bar{\delta} := \bar{\delta}(\delta) := \delta / (\kappa_{\text{err}} \sqrt{1 - 1/M})$ one has

$$\delta^{-1}(\widehat{S}_{\bar{\delta}}(F) - \mathbb{E}[F(X)]) \Rightarrow \mathcal{N}(0, 1) \quad \text{as } \delta \downarrow 0.$$

Hence, the asymptotics of $\widehat{S}_{\bar{\delta}}(F)$ do not depend on the choice of M and we can compare the efficiency of different choices of M by looking at the cost of a simulation of $\widehat{S}_{\bar{\delta}}(F)$. It is of order

$$\begin{aligned} & (1 + o(1))\kappa_{\text{cost}}L(\bar{\delta})^2(M + \beta)\bar{\delta}^{-2} \\ &= (1 + o(1))\frac{\kappa_{\text{cost}}\kappa_{\text{err}}^2}{\alpha^2}\frac{(M - 1)(M + \beta)}{M(\log M)^2}\delta^{-2}(\log \delta^{-1})^2. \end{aligned}$$

A plot illustrating the dependence on the choice of M is provided in Figure 1. There we plot the function $M \mapsto \frac{(M-1)(M+\beta)}{M(\log M)^2}$ for β being 0 or 1. The plot indicates that in both cases 6 is a good choice for M . In particular, it is not necessary to know β explicitly in order to find a “good” M . For numerical tests concerning appropriate choices of β , we refer the reader to [11].

The article is outlined as follows. In Section 2, we analyse the error process and prove Theorem 1.5. In Section 3, we prepare the proofs of the central limit theorems for integral averages for the piecewise constant approximations and for supremum dependent functionals. In Section 4, we provide the proofs of all remaining theorems, in particular, of all central limit theorems. The article ends with an [Appendix](#) where we summarise known and auxiliary results. In particular, we provide a brief introduction to stable convergence and perturbation estimates mainly developed in articles by Jacod and Protter.

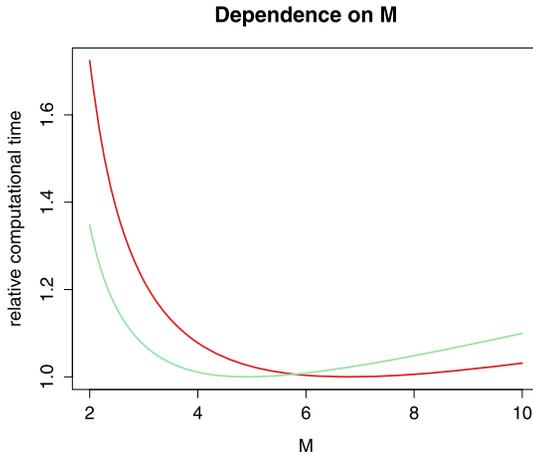


FIG. 1. Impact of M on the computational cost for $\beta = 0$ (green) and $\beta = 1$ (red).

2. The error process (Theorem 1.5). In this section, we prove Theorem 1.5. We assume that properties (ML1) and (ML2) are fulfilled. At first, we introduce the necessary notation and outline our strategy of proof. All intermediate results will be stated as propositions and their proofs are deferred to later subsections. We denote for $n \in \mathbb{N}$ and $t \in [0, T]$

$$\iota_n(t) = \sup[0, t] \cap \mathbb{I}_n,$$

where $\mathbb{I}_n = \{s \in (0, T] : \Delta Y_s^{h_n} \neq 0\} \cup (\varepsilon_n \mathbb{Z} \cap [0, T])$ is the random set of update times and recall that X^n solves

$$(2.1) \quad dX_t^n = a(X_{\iota_n(t-)}^n) dY_t$$

with $X_0^n = x_0$. We analyse the (normalised) *error process* of two consecutive X^n -levels that is the process $U^{n,n+1} = (U_t^{n,n+1})_{t \in [0, T]}$ given by

$$U_t^{n,n+1} = \varepsilon_n^{-1/2} (X^{n+1} - X^n).$$

The error process satisfies the SDE

$$\begin{aligned} dU_t^{n,n+1} &= \varepsilon_n^{-1/2} (a(X_{t-}^{n+1}) - a(X_{t-}^n)) dY_t + \varepsilon_n^{-1/2} (a(X_{\iota_n(t-)}^n) - a(X_{\iota_n(t-)}^n)) dY_t \\ &\quad - \varepsilon_n^{-1/2} (a(X_{\iota_{n+1}(t-)}^{n+1}) - a(X_{\iota_{n+1}(t-)}^n)) dY_t. \end{aligned}$$

In order to rewrite the SDE, we introduce some more notation. We let

$$\nabla a(u, v) = \begin{cases} \frac{a(v) - a(u)}{v - u}, & \text{if } u \neq v, \\ a'(u), & \text{if } u = v \end{cases}$$

for $u, v \in \mathbb{R}$ and consider the processes

$$\begin{aligned} (D_t^n) &= (\nabla a(X_{\iota_n(t)}^n, X_t^n)), & (D_t^{n,n+1}) &= (\nabla a(X_t^n, X_t^{n+1})), \\ (A_t^n) &= a(X_{\iota_n(t)}^n). \end{aligned}$$

In terms of the new notation, we have

$$(2.2) \quad \begin{aligned} dU_t^{n,n+1} &= D_{t-}^{n,n+1} U_{t-}^{n,n+1} dY_t + \varepsilon_n^{-1/2} D_{t-}^n A_{t-}^n (Y_{t-} - Y_{\iota_n(t-)}) dY_t \\ &\quad - \varepsilon_n^{-1/2} D_{t-}^{n+1} A_{t-}^{n+1} (Y_{t-} - Y_{\iota_{n+1}(t-)}) dY_t. \end{aligned}$$

Clearly, the processes (D_t^n) and $(D_t^{n,n+1})$ converge in ucp to $(D_t) := (a'(X_t))_{t \in [0, T]}$ and the processes (A_t^n) to $(A_t) := (a(X_t))_{t \in [0, T]}$. It often will be useful that the processes $D^{n,n+1}$ and D are uniformly bounded by the Lipschitz constant of the coefficient a .

For technical reasons, we introduce a further approximation. For every $\varepsilon > 0$, we denote by $U^{n,n+1,\varepsilon} = (U_t^{n,n+1,\varepsilon})_{t \in [0, T]}$ the solution of the SDE

$$(2.3) \quad \begin{aligned} dU_t^{n,n+1,\varepsilon} &= D_{t-} U_{t-}^{n,n+1,\varepsilon} dY_t \\ &\quad + \varepsilon_n^{-1/2} D_{t-} A_{t-} \sigma(W_{\iota_{n+1}(t-)} - W_{\iota_n(t-)}) dY_t^\varepsilon \end{aligned}$$

with $U_0^{n,n+1,\varepsilon} = 0$, where Y^ε is as in (1.5). Further, let $U^\varepsilon = (U_t^\varepsilon)_{t \in [0, T]}$ denote the solution of

$$(2.4) \quad \begin{aligned} U_t^\varepsilon = & \int_0^t D_{s-} U_{s-}^\varepsilon dY_s + \sigma^2 \Upsilon \int_0^t D_{s-} A_{s-} dB_s \\ & + \sum_{s \in (0, t]: \Delta Y_s^\varepsilon \neq 0} \sigma_s \xi_s D_{s-} A_{s-} \Delta Y_s^\varepsilon. \end{aligned}$$

We will show that the processes $U^\varepsilon, U^{1,2,\varepsilon}, U^{2,3,\varepsilon}, \dots$ are good approximations for the processes $U, U^{1,2}, U^{2,3}, \dots$ in the sense of Remark A.7. As a consequence of Lemma A.6, we then get:

PROPOSITION 2.1. *If for every $\varepsilon > 0$,*

$$(Y, U^{n,n+1,\varepsilon}) \Rightarrow (Y, U^\varepsilon) \quad \text{in } \mathbb{D}(\mathbb{R}^2),$$

then one has

$$(Y, U^{n,n+1}) \Rightarrow (Y, U) \quad \text{in } \mathbb{D}(\mathbb{R}^2).$$

The proof of the proposition is carried out in Section 2.1. It then remains to prove the following proposition which is the task of Section 2.2.

PROPOSITION 2.2. *For every $\varepsilon > 0$,*

$$(Y, U^{n,n+1,\varepsilon}) \Rightarrow (Y, U^\varepsilon) \quad \text{in } \mathbb{D}(\mathbb{R}^2).$$

2.1. *The approximations $U^{n,n+1,\varepsilon}$ are good.* In this subsection, we prove Proposition 2.1. By Lemma A.6, it suffices to show that the approximations are good in the sense of Remark A.7. In this section, we will work with an additional auxiliary process: for $n \in \mathbb{N}$ and $\varepsilon > 0$ we denote by $\bar{U}^{n,n+1,\varepsilon} := (\bar{U}_t^{n,n+1,\varepsilon})_{t \in [0, T]}$ the solution of

$$(2.5) \quad \begin{aligned} d\bar{U}_t^{n,n+1,\varepsilon} = & D_{t-}^{n,n+1} \bar{U}_{t-}^{n,n+1,\varepsilon} dY_t + \varepsilon_n^{-1/2} D_{t-}^n A_{t-}^n \sigma (W_{t-} - W_{l_n(t-)}) dY_t^\varepsilon \\ & - \varepsilon_n^{-1/2} D_{t-}^{n+1} A_{t-}^{n+1} (W_{t-} - W_{l_{n+1}(t-)}) dY_t^\varepsilon \end{aligned}$$

with $\bar{U}_0^{n,n+1,\varepsilon} = 0$.

LEMMA 2.3. *For every $\delta, \varepsilon > 0$, we have:*

1. $\lim_{\varepsilon \downarrow 0} \limsup_{n \rightarrow \infty} \mathbb{E}[\sup_{t \in [0, T]} |U_t^{n,n+1} - \bar{U}_t^{n,n+1,\varepsilon}|^2] = 0,$
2. $\lim_{n \rightarrow \infty} \mathbb{P}(\sup_{t \in [0, T]} |\bar{U}_t^{n,n+1,\varepsilon} - U_t^{n,n+1,\varepsilon}| > \delta) = 0,$
3. $\lim_{\varepsilon \downarrow 0} \mathbb{P}(\sup_{t \in [0, T]} |U_t - U_t^\varepsilon| > \delta) = 0.$

It is straightforward to verify that Lemma 2.3 implies that the approximations are good.

PROOF OF LEMMA 2.3. (1) Recalling (2.2) and (2.5) and noting that $D^{n,n+1}$ is uniformly bounded, we conclude with Lemma A.14 that the first statement is true if

$$(2.6) \quad \lim_{\varepsilon \downarrow 0} \limsup_{n \rightarrow \infty} \varepsilon_n^{-1} \mathbb{E} \left[\sup_{t \in [0, T]} \left| \int_0^t D_{s-}^n A_{s-}^n (Y_{s-} - Y_{l_n(s-)}) dY_s - \sigma \int_0^t D_{s-}^n A_{s-}^n (W_{s-} - W_{l_n(s-)}) dY_s^\varepsilon \right|^2 \right] = 0.$$

Let M^ε denote the martingale $Y - Y^\varepsilon$. The above term can be estimated against the sum of

$$(2.7) \quad \varepsilon_n^{-1} \mathbb{E} \left[\sup_{t \in [0, T]} \left| \int_0^t D_{s-}^n A_{s-}^n (Y_{s-} - Y_{l_n(s-)} - \sigma W_{s-} + \sigma W_{l_n(s-)}) dY_s \right|^2 \right]$$

and

$$(2.8) \quad \varepsilon_n^{-1} \sigma^2 \mathbb{E} \left[\sup_{t \in [0, T]} \left| \int_0^t D_{s-}^n A_{s-}^n (W_{s-} - W_{l_n(s-)}) dM_s^\varepsilon \right|^2 \right].$$

We start with estimating the former expression. For $t \in [0, T]$, one has

$$Y_t - Y_{l_n(t)} = \sigma(W_t - W_{l_n(t)}) + M_t^{h_n} - M_{l_n(t)}^{h_n} + \left(b - \int_{B(0, h_n)^c} x \nu(dx) \right) (t - l_n(t)).$$

By Lemma A.10, one has

$$\begin{aligned} & \varepsilon_n^{-1} \mathbb{E} [|Y_t - Y_{l_n(t)} - \sigma W_t + \sigma W_{l_n(t)}|^2 | l_n] \\ & \leq 2 \int_{B(0, h_n)} x^2 \nu(dx) + 2 \left(b - \int_{B(0, h_n)^c} x \nu(dx) \right)^2 \varepsilon_n \\ & =: \delta_n \rightarrow 0 \end{aligned}$$

as $n \rightarrow \infty$. Further, by Lemma A.11 and the uniform boundedness of D^n , there is a constant κ_1 not depending on n such that

$$(2.9) \quad \begin{aligned} & \varepsilon_n^{-1} \mathbb{E} \left[\sup_{t \in [0, T]} \left| \int_0^t D_{s-}^n A_{s-}^n (Y_{s-} - Y_{l_n(s-)} - \sigma W_{s-} + \sigma W_{l_n(s-)}) dY_s \right|^2 \right] \\ & \leq \kappa_1 \varepsilon_n^{-1} \int_0^T \mathbb{E} [(A_{s-}^n)^2 (Y_{s-} - Y_{l_n(s-)} - \sigma W_{s-} + \sigma W_{l_n(s-)})^2] ds \\ & \leq \kappa_1 \delta_n \int_0^T \mathbb{E} [|A_{s-}^n|^2] ds, \end{aligned}$$

where we have used conditional independence of A_{s-}^n and $Y_{s-} - Y_{\iota_n(s-)} - \sigma W_{s-} + \sigma W_{\iota_n(s-)}$ given ι_n in the last transformation. By Lemma A.12 and the Lipschitz continuity of a , the latter integral is uniformly bounded over all $n \in \mathbb{N}$ so that (2.7) tends to zero as $n \rightarrow \infty$.

Next, consider (2.8). Note that M^ε is a Lévy martingale with triplet $(0, 0, \nu|_{B(0,\varepsilon)})$. By Lemma A.11 and the uniform boundedness of D^n , there exists a constant κ_2 not depending on ε and n such that

$$\begin{aligned}
 (2.10) \quad & \varepsilon_n^{-1} \mathbb{E} \left[\sup_{t \in [0, T]} \left| \int_0^t D_{s-}^n A_{s-}^n (W_{s-} - W_{\iota_n(s-)}) dM_s^\varepsilon \right|^2 \right] \\
 & \leq \kappa_2 \varepsilon_n^{-1} \int_{B(0,\varepsilon)} x^2 \nu(dx) \int_0^T \mathbb{E}[|A_{s-}^n|^2 | W_{s-} - W_{\iota_n(s-)}|^2] ds \\
 & \leq \kappa_2 \int_{B(0,\varepsilon)} x^2 \nu(dx) \int_0^T \mathbb{E}[|A_{s-}^n|^2] ds,
 \end{aligned}$$

where we used in the last step that conditionally on ι_n the random variables A_{s-}^n and $W_{s-} - W_{\iota_n(s-)}$ are independent and $\mathbb{E}[(W_{s-} - W_{\iota_n(s-)})^2 | \iota_n] = s - \iota_n(s) \leq \varepsilon_n$. As noted above, $\int_0^T \mathbb{E}[|A_{s-}^n|^2] ds$ is uniformly bounded, and hence (2.8) tends uniformly to zero over all $n \in \mathbb{N}$ as $\varepsilon \downarrow 0$.

(2) We will use Lemma A.15 to prove that

$$(2.11) \quad \bar{U}^{n,n+1,\varepsilon} - U^{n,n+1,\varepsilon} \rightarrow 0 \quad \text{in ucp, as } n \rightarrow \infty.$$

We rewrite the SDE (2.3) as

$$\begin{aligned}
 dU_t^{n,n+1,\varepsilon} &= D_{t-} U_{t-}^{n,n+1,\varepsilon} dY_t + \varepsilon_n^{-1/2} D_{t-} A_{t-} \sigma (W_{t-} - W_{\iota_{n+1}(t-)}) dY_t^\varepsilon \\
 &\quad - \varepsilon_n^{-1/2} D_{t-} A_{t-} \sigma (W_{t-} - W_{\iota_{n+1}(t-)}) dY_t^\varepsilon.
 \end{aligned}$$

Recalling (2.5), it suffices by part one of Lemma A.15 to show that:

1. $D^{n,n+1} \rightarrow D$, in ucp,
2. $\varepsilon_n^{-1/2} \int_0^\cdot (D_{s-}^n A_{s-}^n - D_{s-} A_{s-}) (W_{s-} - W_{\iota_n(s-)}) dY_s^\varepsilon \rightarrow 0$, in ucp,
3. the families $(\sup_{t \in [0, T]} |D_t^{n,n+1}| : n \in \mathbb{N})$ and

$$\left(\varepsilon_n^{-1/2} \sup_{t \in [0, T]} \left| \int_0^t D_{s-}^n A_{s-}^n (W_{s-} - W_{\iota_n(s-)}) dY_s^\varepsilon \right| : n \in \mathbb{N} \right)$$

are tight.

The tightness of $(\sup_{t \in [0, T]} |D_t^{n,n+1}| : n \in \mathbb{N})$ follows by uniform boundedness. Further, the tightness of the second family follows by observing that in analogy to the proof of (1) one has

$$\begin{aligned}
 & \varepsilon_n^{-1} \mathbb{E} \left[\sup_{t \in [0, T]} \left| \int_0^t D_{s-}^n A_{s-}^n (W_{s-} - W_{\iota_n(s-)}) dY_s^\varepsilon \right|^2 \right] \\
 & \leq \kappa_3 \varepsilon_n^{-1} \int_0^T \mathbb{E}[|A_{s-}^n|^2 | W_{s-} - W_{\iota_n(s-)}|^2] ds \leq \kappa_3 \int_0^T \mathbb{E}[|A_{s-}^n|^2] ds
 \end{aligned}$$

for an appropriate constant κ_3 not depending on n . Furthermore, convergence $D^{n,n+1} \rightarrow D$ follows from ucp convergence of $X^n \rightarrow X$ and Lipschitz continuity of a . To show the remaining property, we let $\delta > 0$ and $T_{n,\delta}$ denote the stopping time

$$T_{n,\delta} = \inf\{s \in [0, T] : |D_s^n A_s^n - D_s A_s| \geq \delta\}.$$

Then by Lemma A.11, there exists a constant κ_4 not depending on n and δ with

$$\begin{aligned} & \mathbb{E} \left[\sup_{t \in [0, T \wedge T_{n,\delta}]} \varepsilon_n^{-1} \left(\int_0^t (D_{s-}^n A_{s-}^n - D_{s-} A_{s-}) (W_{s-} - W_{t_n(s-)}) dY_s^\varepsilon \right)^2 \right] \\ & \leq \kappa_4 \delta^2 \varepsilon_n^{-1} \int_0^T \mathbb{E}[(W_{s-} - W_{t_n(s-)})^2] ds \leq \kappa_4 \delta^2 T. \end{aligned}$$

Since for any $\delta > 0$, $\mathbb{P}(T_{n,\delta} = \infty) \rightarrow 1$ by ucp convergence $D^n A^n - DA \rightarrow 0$, we immediately get the remaining property by choosing $\delta > 0$ arbitrarily small and applying the Markov inequality.

(3) The proof of the third statement can be achieved by a simplified version of the proof of the first statement. It is therefore omitted. \square

2.2. Weak convergence of $U^{n,n+1,\varepsilon}$. In this subsection, we prove Proposition 2.2 for fixed $\varepsilon > 0$. We first outline the proof. We will make use of results of [17] summarised in the Appendix; see Section A.1. We consider processes $Z^{n,\varepsilon} = (Z_t^{n,\varepsilon})_{t \in [0, T]}$ and $Z^\varepsilon = (Z_t^\varepsilon)_{t \in [0, T]}$ given by

$$(2.12) \quad Z_t^{n,\varepsilon} = \varepsilon_n^{-1/2} \int_0^t (W_{t_{n+1}(s-)} - W_{t_n(s-)}) dY_s^\varepsilon$$

and

$$(2.13) \quad Z_t^\varepsilon = \Upsilon B_t + \sum_{s \in (0, t] : |\Delta Y_s| \geq \varepsilon} \frac{\sigma_s}{\sigma} \xi_s \Delta Y_s,$$

where (σ_s) and (ξ_s) are the marks of the point process Π as introduced in Section 1.1.

In view of Theorem A.5, the statement of Proposition 2.2 follows, if we show that

$$\left(Y, \int_0^\cdot D_{t-} dY_t, \int_0^\cdot D_{t-} A_{t-} dZ_t^{n,\varepsilon} \right) \Rightarrow \left(Y, \int_0^\cdot D_{t-} dY_t, \int_0^\cdot D_{t-} A_{t-} dZ_t^\varepsilon \right) \quad \text{in } \mathbb{D}(\mathbb{R}^3).$$

Further, by Theorem A.4, this statement follows once we showed that $(Z^{n,\varepsilon} : n \in \mathbb{N})$ is uniformly tight and

$$(2.14) \quad (Y, D, DA, Z^{n,\varepsilon}) \Rightarrow (Y, D, DA, Z^\varepsilon) \quad \text{in } \mathbb{D}(\mathbb{R}^4).$$

We first prove that $((Y, D, DA, Z^{n,\varepsilon}) : n \in \mathbb{N})$ is tight which shows that, in particular, $(Z^{n,\varepsilon} : n \in \mathbb{N})$ is uniformly tight; see Lemma 2.4. Note that (Y, D, DA) is $\sigma(Y)$ -measurable. To identify the limit and complete the proof of (2.14), it suffices to prove stable convergence

$$Z^{n,\varepsilon} \xrightarrow{\text{stably}} Z^\varepsilon$$

with respect to the σ -field $\sigma(Y)$; see Section A.1 in the Appendix for a brief introduction of stable convergence. The latter statement is equivalent to

$$(Y, Z^{n,\varepsilon}) \Rightarrow (Y, Z^\varepsilon) \quad \text{in } \mathbb{D}(\mathbb{R}) \times \mathbb{D}(\mathbb{R}),$$

by Theorem A.2. We prove the stronger statement that this is even true in the finer topology $\mathbb{D}(\mathbb{R}^2)$: the sequence $((Y, Z^{n,\varepsilon}) : n \in \mathbb{N})$ is tight by Lemma 2.4 and we will prove convergence of finite-dimensional marginals in Lemma 2.6. The proof of the latter lemma is based on a perturbation result provided by Lemma 2.5.

LEMMA 2.4. *For $\varepsilon > 0$, the family $((Y, D, DA, Z^{n,\varepsilon}) : n \in \mathbb{N})$ taking values in $\mathbb{D}(\mathbb{R}^4)$ is tight. In particular, $(Z^{n,\varepsilon} : n \in \mathbb{N})$ is uniformly tight.*

PROOF. One has by Lemma A.11

$$\mathbb{E} \left[\sup_{t \in [0, T]} (Z_t^{n,\varepsilon})^2 \right] \leq \kappa_1 \varepsilon_n^{-1} \int_0^t \mathbb{E}[(W_{l_{n+1}(t-)} - W_{l_n(t-)})^2] dt \leq \kappa_1$$

for an appropriate constant κ_1 so that by the Markov inequality

$$\lim_{K \rightarrow \infty} \sup_{n \in \mathbb{N}} \mathbb{P} \left(\sup_{t \in [0, T]} |Y_t| \vee |Z_t^{n,\varepsilon}| \vee |D_t| \vee |D_t A_t| \geq K \right) = 0.$$

It remains to verify Aldous' criterion for tightness [18], Theorem VI.4.5, which can be checked componentwise. It is certainly fulfilled for Y, A and DA and it remains to show that for every $K > 0$ there exists for every $\delta > 0$ a constant $c_\delta > 0$ such that for arbitrary stopping times S_1, S_2, \dots

$$\limsup_{n \rightarrow \infty} \mathbb{P} \left(\sup_{t \in [S_n, (S_n + \delta) \wedge T]} |Z_t^{n,\varepsilon} - Z_{S_n}^{n,\varepsilon}| \geq K \right) \leq c_\delta$$

and $\lim_{\delta \downarrow 0} c_\delta = 0$.

First, suppose that S_1, S_2, \dots denote stopping times taking values in the respective sets $\varepsilon_n \mathbb{Z}$. Then as above

$$\begin{aligned} & \mathbb{E} \left[\sup_{t \in [S_n, (S_n + \delta) \wedge T]} |Z_t^{n,\varepsilon} - Z_{S_n}^{n,\varepsilon}|^2 \right] \\ & \leq \kappa_1 \varepsilon_n^{-1} \int_0^T \mathbb{E}[\mathbb{1}_{[S_n, (S_n + \delta)]}(t) (W_{l_{n+1}(t-)} - W_{l_n(t-)})^2] dt \\ (2.15) \quad & \leq \kappa_1 \varepsilon_n^{-1} \int_0^T \mathbb{E}[\mathbb{1}_{[S_n, (S_n + \delta)]}(l_n(t)) (W_{l_{n+1}(t-)} - W_{l_n(t-)})^2] dt \end{aligned}$$

$$\begin{aligned} &\leq \kappa_1 \mathbb{E} \left[\int_0^T \mathbb{1}_{[S_n, (S_n + \delta)]}(\iota_n(t)) \, dt \right] \\ &\leq \kappa_1 (\varepsilon_n + \delta) \rightarrow \kappa_1 \delta, \end{aligned}$$

where we have used that $\mathbb{E}[(W_{\iota_{n+1}(t-)} - W_{\iota_n(t-)})^2 | \mathcal{F}_{\iota_n(t)}] \leq \varepsilon_n$ and $\mathbb{1}_{[S_n, (S_n + \delta)]}(t)$ is $\mathcal{F}_{\iota_n(t)}$ -measurable. It remains to estimate for general stopping times S_1, S_2, \dots

$$\mathbb{E} \left[\sup_{t \in [S_n, \bar{S}_n]} |Z_t^{n, \varepsilon} - Z_{S_n}^{n, \varepsilon}|^2 \right],$$

where $\bar{S}_n = \inf[S_n, \infty) \cap \varepsilon_n \mathbb{Z}$. As in (2.15), we conclude with $\bar{S}_n - S_n \leq \varepsilon$ that

$$\begin{aligned} &\mathbb{E} \left[\sup_{t \in [S_n, \bar{S}_n]} |Z_t^{n, \varepsilon} - Z_{S_n}^{n, \varepsilon}|^2 \right] \\ &\leq \kappa_1 \varepsilon_n^{-1} \mathbb{E} \left[\int_0^T \mathbb{1}_{[S_n, \bar{S}_n]}(t) (W_{\iota_{n+1}(t-)} - W_{\iota_n(t-)})^2 \, dt \right] \\ &\leq \kappa_1 \mathbb{E} \left[\sup_{\substack{k=1, \dots, \varepsilon_n^{-1} \\ s, t \in [(k-1)\varepsilon_n^{-1}, k\varepsilon_n^{-1}]} |W_s - W_t|^2 \right] \rightarrow 0. \end{aligned}$$

By the Markov inequality, this estimate together with (2.15) imply Aldous' criterion. \square

To control perturbations, we will use the following lemma.

LEMMA 2.5. For $j = 1, 2$, let $(\alpha_t^{(j)})_{t \in [0, T]}$ and $(\beta_t^{(j)})_{t \in [0, t]}$ optional processes being square integrable with respect to $\mathbb{P} \otimes \ell_{[0, T]}$ and let

$$\Upsilon_t^{n, j} = \varepsilon_n^{-1/2} \int_0^t (\bar{W}_{\iota_{n+1}(s-)}^{(j)} - \bar{W}_{\iota_n(s-)}^{(j)}) \, d\bar{Y}_s^{(j)},$$

where

$$\bar{W}_t^{(j)} = W_t + \int_0^t \alpha_s^{(j)} \, ds, \quad \bar{Y}_t^{(j)} = M_t + \int_0^t \beta_s^{(j)} \, ds$$

and

$$M_t = \sigma W_t + \int_{(0, t] \times B(0, \varepsilon)^c} x \, d\bar{\Pi}(s, x).$$

For $t \in \mathcal{D} = \bigcup_{n \in \mathbb{N}} \varepsilon_n \mathbb{Z} \cap [0, T]$, the sequences $(\Upsilon_t^{n, 1})_{n \in \mathbb{N}}$ and $(\Upsilon_t^{n, 2})_{n \in \mathbb{N}}$ are equivalent in probability, that is, for every $\delta > 0$

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\Upsilon_t^{n, 1} - \Upsilon_t^{n, 2}| > \delta) = 0.$$

PROOF. We prove the statement in three steps.

1st step. First, we show a weaker perturbation estimate. Using the bilinearity of the stochastic integral, we get that

$$\begin{aligned}
 \Upsilon_t^{n,1} - \Upsilon_t^{n,2} &= \varepsilon_n^{-1/2} \int_0^t \int_{\iota_n(s-)}^{\iota_{n+1}(s-)} (\alpha_u^{(1)} - \alpha_u^{(2)}) du dM_s \\
 &\quad + \varepsilon_n^{-1/2} \int_0^t (W_{\iota_{n+1}(s-)} - W_{\iota_n(s-)})(\beta_s^{(1)} - \beta_s^{(2)}) ds \\
 &\quad + \varepsilon_n^{-1/2} \int_0^t \int_{\iota_n(s-)}^{\iota_{n+1}(s-)} (\alpha_u^{(1)} - \alpha_u^{(2)}) du \beta_s^{(1)} ds \\
 &\quad + \varepsilon_n^{-1/2} \int_0^t \int_{\iota_n(s-)}^{\iota_{n+1}(s-)} \alpha_u^{(2)} du (\beta_s^{(1)} - \beta_s^{(2)}) ds.
 \end{aligned}
 \tag{2.16}$$

We analyse the terms individually. By Itô's isometry, the fact that $s - \varepsilon_n \leq \iota_n(s-) \leq \iota_{n+1}(s-) \leq s$ and Fubini's theorem one has that for $\kappa = \sigma^2 + \int_{B(0,\varepsilon)^c} x^2 \nu(dx)$

$$\begin{aligned}
 &\mathbb{E} \left[\left(\varepsilon_n^{-1/2} \int_0^t \int_{\iota_n(s-)}^{\iota_{n+1}(s-)} (\alpha_u^{(1)} - \alpha_u^{(2)}) du dM_s \right)^2 \right] \\
 &= \kappa \varepsilon_n^{-1} \mathbb{E} \left[\int_0^t \left(\int_{\iota_n(s-)}^{\iota_{n+1}(s-)} (\alpha_u^{(1)} - \alpha_u^{(2)}) du \right)^2 ds \right] \\
 &\leq \kappa \mathbb{E} \left[\int_0^t \int_{\iota_n(s-)}^{\iota_{n+1}(s-)} (\alpha_u^{(1)} - \alpha_u^{(2)})^2 du ds \right] \\
 &\leq \kappa \mathbb{E} \left[\int_0^t \int_{(s-\varepsilon_n) \vee 0}^s (\alpha_u^{(1)} - \alpha_u^{(2)})^2 du ds \right] \\
 &\leq \kappa \varepsilon_n \mathbb{E} \left[\int_0^t (\alpha_s^{(1)} - \alpha_s^{(2)})^2 ds \right].
 \end{aligned}
 \tag{2.17}$$

By the Cauchy–Schwarz inequality and Fubini, it follows that the second term satisfies

$$\begin{aligned}
 &\mathbb{E} \left[\varepsilon_n^{-1/2} \left| \int_0^t (W_{\iota_{n+1}(s-)} - W_{\iota_n(s-)})(\beta_s^{(1)} - \beta_s^{(2)}) ds \right| \right] \\
 &\leq \varepsilon_n^{-1/2} \mathbb{E} \left[\int_0^t (W_{\iota_{n+1}(s-)} - W_{\iota_n(s-)})^2 ds \right]^{1/2} \mathbb{E} \left[\int_0^t (\beta_s^{(1)} - \beta_s^{(2)})^2 ds \right]^{1/2} \\
 &\leq t \mathbb{E} \left[\int_0^t (\beta_s^{(1)} - \beta_s^{(2)})^2 ds \right]^{1/2},
 \end{aligned}$$

where we have used in the last step that $\iota_{s-}^{n+1} - \iota_{s-}^n$ is independent of the Brownian motion and smaller or equal to ε_n . The third term is estimated similarly as the first

term:

$$\begin{aligned}
& \mathbb{E} \left[\varepsilon_n^{-1/2} \left| \int_0^t \int_{\iota_n(s-)}^{\iota_{n+1}(s-)} (\alpha_u^{(1)} - \alpha_u^{(2)}) \, du \, \beta_s^{(1)} \, ds \right| \right] \\
& \leq \varepsilon_n^{1/2} \mathbb{E} \left[\int_0^t \left(\int_{\iota_n(s-)}^{\iota_{n+1}(s-)} (\alpha_u^{(1)} - \alpha_u^{(2)}) \, du \right)^2 \, ds \right]^{1/2} \mathbb{E} \left[\int_0^T (\beta_s^{(1)})^2 \, ds \right]^{1/2} \\
& \leq \mathbb{E} \left[\int_0^t \int_{\iota_n(s-)}^{\iota_{n+1}(s-)} (\alpha_u^{(1)} - \alpha_u^{(2)})^2 \, du \, ds \right]^{1/2} \mathbb{E} \left[\int_0^T (\beta_s^{(1)})^2 \, ds \right]^{1/2} \\
& \leq \varepsilon_n^{1/2} \mathbb{E} \left[\int_0^t (\alpha_s^{(1)} - \alpha_s^{(2)})^2 \, ds \right]^{1/2} \mathbb{E} \left[\int_0^T (\beta_s^{(1)})^2 \, ds \right]^{1/2}.
\end{aligned}$$

In complete analogy, the fourth term satisfies

$$\begin{aligned}
& \mathbb{E} \left[\varepsilon_n^{-1/2} \left| \int_0^t \int_{\iota_n(s-)}^{\iota_{n+1}(s-)} \alpha_u^{(2)} \, du (\beta_s^{(1)} - \beta_s^{(2)}) \, ds \right| \right] \\
& \leq \varepsilon_n^{1/2} \mathbb{E} \left[\int_0^t (\alpha_s^{(2)})^2 \, ds \right]^{1/2} \mathbb{E} \left[\int_0^T (\beta_s^{(1)} - \beta_s^{(2)})^2 \, ds \right]^{1/2}.
\end{aligned}$$

By the Markov inequality, the first, third and fourth term of (2.16) tend to zero in probability as $n \rightarrow \infty$.

2nd step. Next, we analyse the case where $\beta^{(2)} = 0$ and $\beta := \beta^{(1)}$ is simple in the following sense. There exist $l \in \mathbb{N}$, increasingly ordered times $0 = t_0, t_1, \dots, t_l = t \in \mathcal{D} = \bigcup_{n \in \mathbb{N}} \varepsilon_n \mathbb{Z} \cap [0, T]$ such that β is almost surely constant on each of the time intervals $[t_0, t_1), \dots, [t_{l-1}, t_l)$. For $n \in \mathbb{N}$ and $j = 1, \dots, l$, we let

$$M_{j,n} := \varepsilon_n^{-1/2} \int_{t_{j-1}}^{t_j} (W_{\iota_{n+1}(s-)} - W_{\iota_n(s-)}) \, ds.$$

We suppose that $n \in \mathbb{N}$ is sufficiently large to ensure that $\{t_1, \dots, t_l\} \subset \varepsilon_n \mathbb{Z}$. The Brownian motion W is independent of Π so that for $u, s \in [0, t]$

$$\begin{aligned}
& \mathbb{E}[(W_{\iota_{n+1}(s-)} - W_{\iota_n(s-)})(W_{\iota_{n+1}(u-)} - W_{\iota_n(u-)})|\Pi] \\
& = \ell([\iota_n(s-), \iota_{n+1}(s-)] \cap [\iota_n(u-), \iota_{n+1}(u-)]) \\
& \leq \varepsilon_n \mathbb{1}_{\{|s-u| \leq \varepsilon_n\}}.
\end{aligned}$$

Consequently, we obtain with Fubini that

$$\begin{aligned}
\mathbb{E}[M_{j,n}^2] &= \varepsilon_n^{-1} \mathbb{E} \left[\int_{t_{j-1}}^{t_j} \int_{t_{j-1}}^{t_j} (W_{\iota_{n+1}(s-)} - W_{\iota_n(s-)})(W_{\iota_{n+1}(u-)} - W_{\iota_n(u-)}) \, ds \, du \right] \\
&\leq 2\varepsilon_n(t_j - t_{j-1}).
\end{aligned}$$

Since $M_{j,n}$ is independent of $\mathcal{F}_{t_{j-1}}$ and has mean zero, we conclude that $(\sum_{j=1}^k \beta_{t_{j-1}} M_{j,n})_{k=0,\dots,l}$ is a square integrable martingale so that

$$\begin{aligned} & \mathbb{E}\left[\left(\varepsilon_n^{-1/2} \int_0^t (W_{t_{n+1}(s-)} - W_{t_n(s-)}) \beta_s \, ds\right)^2\right] \\ &= \sum_{j=1}^l \mathbb{E}[\beta_{t_{j-1}}^2 M_{j,n}^2] = \sum_{j=1}^l \mathbb{E}[\beta_{t_{j-1}}^2] \mathbb{E}[M_{j,n}^2] \\ &\leq 2\varepsilon_n \sum_{j=1}^l \beta_{t_{j-1}}^2 (t_j - t_{j-1}) = 2\varepsilon_n \mathbb{E}\left[\int_0^t \beta_s^2 \, ds\right]. \end{aligned}$$

3rd step. We combine the first and second step. Let $\alpha^{(2)}$ and $\beta^{(2)}$ be as in the statement of the theorem and let $\delta > 0$ be arbitrary. The simple functions as defined in step two are dense in the space of previsible processes with finite L^2 -norm with respect to $\mathbb{P} \otimes \ell_{[0,T]}$. By part one, we can choose $\alpha^{(1)} = 0$ and a simple process $\beta^{(1)}$ such that

$$\mathbb{P}(|\Upsilon_t^{n,2} - \Upsilon_t^{n,1}| \geq \delta/2) \leq \delta/2$$

for n sufficiently large. Next, let $\Upsilon^{n,0}$ denote the process that is obtained in analogy to $\Upsilon^{n,1}$ and $\Upsilon^{n,2}$ when choosing $\alpha = \beta = 0$. By the second step, $(\Upsilon_t^{n,1} : n \in \mathbb{N})$ and $(\Upsilon_t^{n,0} : n \in \mathbb{N})$ are asymptotically equivalent in probability implying that

$$\mathbb{P}(|\Upsilon_t^{n,1} - \Upsilon_t^{n,0}| \geq \delta/2) \leq \delta/2$$

for sufficiently large $n \in \mathbb{N}$. Altogether, we arrive at

$$\mathbb{P}(|\Upsilon_t^{n,2} - \Upsilon_t^{n,0}| \geq \delta) \leq \delta$$

for sufficiently large $n \in \mathbb{N}$. Since $\delta > 0$ is arbitrary, $(\Upsilon_t^{n,2} : n \in \mathbb{N})$ and $(\Upsilon_t^{n,0} : n \in \mathbb{N})$ are equivalent in probability. The general statement follows by transitivity of equivalence in probability. \square

LEMMA 2.6. *For any finite subset $\mathbb{T} \subset \mathcal{D} = \bigcup_{n \in \mathbb{N}} \varepsilon_n \mathbb{N}_0$, one has convergence*

$$(Y_t, Z_t^{n,\varepsilon})_{t \in \mathbb{T}} \Rightarrow (Y_t, Z_t^\varepsilon)_{t \in \mathbb{T}}.$$

PROOF. *1st step.* In the first step, we derive a simpler sufficient criterion which implies the statement. Fix $l \in \mathbb{N}$, increasing times $0 = t_0 \leq t_1 < \dots < t_l \leq T$ and consider $\mathbb{T} = \{t_1, \dots, t_l\}$. The statement follows if for $A \in \sigma(Y_t : t \in \mathbb{T})$ and continuous compactly supported $f : \mathbb{R}^l \rightarrow \mathbb{R}$

$$\mathbb{E}[\mathbb{1}_A f(Z_{t_1}^{n,\varepsilon}, \dots, Z_{t_l}^{n,\varepsilon})] \rightarrow \mathbb{E}[\mathbb{1}_A f(Z_{t_1}^\varepsilon, \dots, Z_{t_l}^\varepsilon)].$$

By the Stone–Weierstrass theorem, the linear hull of functions of the form

$$\mathbb{R}^l \rightarrow \mathbb{R}, \quad x \mapsto f_1(x_1) \times \dots \times f_l(x_l)$$

with continuous compactly supported functions $f_1, \dots, f_l: \mathbb{R} \rightarrow \mathbb{R}$ is dense in the space of compactly supported continuous functions on \mathbb{R}^l equipped with supremum norm. Hence, it suffices to verify that

$$(2.18) \quad \begin{aligned} & \mathbb{E}[\mathbb{1}_A f_1(Z_{t_1}^{n,\varepsilon}) \cdots f_l(Z_{t_l}^{n,\varepsilon} - Z_{t_{l-1}}^{n,\varepsilon})] \\ & \rightarrow \mathbb{E}[\mathbb{1}_A f_1(Z_{t_1}^\varepsilon) \cdots f_l(Z_{t_l}^\varepsilon - Z_{t_{l-1}}^\varepsilon)] \end{aligned}$$

for arbitrary continuous compactly supported functions $f_1, \dots, f_l: \mathbb{R} \rightarrow \mathbb{R}$.

For fixed set \mathbb{T} , the family of sets $A \in \sigma(Y_t: t \in \mathbb{T})$ for which (2.18) is valid is a Dynkin system provided that the statement is true for $A = \Omega$. Consequently, it suffices to prove (2.18) on the \cap -stable generator

$$\mathcal{E} = \{A_1 \cap \cdots \cap A_l: A_0 \in \mathcal{A}_0, \dots, A_l \in \mathcal{A}_l\},$$

where $\mathcal{A}_1 = \sigma(Y_{t_1}), \dots, \mathcal{A}_l = \sigma(Y_{t_l} - Y_{t_{l-1}})$. We note that for $A = A_1 \cap \cdots \cap A_l \in \mathcal{E}$ the random variables

$$\mathbb{1}_{A_1} f_1(Z_{t_1}^{n,\varepsilon}), \dots, \mathbb{1}_{A_l} f_l(Z_{t_l}^{n,\varepsilon} - Z_{t_{l-1}}^{n,\varepsilon})$$

are independent if $\mathbb{T} \subset \varepsilon_n \mathbb{N}_0$ which is fulfilled for sufficiently large n since \mathbb{T} is finite and a subset of \mathcal{D} . Likewise this holds for $(Z_t^{n,\varepsilon})$ replaced by (Z_t^ε) . Consequently, it suffices to prove that for $k = 1, \dots, l$

$$\mathbb{E}[\mathbb{1}_{A_k} f_k(Z_{t_k}^{n,\varepsilon} - Z_{t_{k-1}}^{n,\varepsilon})] \rightarrow \mathbb{E}[\mathbb{1}_{A_k} f_k(Z_{t_k}^\varepsilon - Z_{t_{k-1}}^\varepsilon)].$$

Due to the time homogeneity of the problem, we can and will restrict attention to the case $k = 1$ and set $t = t_1$. Note that $\sigma(W) \cap \bigcup_{\varepsilon' > 0} \sigma(\sum_{s \in (0,t]: |\Delta Y_s| \geq \varepsilon'} \delta_{\Delta Y_s})$ is \cap -stable, contains Ω and generates a σ -field that contains $\sigma(Y_t)$.

We conclude that the statement of the lemma is true, if for all $t \in \mathcal{D}$, $\varepsilon' > 0$, all $A \in \sigma(W)$ and $A' \in \sigma(\sum_{s \in (0,t]: |\Delta Y_s| \geq \varepsilon'} \delta_{\Delta Y_s})$ and all continuous compactly supported $f: \mathbb{R} \rightarrow \mathbb{R}$, one has

$$(2.19) \quad \lim_{n \rightarrow \infty} \mathbb{E}[\mathbb{1}_{A \cap A'} f(Z_t^{n,\varepsilon})] = \mathbb{E}[\mathbb{1}_{A \cap A'} f(Z_t^\varepsilon)].$$

2nd step. In this step, we prove that for $A \in \sigma(W)$ and $A' \in \sigma(\Pi)$

$$\lim_{n \rightarrow \infty} |\mathbb{E}[\mathbb{1}_{A \cap A'} f(Z_t^{n,\varepsilon})] - \mathbb{P}(A) \mathbb{E}[\mathbb{1}_{A'} f(\bar{Z}_t^{n,\varepsilon})]| = 0,$$

where (\bar{Y}_s^ε) and $(\bar{Z}_s^{n,\varepsilon})$ are given by

$$\bar{Y}_s^\varepsilon = \sigma W_s + \int_{(0,s] \times B(0,\varepsilon)^c} x \, d\Pi(u, x)$$

and

$$\bar{Z}_s^{n,\varepsilon} = \varepsilon_n^{-1/2} \int_0^s (W_{l_{n+1}(u-)} - W_{l_n(u-)}) \, d\bar{Y}_u^\varepsilon.$$

It suffices to consider the case $\mathbb{P}(A) > 0$. We use results of enlargements of filtrations; see [19], Theorem 2, page 47, or [1], Example 2: there exists a previsible

process $(\alpha_s)_{s \in [0, T]}$ being square integrable with respect to $\mathbb{P} \otimes \ell_{[0, T]}$ such that given A the process $(W_s^A)_{s \in [0, T]}$

$$W_s^A := W_s - \int_0^s \alpha_u \, du$$

is a Wiener process. By Lemma 2.5, the processes $(Z_t^{n, \varepsilon})$ and

$$\bar{Z}_s^{n, \varepsilon, A} = \varepsilon_n^{-1/2} \int_0^s (W_{l_{n+1}(u-)}^A - W_{l_n(u-)}^A) \, d\bar{Y}_u^{\varepsilon, A}$$

with $\bar{Y}^{\varepsilon, A} = (\sigma W_s^A + \int_{(0, s] \times B(0, \varepsilon)^c} x \, d\Pi(u, x))_{s \in [0, T]}$ are equivalent in probability. Hence,

$$|\mathbb{E}[\mathbb{1}_{A \cap A'} f(Z_t^{n, \varepsilon})] - \mathbb{E}[\mathbb{1}_{A \cap A'} f(\bar{Z}_t^{n, \varepsilon, A})]| \rightarrow 0.$$

The set A is independent of Π . Further, conditionally on A the process W^A is a Brownian motion that is independent of Π which implies that

$$\mathbb{E}[\mathbb{1}_{A \cap A'} f(\bar{Z}_t^{n, \varepsilon, A})] = \mathbb{P}(A) \mathbb{E}[\mathbb{1}_{A'} f(\bar{Z}_t^{n, \varepsilon})].$$

3rd step. Let Γ denote the finite Poisson point process on $B(0, \varepsilon')^c$ with

$$\Gamma = \sum_{\substack{s \in (0, t] \\ |\Delta Y_s| \geq \varepsilon'}} \delta_{\Delta Y_s} = \int_{(0, t] \times B(0, \varepsilon)^c} \delta_x \, d\Pi(u, x).$$

In the third step, we prove that for every $A' \in \sigma(\Gamma)$ and every continuous and bounded function $f: \mathbb{R} \rightarrow \mathbb{R}$ one has

$$\lim_{n \rightarrow \infty} \mathbb{E}[\mathbb{1}_{A'} f(\bar{Z}_t^{n, \varepsilon})] = \mathbb{E}[\mathbb{1}_{A'} f(Z_t^\varepsilon)].$$

By dominated convergence, it suffices to show that, almost surely,

$$(2.20) \quad \lim_{n \rightarrow \infty} \mathbb{E}[f(\bar{Z}_t^{n, \varepsilon}) | \Gamma] = \mathbb{E}[f(Z_t^\varepsilon) | \Gamma].$$

The regular conditional probability of $\Pi|_{(0, t] \times B(0, \varepsilon')^c}$ given Γ can be made precise: the distribution of $\Pi|_{(0, t] \times B(0, \varepsilon')^c}$ given $\{\Gamma = \gamma := \sum_{k=1}^m \delta_{y_k}\}$ with $m \in \mathbb{N}$ and $y_1, \dots, y_m \in B(0, \varepsilon')^c$ is the same as the distribution of

$$\sum_{k=1}^m \delta_{S_k, y_k}$$

with independent on $(0, t]$ uniformly distributed random variables S_1, \dots, S_m . Since, furthermore, $\Pi|_{(0, t] \times B(0, \varepsilon')^c}$ is independent of $\Pi|_{(0, t] \times B(0, \varepsilon') \setminus \{0\}}$ and the Brownian motion W , we conclude that the distribution of $\bar{Z}_t^{n, \varepsilon}$ conditioned on $\{\Gamma = \gamma\}$ equals the distribution of the random variable

$$\bar{Z}_t^{n, \varepsilon, \gamma} = \varepsilon_n^{-1/2} \int_0^t (W_{l_{n+1}^\gamma(u-)} - W_{l_n^\gamma(u-)}^{\gamma}) \, d\bar{Y}_u^{n, \varepsilon}$$

with $\overline{Y}_s^{n,\gamma} = \sigma W_s + \sum_{k=1}^m y_m \mathbb{1}_{\{|y_m| \geq \varepsilon\}} \mathbb{1}_{\{S_k \leq s\}}$ and

$$t_n^\gamma(s) = \sup[(\varepsilon_n \mathbb{Z} \cap [0, t]) \cup \{s \in (0, t] : h_n \leq |\Delta Y_s| < \varepsilon'\} \cup \{S_1, \dots, S_m\}].$$

Here, the random variables S_1, \dots, S_m are independent of $\Pi|_{(0,t] \times B(0,\varepsilon')}$ and W . Likewise the random variable Z_t^ε given $\{\Gamma = \gamma\}$ has the same distribution as the unconditional random variable

$$Z_t^{\varepsilon,\gamma} = \Upsilon B_t + \sum_{j=1}^m \frac{\sigma_j}{\sigma} \xi_j y_j \mathbb{1}_{\{|y_j| \geq \varepsilon\}}$$

with $\sigma_1, \dots, \sigma_m$ and ξ_1, \dots, ξ_m being independent (also of B) with the same distribution as the marks of the point process Π . Consequently, statement (2.20) follows if for every γ as above,

$$\lim_{n \rightarrow \infty} \mathbb{E}[f(\overline{Z}_t^{n,\varepsilon,\gamma})] = \mathbb{E}[f(Z_t^{\varepsilon,\gamma})].$$

We keep γ fixed and analyse $\overline{Z}_t^{n,\varepsilon,\gamma}$ for $n \in \mathbb{N}$ sufficiently large, that is, with $t \in \varepsilon_n \mathbb{Z}$. We partition $(0, t]$ into t/ε_n n -windows. We call the k th n -window to be *occupied* by S_j if S_j is the only time in the window $((k-1)\varepsilon_n, k\varepsilon_n]$. Further, we call a window to be *empty*, if none of the times S_1, \dots, S_m is in the window. For each window $k = 1, \dots, t/\varepsilon_n$ that is empty, we set

$$\mathcal{Z}_k^{n,\gamma} = \varepsilon_n^{-1/2} \sigma \int_{(k-1)\varepsilon_n}^{k\varepsilon_n} (W_{t_{n+1}^\gamma(u-)} - W_{t_n^\gamma(u-)}^{\gamma}) dW_u,$$

and for a window $((k-1)\varepsilon_n, k\varepsilon_n]$ being occupied by j

$$\mathcal{Z}_k^{n,\gamma} = \varepsilon_n^{-1/2} (W_{t_{n+1}^\gamma(S_j-)}^{\gamma} - W_{t_n^\gamma(S_j-)}^{\gamma}) y_j \mathbb{1}_{\{|y_j| \geq \varepsilon\}}.$$

The remaining $\mathcal{Z}_k^{n,\gamma}$ can be defined arbitrarily since we will make use of the fact that the event \mathcal{T}_n that all windows are either empty or occupied satisfies $\mathbb{P}(\mathcal{T}_n) \rightarrow 1$.

We first analyse the contribution of the occupied windows. Given that \mathcal{T}_n occurs and that S_1, \dots, S_m are in windows k_1, \dots, k_m , the random variables $\mathcal{Z}_{k_1}^{n,\gamma}, \dots, \mathcal{Z}_{k_m}^{n,\gamma}$ are independent. We consider their conditional distributions: conditionally, each S_j is uniformly distributed on the respective window and the last displacement in $B(0, \varepsilon') \setminus B(0, h_n)$, respectively, $B(0, h_n) \setminus B(0, h_{n+1})$ has occurred an independent exponentially distributed amount of time ago; with parameter $\lambda_n = \nu(B(0, \varepsilon') \setminus B(0, h_n))$, respectively, $\lambda_{n+1} - \lambda_n$. Therefore, the conditional distribution of $(S_j - t_n(S_j), S_j - t_{n+1}(S_j))$ is the same as the one of

$$\left(\min(\mathcal{U}^{\varepsilon_n}, \mathcal{E}^{\lambda_n}), \sum_{i=1}^M \mathbb{1}_{((i-1)\varepsilon_n, i\varepsilon_n]}(\mathcal{U}^{\varepsilon_n}) \min\left(\mathcal{U}^{\varepsilon_n} - \frac{i-1}{M}, \mathcal{E}^{\lambda_n}, \mathcal{E}^{\lambda_{n+1}-\lambda_n}\right) \right),$$

where $\mathcal{U}^{\varepsilon_n}$, \mathcal{E}^{λ_n} and $\mathcal{E}^{\lambda_{n+1}}$ are independent random variables with $\mathcal{U}^{\varepsilon_n}$ being uniformly distributed on $[0, \varepsilon_n]$ and $\mathcal{E}^{\lambda_n}, \mathcal{E}^{\lambda_{n+1}-\lambda_n}$ being exponentially distributed

with parameters λ_n and $\lambda_{n+1} - \lambda_n$. Consequently, conditionally, one has that

$$\begin{aligned} \mathcal{Z}_{k_j}^{n,\gamma} &\stackrel{d}{=} \varepsilon_n^{-1/2} \left(\min(\mathcal{U}^{\varepsilon_n}, \mathcal{E}^{\lambda_n}) \right. \\ &\quad \left. - \min \left(\sum_{i=1}^M \mathbb{1}_{((i-1)\varepsilon_n, i\varepsilon_n]}(\mathcal{U}^{\varepsilon_n}) \left(\mathcal{U}^{\varepsilon_n} - \frac{i-1}{M} \right), \mathcal{E}^{\lambda_n}, \mathcal{E}^{\lambda_{n+1}-\lambda_n} \right) \right)^{1/2} \\ &\quad \times \xi y_j \mathbb{1}_{\{|y_j| \geq \varepsilon\}}, \end{aligned}$$

where ξ denotes an independent standard normal. By assumption, $\lambda_n/\varepsilon_n \rightarrow \theta$ as $n \rightarrow \infty$ so that the latter distribution converges to the one of $\frac{\sigma_j}{\sigma} \xi_j y_j$. Hence, conditionally on \mathcal{T}_n one has

$$\sum_{\substack{k \in \mathbb{N} \cap [0, t/\varepsilon_n] \\ k \text{th } n\text{-window occupied}}} \mathcal{Z}_k^{n,\gamma} \Rightarrow \sum_{j=1}^m \frac{\sigma_j}{\sigma} \xi_j y_j \mathbb{1}_{\{|y_j| \geq \varepsilon\}}.$$

Next, we analyse the contribution of all empty windows. Given \mathcal{T}_n , there are $t/\varepsilon_n - m$ empty windows and the corresponding random variables $\mathcal{Z}_k^{n,\gamma}$ are independent and identically distributed. We have

$$\mathbb{E}[\mathcal{Z}_1^{n,\gamma} | (0, \varepsilon_n] \text{ empty}, \mathcal{T}_n] = 0$$

since W is independent of the event we condition on. Further, by Itô's isometry and the scaling properties of Brownian motion one has

$$\begin{aligned} &\text{Var}(\mathcal{Z}_1^{n,\gamma} | (0, \varepsilon_n] \text{ empty}, \mathcal{T}_n) \\ &= \varepsilon_n^{-1} \sigma^2 \mathbb{E} \left[\int_0^{\varepsilon_n} (W_{t_{n+1}^\gamma(u)}^\gamma - W_{t_n^\gamma(u)}^\gamma)^2 du | (0, \varepsilon_n] \text{ empty}, \mathcal{T}_n \right] \\ (2.21) \quad &= \varepsilon_n \sigma^2 \mathbb{E} [(W_{\varepsilon_n^{-1} t_{n+1}^\gamma(\mathcal{U}^{\varepsilon_n})}^\gamma - W_{\varepsilon_n^{-1} t_n^\gamma(\mathcal{U}^{\varepsilon_n})}^\gamma)^2 | (0, \varepsilon_n] \text{ empty}, \mathcal{T}_n] \\ &= \varepsilon_n \sigma^2 \mathbb{E} [\varepsilon_n^{-1} t_{n+1}^\gamma(\mathcal{U}^{\varepsilon_n}) - \varepsilon_n^{-1} t_n^\gamma(\mathcal{U}^{\varepsilon_n}) | (0, \varepsilon_n] \text{ empty}, \mathcal{T}_n]. \end{aligned}$$

Here, we denote again by $\mathcal{U}^{\varepsilon_n}$ an independent uniform random variable on $[0, \varepsilon_n]$ and we used that conditionally the processes t_n^γ and t_{n+1}^γ are independent of the Brownian motion W . As above, we note that the distributions of $\varepsilon_n^{-1} t_{n+1}^\gamma(\mathcal{U}^{\varepsilon_n})$ and $\varepsilon_n^{-1} t_n^\gamma(\mathcal{U}^{\varepsilon_n})$ are identically distributed as

$$\varepsilon_n^{-1} \left(\mathcal{E}^{\lambda_{n+1}} \wedge \frac{\mathcal{U}^{\varepsilon_n}}{M} \right) \quad \text{and} \quad \varepsilon_n^{-1} (\mathcal{E}^{\lambda_n} \wedge \mathcal{U}^{\varepsilon_n}).$$

By assumption (ML2), these converge in L^1 to $\mathcal{E}^{M\theta} \wedge \mathcal{U}^{1/M}$ and $\mathcal{E}^\theta \wedge \mathcal{U}^1$, respectively. Hence, computing the respective expectations gives with (2.21)

$$\varepsilon_n^{-1} \text{Var}(\mathcal{Z}_1^{n,\gamma} | (0, \varepsilon_n] \text{ empty}, \mathcal{T}_n) \rightarrow \sigma^2 \frac{M-1}{M} \frac{e^{-\theta} - (1-\theta)}{\theta^2} =: \Upsilon^2.$$

The uniform L^2 -integrability of $\mathcal{L}(\varepsilon_n^{-1/2} \mathcal{Z}_1^{n,\gamma} | (0, \varepsilon_n] \text{ empty}, \mathcal{T}_n)$ follows by noticing that by the Burkholder–Davis–Gundy inequality there exists a universal constant κ such that

$$\begin{aligned} & \mathbb{E}((\mathcal{Z}_1^{n,\gamma})^4 | (0, \varepsilon_n] \text{ empty}, \mathcal{T}_n) \\ & \leq \kappa \varepsilon_n^{-2} \sigma^4 \mathbb{E} \left[\left(\int_0^{\varepsilon_n} (W_{i_{n+1}^\gamma(u)} - W_{i_n^\gamma(u)})^2 du \right)^2 \middle| (0, \varepsilon_n] \text{ empty}, \mathcal{T}_n \right] \\ & \leq 4\kappa \sigma^4 \mathbb{E} \left[\sup_{u \in [0, \varepsilon_n]} W_u^4 \right] = 4\kappa \sigma^4 \varepsilon_n^2 \mathbb{E} \left[\sup_{u \in [0, 1]} W_u^4 \right]. \end{aligned}$$

Hence, conditionally on \mathcal{T}_n one has

$$\sum_{\substack{k \in \mathbb{N} \cap (0, t/\varepsilon_n] \\ k\text{th } n\text{-window empty}}} \mathcal{Z}_k^{n,\gamma} \Rightarrow \mathcal{N}(0, \Upsilon^2 t).$$

Given \mathcal{T}_n the contribution of the empty and occupied windows are independent, so that since $\mathbb{P}(\mathcal{T}_n) \rightarrow 1$, generally

$$\sum_{k=1}^{t/\varepsilon_n} \mathcal{Z}_k^{n,\gamma} \Rightarrow Z_t^{\varepsilon,\gamma}.$$

It remains to show that

$$\lim_{n \rightarrow \infty} \left(\bar{Z}_t^{n,\varepsilon,\gamma} - \sum_{k=1}^{t/\varepsilon_n} \mathcal{Z}_k^{n,\gamma} \right) = 0 \quad \text{in probability.}$$

This follows immediately by noticing that, given \mathcal{T}_n , one has

$$\begin{aligned} & \bar{Z}_t^{n,\varepsilon,\gamma} - \sum_{k=1}^{t/\varepsilon_n} \mathcal{Z}_k^{n,\gamma} \\ & = \sigma \varepsilon_n^{-1/2} \sum_{\substack{k \in \mathbb{N} \cap [0, t/\varepsilon_n] \\ k\text{th } n\text{-window occupied}}} \int_{(k-1)\varepsilon_n}^{k\varepsilon_n} (W_{i_{n+1}(u-)} - W_{i_n(u-)}) dW_u, \end{aligned}$$

where the sum on the right-hand side is over m independent and identically distributed summands each having second moment smaller than ε_n^2 .

4th step. In the last step, we combine the results of the previous steps. By step one, it suffices to verify equation (2.19). Provided that the statement is true for $A = \Omega$, the system of sets A for which (2.19) is satisfied is a Dynkin system. Consequently, it suffices to verify validity for sets $A \cap A'$ with $A \in \sigma(W_t)$ and $A' \in \sigma(\Gamma)$. By step two, one has

$$\lim_{n \rightarrow \infty} |\mathbb{E}[\mathbb{1}_{A \cap A'} f(Z_t^{n,\varepsilon}) - \mathbb{P}(A) \mathbb{E}[\mathbb{1}_{A'} f(\bar{Z}_t^{n,\varepsilon})]]| \rightarrow 0$$

and by step three

$$\lim_{n \rightarrow \infty} \mathbb{E}[\mathbb{1}_{A'} f(\bar{Z}^{n,\varepsilon})] = \mathbb{E}[\mathbb{1}_{A'} f(Z_t^\varepsilon)]$$

so that

$$\lim_{n \rightarrow \infty} \mathbb{E}[\mathbb{1}_{A \cap A'} f(Z_t^{n,\varepsilon})] = \mathbb{P}(A) \mathbb{E}[\mathbb{1}_{A'} f(Z_t^\varepsilon)].$$

The proof is complete by noticing that $\sigma(W_t)$ is independent of $\sigma(\Gamma, Z_t^\varepsilon)$ so that

$$\mathbb{P}(A) \mathbb{E}[\mathbb{1}_{A'} f(\bar{Z}_t^\varepsilon)] = \mathbb{E}[\mathbb{1}_{A \cap A'} f(\bar{Z}_t^\varepsilon)]. \quad \square$$

3. Scaled errors of derived quantities. In this section, we collect results that will enable us to deduce the main central limit theorems with the help of Theorem 1.5.

3.1. *The integrated processes.* The following lemma is central to the proof of Theorems 1.9 and 1.10.

LEMMA 3.1. *If assumptions (ML1) and (ML2) hold, then one has*

$$\lim_{n \rightarrow \infty} \varepsilon_n^{-1} \mathbb{E} \left[\left| \int_0^T (\widehat{X}_t^n - \bar{X}_t^n) dt \right|^2 \right] = 0.$$

PROOF. With $b_n := b - \int_{B(0, h_n)^c} x \nu(dx)$ we have for $t \in [0, T]$

$$\begin{aligned} \widehat{X}_t^n - \bar{X}_t^n &= a(\widehat{X}_{\iota_n(t)}) (Y_t^h - Y_{\iota_n(t)}^h) \\ (3.1) \quad &= a(\widehat{X}_{\iota_n(t)}) (b_n(t - \iota_n(t)) + \sigma(W_t - W_{\iota_n(t)})). \end{aligned}$$

We estimate

$$\mathbb{E} \left[\left| \int_0^T a(\widehat{X}_{\iota_n(t)}) b_n(t - \iota_n(t)) dt \right|^2 \right] \leq b_n^2 \varepsilon_n^2 T \mathbb{E} \left[\int_0^T |a(\widehat{X}_{\iota_n(t-)})|^2 dt \right].$$

The latter expectation is uniformly bounded over all n ; see Lemma A.12. Further, $b_n^2 = o(\varepsilon_n^{-1})$ by Lemma A.10. Consequently, the first term is of order $o(\varepsilon_n)$. By Fubini,

$$\begin{aligned} &\mathbb{E} \left[\left(\int_0^T a(\widehat{X}_{\iota_n(t)}) \sigma(W_t - W_{\iota_n(t)}) dt \right)^2 \right] \\ &= \sigma^2 \int_0^T \int_0^T \mathbb{E} [a(\widehat{X}_{\iota_n(t)}) (W_t - W_{\iota_n(t)}) a(\widehat{X}_{\iota_n(u)}) (W_u - W_{\iota_n(u)})] dt du. \end{aligned}$$

Further, for $0 \leq t \leq u \leq T$,

$$\begin{aligned} &\mathbb{E} [a(\widehat{X}_{\iota_n(t)}) (W_t - W_{\iota_n(t)}) a(\widehat{X}_{\iota_n(u)}) (W_u - W_{\iota_n(u)}) | \iota_n, \widehat{X}_{\iota_n(t)}] \\ &= \mathbb{1}_{\{\iota_n(t) = \iota_n(u)\}} a(\widehat{X}_{\iota_n(t)})^2 ((t \wedge u) - \iota_n(t)) \end{aligned}$$

and since the statement is symmetric in the variables t, u also for $0 \leq u \leq t \leq T$. Consequently,

$$\mathbb{E} \left[\left(\int_0^T a(\widehat{X}_{t_n(t)}) \sigma(W_t - W_{t_n(t)}) dt \right)^2 \right] \leq 2\varepsilon_n^2 \sigma^2 \int_0^T \mathbb{E}[a(\widehat{X}_{t_n(t)})^2] dt.$$

We recall that the latter expectation is uniformly bounded so that this term is also of order $o(\varepsilon_n)$. \square

3.2. The supremum. The results of this subsection are central to the proof of Theorem 1.8. We first give some qualitative results for solutions $X = (X_t)_{t \in [0, T]}$ of the stochastic differential equation

$$dX_t = a(X_{t-}) dY_t$$

with arbitrary starting value. We additionally assume that a does *not* attain zero.

LEMMA 3.2. *One has for every $t \in [0, T]$ that, almost surely,*

$$\sup_{s \in [0, t]} X_s > X_0 \vee X_t.$$

PROOF. We only prove that

$$\sup_{s \in [0, t]} X_s > X_t$$

and remark that the remaining statement follows by similar simpler considerations.

1st step. In the first step, we show that

$$\frac{1}{\sqrt{\varepsilon}} (X_{t-\varepsilon+\varepsilon s} - X_{t-\varepsilon})_{s \in [0, 1]} \xrightarrow{\text{stably}} (\sigma a(X_t) B_s)_{s \in [0, 1]}.$$

We show the statement in two steps: first note that

$$\frac{1}{\sqrt{\varepsilon}} (X_{t-\varepsilon+\varepsilon s} - X_{t-\varepsilon})_{s \in [0, 1]} \quad \text{and} \quad \frac{1}{\sqrt{\varepsilon}} (a(X_{t-\varepsilon})(Y_{t-\varepsilon+\varepsilon s} - Y_{t-\varepsilon}))_{s \in [0, 1]}$$

are equivalent in ucp. Further, $Z^\varepsilon := (\varepsilon^{-1/2}(Y_{T-\varepsilon+\varepsilon s} - Y_{t-\varepsilon}))_{s \in [0, 1]}$ is independent of $a(X_{t-\varepsilon})$ and $a(X_{t-\varepsilon})$ tends to $a(X_t)$, almost surely. Hence, it remains to show that Z^ε converges for $\varepsilon \downarrow 0$ in distribution to σB . Note that Z^ε is a Lévy-process with triplet $(b\sqrt{\varepsilon}, \sigma^2, \nu_\varepsilon)$, where $\nu_\varepsilon(A) = \varepsilon \nu(\sqrt{\varepsilon}A)$ for Borel sets $A \subset \mathbb{R} \setminus \{0\}$. It suffices to show that Lévy-processes \overline{Z}^ε with triplet $(0, 0, \nu_\varepsilon)$ converge to the zero process.

We uniquely represent \overline{Z}^ε as

$$\overline{Z}_t^\varepsilon = \overline{Z}_t^{\varepsilon, r} + \overline{\overline{Z}}^{\varepsilon, r} - b_{\varepsilon, r} t$$

with independent Lévy processes $\bar{Z}_t^{\varepsilon,r}$ and $\bar{\bar{Z}}^{\varepsilon,r}$, the first one with triplet $(0, 0, \nu_\varepsilon|_{B(0,r)})$, the second one being a compound Poisson process with intensity $\nu|_{B(0,r)^c}$, and with $b_{\varepsilon,r} := \int_{B(0,r)^c} x \, d\nu_\varepsilon(x)$. Clearly, for $\delta > 0$

$$(3.2) \quad \mathbb{P}\left(\sup_{t \in [0,1]} |\bar{Z}_t^\varepsilon| > \delta\right) \leq \mathbb{1}_{\{|b_{\varepsilon,r}| > \delta/2\}} + \mathbb{P}\left(\sup_{t \in [0,1]} |\bar{Z}_t^{\varepsilon,r}| > \delta/2\right) + \mathbb{P}(\bar{\bar{Z}}^{\varepsilon,r} \neq 0).$$

For $r > 0$, one has

$$\begin{aligned} r\nu(B(0,r)^c) &\leq \int_{B(0,r)^c} |x| \, d\nu_\varepsilon(x) = \varepsilon \int_{B(0,\sqrt{\varepsilon r})^c} \frac{|x|}{\sqrt{\varepsilon}} \, d\nu(x) \\ &\leq \sqrt{\varepsilon} \int_{B(0,\sqrt{\varepsilon r})} \frac{x^2}{\sqrt{\varepsilon r}} \nu(dx) \leq \frac{1}{r} \int x^2 \nu(dx). \end{aligned}$$

Hence, $|b_{\varepsilon,r}| \leq \delta/2$, for sufficiently large r , and $\mathbb{P}(\bar{\bar{Z}}^{\varepsilon,r} \neq 0) \leq \nu(B(0,r)^c) \leq \frac{1}{r^2} \int x^2 \nu(dx)$. Further,

$$\int_{B(0,r)} x^2 \, d\nu_\varepsilon(x) = \int_{B(0,\sqrt{\varepsilon r})} x^2 \, d\nu(x) \rightarrow 0$$

so that Doob's L^2 -inequality yields

$$\lim_{\varepsilon \downarrow 0} \mathbb{P}\left(\sup_{t \in [0,1]} |\bar{Z}_t^{\varepsilon,r}| > \delta/2\right) = 0.$$

Plugging these estimates into (3.2) gives

$$\limsup_{\varepsilon \downarrow 0} \mathbb{P}\left(\sup_{t \in [0,1]} |\bar{Z}_t^\varepsilon| > \delta\right) \leq \frac{1}{r^2} \int x^2 \nu(dx)$$

and the statement of step one follows by noticing that $r > 0$ can be chosen arbitrarily large.

2nd step. Clearly, for $\varepsilon \in (0, t]$,

$$\mathbb{P}\left(\sup_{s \in [0,t]} X_s = X_t\right) \leq \mathbb{P}\left(\varepsilon^{-1/2} \sup_{s \in [0,1]} (X_{t-\varepsilon+\varepsilon s} - X_{t-\varepsilon}) = \varepsilon^{-1/2} (X_t - X_{t-\varepsilon})\right).$$

The set of all càdlàg functions $x : [0, 1] \rightarrow \mathbb{R}$ with $\sup_{s \in [0,1]} x_s = x_1$ is closed in the Skorokhod space so that

$$\begin{aligned} &\mathbb{P}\left(\sup_{s \in [0,t]} X_s = X_t\right) \\ &\leq \limsup_{\varepsilon \downarrow 0} \mathbb{P}\left(\varepsilon^{-1/2} \sup_{s \in [0,1]} (X_{t-\varepsilon+\varepsilon s} - X_{t-\varepsilon}) = \varepsilon^{-1/2} (X_t - X_{t-\varepsilon})\right) \\ &\leq \mathbb{P}\left(a(X_t) \sup_{s \in [0,1]} \sigma B_s = a(X_t) B_1\right) = 0. \end{aligned}$$

□

LEMMA 3.3. *Suppose that $a(x) \neq 0$ for all $x \in \mathbb{R}$. There is a unique random time S (up to indistinguishability) such that, almost surely,*

$$\sup_{s \in [0, T]} X_s = X_S$$

and one has $\Delta X_S = 0$. Further, for every $\varepsilon > 0$, almost surely,

$$\sup_{s \in [0, S]: |s-S| \geq \varepsilon} X_s < X_S.$$

PROOF. *1st step.* First we prove that the supremum $\sup_{t \in [0, T]} X_t$ is almost surely attained at some random time S with $\Delta X_S = 0$. By compactness of the time domain, we can find an almost surely convergent $[0, T]$ -valued sequence $(S_n)_{n \in \mathbb{N}}$ of random variables, say with limit S , with

$$\lim_{n \rightarrow \infty} X_{S_n} = \sup_{t \in [0, T]} X_t.$$

Let $h > 0$. We represent Y as sum

$$Y_t = Y_t^h + \sum_{k=1}^N \mathbb{1}_{[T_k, T]}(t) \Delta Y_{T_k},$$

where T_1, \dots, T_N are the increasingly ordered times of the discontinuities of Y being larger than h . Further, Y^h is a Lévy process that is independent of $\bar{Y}^h := Y - Y^h$. Given \bar{Y}^h , for every $k = 1, \dots, N$, the process $(X_t)_{t \in [T_{k-1}, T_k]}$ solves the SDE

$$dX_t = a(X_{t-}) dY_t^h$$

and we have, almost surely, that

$$X_{T_k-} = X_{T_{k-1}} + \int_{T_{k-1}}^{T_k} a(X_s) dY_s^h.$$

Consequently, we can apply Lemma 3.2 and conclude that, almost surely, for each $k = 1, \dots, N+1$,

$$\sup_{s \in [T_{k-1}, T_k)} X_s > X_{T_{k-1}} \vee X_{T_k-}$$

with $T_0 = 0$ and $T_{N+1} = T$. Hence, almost surely,

$$\sup_{s \in [0, T]} X_s > \sup_{k=1, \dots, N+1} X_{T_{k-1}} \vee X_{T_k-}.$$

Consequently, S is almost surely not equal to 0 or T or a time with displacement larger than h . Since $h > 0$ was arbitrary, we get that, almost surely, $\Delta X_S = 0$, so that

$$X_S = \lim_{n \rightarrow \infty} X_{S_n} = \sup_{t \in [0, T]} X_t \quad \text{almost surely.}$$

2nd step. We prove that for every $t \in [0, T]$ the distribution of $\sup_{s \in [0, t]} X_s$ has no atom. Suppose that it has an atom in $z \in \mathbb{R}$. We consider the stopping time

$$T_{\{z\}} = \inf\{t \in [0, T] : X_t = z\}$$

with the convention $T_{\{z\}} = \infty$ in the case when z is not hit. For $\varepsilon > 0$, conditionally on the event $\{T_{\{z\}} \leq T - \varepsilon\}$ the process $(\tilde{X}_s)_{s \in [0, \varepsilon]}$ with

$$\tilde{X}_s = X_{T_{\{z\}}+s}$$

starts in z and solves $d\tilde{X}_s = a(\tilde{X}_s) d\tilde{Y}_s$ with \tilde{Y} denoting the $T_{\{z\}}$ -shifted Lévy process Y . Hence, by Lemma 3.2, one has almost surely on $\{T_{\{z\}} \leq T - \varepsilon\}$ that

$$z = \tilde{X}_0 < \sup_{s \in [0, \varepsilon]} \tilde{X}_s \leq \sup_{s \in [0, T]} X_s.$$

Since $\varepsilon > 0$ is arbitrary and X does not attain its supremum in T , it follows that $\mathbb{P}(\sup_{s \in [0, T]} X_s = z) = 0$.

3rd step. We prove that the supremum over two disjoint time windows $[u, v]$ and $[w, z]$ with $0 \leq u < v \leq w < z \leq T$, satisfies

$$\sup_{s \in [u, v]} X_s \neq \sup_{s \in [w, z]} X_s,$$

almost surely. By the Markov property, the random variables $\sup_{s \in [u, v]} X_s$ and $\sup_{s \in [w, z]} X_s$ are independent given X_w and we get

$$\begin{aligned} & \mathbb{P}\left(\sup_{s \in [u, v]} X_s = \sup_{s \in [w, z]} X_s\right) \\ &= \int \mathbb{P}\left(\sup_{s \in [u, v]} X_s = y \mid X_w = x\right) d\mathbb{P}_{(X_w, \sup_{s \in [u, v]} X_s)}(x, y), \end{aligned}$$

where $\mathbb{P}_{(X_w, \sup_{s \in [u, v]} X_s)}$ denotes the distribution of $(X_w, \sup_{s \in [u, v]} X_s)$. We note that the conditional process $(X_s)_{s \in [w, z]}$ is again a solution of the SDE started in x and by step two the inner conditional probability equals zero.

4th step. We finish the proof of the statement. For given $\varepsilon > 0$, we choose deterministic times $0 = t_0 < t_1 < \dots < t_m = T$ with $t_k - t_{k-1} \leq \varepsilon$. By step three, there is, almost surely, one window in which the supremum is attained, say in $[t_{M-1}, t_M)$, and

$$\sup_{s \in [0, T] : |S-s| \geq \varepsilon} X_s \leq \sup_{k \in \{1, \dots, m\} \setminus \{M\}} \sup_{s \in [t_{k-1}, t_k)} X_s < \sup_{s \in [t_{M-1}, t_M)} X_s = X_S. \quad \square$$

LEMMA 3.4. *Suppose that $a(x) \neq 0$ for all $x \in \mathbb{R}$ and denote by S the random time at which X attains its maximum. One has*

$$\varepsilon_n^{-1/2} \left(\sup_{t \in [0, T]} X_t^{n+1} - \sup_{t \in [0, T]} X_t^n \right) - U_S^{n, n+1} \rightarrow 0 \quad \text{in probability.}$$

PROOF. With Lemma 3.3 we conclude that, for every $\varepsilon > 0$, one has with high probability that

$$\begin{aligned} & \left| \varepsilon_n^{-1/2} \left(\sup_{t \in [0, T]} X_t^{n+1} - \sup_{t \in [0, T]} X_t^n \right) - U_S^{n, n+1} \right| \\ & \leq \sup_{t: |t-S| \leq \varepsilon} \left| \varepsilon_n^{-1/2} (X_t^{n+1} - X_t^n) - U_S^{n, n+1} \right| \\ & = \sup_{t: |t-S| \leq \varepsilon} |U_t^{n, n+1} - U_S^{n, n+1}|. \end{aligned}$$

For $\varepsilon, \delta > 0$, consider

$$A_{\varepsilon, \delta} = \left\{ (s, x) \in [0, T] \times \mathbb{D}(\mathbb{R}) : \sup_{(t, u) : s - \varepsilon \leq t \leq u \leq s + \varepsilon} |x_t - x_u| \geq \delta \right\}.$$

Note that $\text{cl}(A_{\varepsilon, \delta}) \subset A_{2\varepsilon, \delta}$ and recall that $(S, U^{n, n+1}) \Rightarrow (S, U)$. Hence,

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \mathbb{P} \left(\left| \varepsilon_n^{-1/2} \sup_{t \in [0, T]} X_t^{n+1} - \varepsilon_n^{-1/2} \sup_{t \in [0, T]} X_t^n - U_S^{n, n+1} \right| \geq \delta \right) \\ & \leq \limsup_{n \rightarrow \infty} \mathbb{P}((S, U^{n, n+1}) \in A_{\varepsilon, \delta}) \leq \mathbb{P}((S, U) \in A_{2\varepsilon, \delta}). \end{aligned}$$

Note that U is almost surely continuous in S so that for $\varepsilon \downarrow 0$, $\mathbb{P}((S, U) \in A_{2\varepsilon, \delta}) \rightarrow 0$. \square

4. Proofs of the central limit theorems. In this section, we prove all central limit theorems and Theorem 1.11. We will verify the Lindeberg conditions for the summands of the multilevel estimate $\widehat{S}(F)$; see (1.4). As shown in Lemma A.9 in the Appendix, a central limit theorem holds for the idealised approximations X^1, X^2, \dots , if:

- (1) $\lim_{n \rightarrow \infty} \text{Var}(\varepsilon_n^{-1/2} (F(X^{n+1}) - F(X^n))) = \rho^2$ and
- (2) $(\varepsilon_n^{-1/2} (F(X^{n+1}) - F(X^n))) : k \in \mathbb{N}$ is uniformly L^2 -integrable.

The section is organised as follows. In Section 4.1, we verify uniform L^2 -integrability of the error process in supremum norm which will allow us to verify property (2) in the central limit theorems. In Section 4.2, we prove Theorems 1.6 and 1.9, essentially by verifying property (1).

It remains to deduce Theorems 1.8 and 1.10 from the respective theorems for the idealised scheme. By Lemmas 1.2, 1.4 and 3.1, switching from the idealised to the continuous or piecewise constant approximation leads to asymptotically equivalent L^2 -errors. Hence, the same error process can be used and, in particular, uniform L^2 -integrability prevails due to Lemma A.8. Consequently, the identical proofs yield the statements.

Finally, we prove Theorem 1.11 in Section 4.3.

4.1. Uniform L^2 -integrability.

PROPOSITION 4.1. *The sequence $(\varepsilon_n^{-1/2} \sup_{t \in [0, T]} |X_t^{n+1} - X_t^n|)_{n \in \mathbb{N}}$ is uniformly L^2 -integrable.*

To prove the proposition, we will make use of the perturbation estimates given in the [Appendix](#); see Section A.4. Recall that $U^{n, n+1} = \varepsilon_n^{-1/2} (X^{n+1} - X^n)$ satisfies the equation

$$\begin{aligned} U_t^{n, n+1} &= \int_0^t D_{s-}^{n, n+1} U_{s-}^{n, n+1} dY_s + \varepsilon_n^{-1/2} \int_0^t D_{s-}^n A_{s-}^n (Y_{s-} - Y_{t_n(s-)}) dY_s \\ &\quad - \varepsilon_n^{-1/2} \int_0^t D_{s-}^{n+1} A_{s-}^{n+1} (Y_{s-} - Y_{t_{n+1}(s-)}) dY_s. \end{aligned}$$

We use approximations indexed by $m \in \mathbb{N}$: we denote by

$$\mathcal{U}^{n, n+1, m} = (\mathcal{U}_t^{n, n+1, m})_{t \in [0, T]}$$

the solution of the equation

$$\begin{aligned} (4.1) \quad \mathcal{U}_t^{n, n+1, m} &= \int_0^t D_{s-}^{n, n+1} \mathcal{U}_s^{n, n+1, m} d\mathcal{Y}_s^m \\ &\quad + \varepsilon_n^{-1/2} \sigma \int_0^t D_{s-}^n \mathcal{A}_{s-}^{n, m} (W_{s-} - W_{t_n(s-)}) d\mathcal{Y}_s^m \\ &\quad - \varepsilon_n^{-1/2} \sigma \int_0^t D_{s-}^{n+1} \mathcal{A}_{s-}^{n, m} (W_{s-} - W_{t_{n+1}(s-)}) d\mathcal{Y}_s^m, \end{aligned}$$

where $\mathcal{Y}^m = (\mathcal{Y}_t^m)_{t \in [0, T]}$ is given by

$$\mathcal{Y}_t^m = bt + \sigma W_t + \lim_{\delta \downarrow 0} \int_{(0, t] \times (B(0, m) \setminus B(0, \delta))} x d\bar{\Pi}(s, x),$$

and $\mathcal{A}^{n, m} = (\mathcal{A}_t^{n, m})_{t \in [0, T]}$ is the simple adapted càdlàg process given by

$$\mathcal{A}_t^{n, m} = \begin{cases} A_t^n, & \text{if } |A_t^n| \leq m, \\ 0, & \text{else.} \end{cases}$$

The proof of the proposition is achieved in two steps. We show that:

1. $\lim_{m \uparrow \infty} \limsup_{n \rightarrow \infty} \mathbb{E}[\sup_{t \in [0, T]} |U_t^{n, n+1} - \mathcal{U}_t^{n, n+1, m}|^2] = 0$ and
2. for every $p \geq 2$ and $m \in \mathbb{N}$, $\mathbb{E}[\sup_{t \in [0, T]} |\mathcal{U}_t^{n, n+1, m}|^p] < \infty$.

Then the uniform L^2 -integrability of $(\sup_{t \in [0, T]} |U_t^{n, n+1}|)_{n \in \mathbb{N}}$ follows with [Lemma A.8](#).

LEMMA 4.2. *One has*

$$\lim_{m \uparrow \infty} \limsup_{n \rightarrow \infty} \mathbb{E} \left[\sup_{t \in [0, T]} |U_t^{n, n+1} - \mathcal{U}_t^{n, n+1, m}|^2 \right] = 0.$$

PROOF. The processes $\mathcal{U}^{n,n+1,m}$ are perturbations of $U^{n,n+1}$ as analysed in Lemma A.14. More explicitly, the result follows if there exists a constant $\kappa > 0$ such that

$$(4.2) \quad \mathbb{E} \left[\sup_{t \in [0, T]} \left| \varepsilon_n^{-1/2} \int_0^t D_{s-}^n \mathcal{A}_{s-}^{n,m} (W_{s-} - W_{l_n(s-)}) d\mathcal{Y}_s^m \right|^2 \right] \leq \kappa,$$

for all $n, m \in \mathbb{N}$, and

$$(4.3) \quad \lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \varepsilon_n^{-1} \mathbb{E} \left[\sup_{t \in [0, T]} \left| \int_0^t D_{s-}^n A_{s-}^n (Y_{s-} - Y_{l_n(s-)}) dY_s - \sigma \int_0^t D_{s-}^n \mathcal{A}_{s-}^{n,m} (W_{s-} - W_{l_n(s-)}) d\mathcal{Y}_s^m \right|^2 \right] = 0.$$

Using Lemma A.11, the uniform boundedness of D^n , conditional independence of $\mathcal{A}_{s-}^{n,m}$ and $W_{s-} - W_{l_n(s-)}$ given l_n , there exists a constant $\kappa_1 > 0$ such that

$$\begin{aligned} & \varepsilon_n^{-1} \mathbb{E} \left[\sup_{0 \leq r \leq t} \left| \int_0^r D_{s-}^n \mathcal{A}_{s-}^{n,m} (W_{s-} - W_{l_n(s-)}) d\mathcal{Y}_s^m \right|^2 \right] \\ & \leq \kappa_1 \int_0^T \varepsilon_n^{-1} \mathbb{E} [|\mathcal{A}_{s-}^{n,m}|^2 | W_{s-} - W_{l_n(s-)}|^2] ds \\ & \leq \kappa_1 \int_0^T \mathbb{E} [|\mathcal{A}_{s-}^{n,m}|^2] ds \leq \kappa_1 \int_0^T \mathbb{E} [|A_{s-}^n|^2] ds \end{aligned}$$

for all $n, m \in \mathbb{N}$. The latter integral is uniformly bounded by Lemma A.12 and the Lipschitz continuity of a .

We proceed with the analysis of (4.3). The expectation in (4.3) is bounded by twice the sum of

$$\begin{aligned} \Sigma_{n,m}^{(1)} & := \varepsilon_n^{-1} \mathbb{E} \left[\sup_{t \in [0, T]} \left| \int_0^t D_{s-}^n A_{s-}^n (Y_{s-} - Y_{l_n(t-)}) dY_s - \sigma \int_0^t D_{s-}^n A_{s-}^n (W_{s-} - W_{l_n(s-)}) d\mathcal{Y}_s^m \right|^2 \right] \end{aligned}$$

and

$$\begin{aligned} \Sigma_{n,m}^{(2)} & := \varepsilon_n^{-1} \mathbb{E} \left[\sup_{t \in [0, T]} \left| \int_0^t D_{s-}^n A_{s-}^n (W_{s-} - W_{l_n(t-)}) d\mathcal{Y}_s^m - \int_0^t D_{s-}^n \mathcal{A}_{s-}^{n,m} (W_{s-} - W_{l_n(t-)}) d\mathcal{Y}_s^m \right|^2 \right]. \end{aligned}$$

The term $\Sigma_{n,m}^{(1)}$ is the same as the one appearing in (2.6) when replacing Y^ε by \mathcal{Y}^m . One can literally translate the proof of (2.6) to obtain that

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \Sigma_{n,m}^{(1)} = 0.$$

By uniform boundedness of D^n and Lemma A.11, there exists a constant κ_2 not depending on $n, m \in \mathbb{N}$ with

$$\begin{aligned} \Sigma_{n,m}^{(2)} &\leq \kappa_2 \varepsilon_n^{-1} \int_0^T \mathbb{E}[(A_{s-}^n - \mathcal{A}_{s-}^{n,m})^2 (W_{s-} - W_{\iota_n(s-)})^2] ds \\ &\leq \kappa_2 \int_0^T \mathbb{E}[(A_{s-}^n - \mathcal{A}_{s-}^{n,m})^2] ds \\ &\leq 2\kappa_2 \int_0^T \mathbb{E}[(A_{s-}^n - a(X_{s-}))^2] ds + 2\kappa_2 \int_0^T \mathbb{E}[(a(X_{s-}) - \mathcal{A}_{s-}^{n,m})^2] ds, \end{aligned}$$

where we have used again that given ι_n the random variables $A_{s-}^n - \mathcal{A}_{s-}^{n,m}$ and $W_{s-} - W_{\iota_n(s-)}$ are independent. The first integral in the previous line tends to zero by Lipschitz continuity of a and L^2 -convergence of $\sup_{t \in [0, T]} |X_t^n - X_t| \rightarrow 0$ (see Proposition 4.1 of [11]). Further, the second integral satisfies

$$\limsup_{n \rightarrow \infty} \int_0^T \mathbb{E}[(a(X_{s-}) - \mathcal{A}_{s-}^{n,m})^2] ds \leq \int_0^T \mathbb{E}[\mathbb{1}_{[m, \infty)}(|X_{s-}|) a(X_{s-})^2] ds$$

which tends to zero as $m \rightarrow \infty$ since $\sup_{t \in [0, T]} |X_t|$ is square integrable. \square

LEMMA 4.3. *For every $m \in \mathbb{N}$ and $p \geq 2$, one has*

$$\sup_{n \in \mathbb{N}} \mathbb{E} \left[\sup_{t \in [0, T]} |\mathcal{U}_t^{n, n+1, m}|^p \right] < \infty.$$

PROOF. Since \mathcal{Y}^h has bounded jumps, it has finite p th moment. $D^{n, n+1}$ is uniformly bounded and by part one of Lemma A.15 it suffices to prove that

$$\mathbb{E} \left[\sup_{t \in [0, T]} \left| \varepsilon_n^{-1/2} \int_0^t D_{s-}^n \mathcal{A}_{s-}^{n,m} (W_{s-} - W_{\iota_n(s-)}) d\mathcal{Y}_s^m \right|^p \right]$$

is uniformly bounded over all $n \in \mathbb{N}$ for fixed $m \in \mathbb{N}$. Using Lemma A.11 and the uniform boundedness of D^n and $\mathcal{A}^{n,m}$ over all $n \in \mathbb{N}$, we conclude existence of a constant κ_3 such that for every $n \in \mathbb{N}$

$$\begin{aligned} &\mathbb{E} \left[\sup_{t \in [0, T]} \left| \int_0^t D_{s-}^n \mathcal{A}_{s-}^{n,m} (W_{s-} - W_{\iota_n(s-)}) d\mathcal{Y}_s^m \right|^p \right] \\ &\leq \kappa_3 \int_0^T \mathbb{E}[|W_{s-} - W_{\iota_n(s-)}|^p] ds \leq \kappa_3 T \varepsilon_n^{p/2}. \quad \square \end{aligned}$$

4.2. *Proof of the central limit theorems for X^1, X^2, \dots* In this section we prove Theorems 1.6 and 1.9. By Proposition 4.1 and the Lipschitz continuity of F with respect to supremum norm, we conclude that $(\varepsilon_n^{-1/2}(F(X^{n+1}) - F(X^n)) : n \in \mathbb{N})$ is uniformly L^2 -integrable in both settings. In view of the discussion at the beginning of Section 4 it suffices to show that

$$\lim_{n \rightarrow \infty} \text{Var}(\varepsilon_n^{-1/2}(F(X^{n+1}) - F(X^n))) = \text{Var}(\nabla f(AX) \cdot AU)$$

in the first setting and

$$\lim_{n \rightarrow \infty} \text{Var}(\varepsilon_n^{-1/2}(F(X^{n+1}) - F(X^n))) = \text{Var}(f'(X_S) \cdot U_S)$$

in the second setting. By dominated convergence it even suffices to show weak convergence of the distributions appearing in the variances. Theorem 1.6 follows from the following lemma.

LEMMA 4.4. *Under the assumptions of Theorem 1.6, one has*

$$\varepsilon_n^{-1/2}(F(X^{n+1}) - F(X^n)) \Rightarrow \nabla f(AX) \cdot AU.$$

PROOF. For $n \in \mathbb{N}$, let $Z_n := AX^n$ and set $Z = AX$. Since $Z \in D_f$, almost surely, we conclude that

$$(4.4) \quad \lim_{n \rightarrow \infty} \varepsilon_n^{-1/2}(f(Z_n) - f(Z) - \nabla f(Z)(Z_n - Z)) = 0 \quad \text{in probability.}$$

Indeed, one has $f(Z_n) - f(Z) - \nabla f(Z)(Z_n - Z) = R_n(Z_n - Z)$ for appropriate random variables R_n that converge in probability to zero since $Z_n - Z \rightarrow 0$, in probability, and f is differentiable in Z . Further, for fixed $\varepsilon > 0$ we choose $\delta > 0$ large and estimate

$$\mathbb{P}(|\varepsilon_n^{-1/2}R_n(Z_n - Z)| > \varepsilon) \leq \mathbb{P}(|R_n| > \varepsilon/\delta) + \mathbb{P}(|\varepsilon_n^{-1/2}(Z_n - Z)| > \delta).$$

The first summand converges to zero as $n \rightarrow \infty$ and the second term can be made uniformly arbitrarily small over n by choosing δ sufficiently large due to tightness of the sequence $(\varepsilon_n^{-1/2}(Z_n - Z))_{n \in \mathbb{N}}$. Equation (4.4) remains true when replacing Z_n by Z_{n+1} and we conclude that

$$\lim_{n \rightarrow \infty} \varepsilon_n^{-1/2}(f(Z_{n+1}) - f(Z_n) - \nabla f(Z)(Z_{n+1} - Z_n)) = 0 \quad \text{in probability.}$$

By Theorem 1.5 and the fact that A is continuous in \mathbb{P}_U -almost every point, we conclude that

$$(Y, A\varepsilon_n^{-1/2}(X^{n+1} - X^n)) \Rightarrow (Y, AU)$$

and, hence,

$$\varepsilon_n^{-1/2}(Z_{n+1} - Z_n) \xrightarrow{\text{stably}} AU,$$

by Lemma A.2. Consequently, since $\nabla f(Z)$ is $\sigma(Y)$ -measurable we get

$$(\nabla f(Z), \varepsilon_n^{-1/2}(Z_{n+1} - Z_n)) \Rightarrow (\nabla f(Z), AU)$$

and the proof is completed by noticing that the scalar product is continuous. \square

Analogously, Theorem 1.9 is a consequence of the following lemma.

LEMMA 4.5. *Under the assumptions of Theorem 1.9, one has*

$$\varepsilon_n^{-1/2} \left(f \left(\sup_{s \in [0, T]} X_s^{n+1} \right) - f \left(\sup_{s \in [0, T]} X_s^n \right) \right) \Rightarrow f'(X_S) \cdot U_S,$$

where S denotes the time where X attains its maximum.

PROOF. By Lemma 3.3, there exists a unique time S at which X attains its maximum and by Lemma 3.4 one has

$$\varepsilon_n^{-1/2} \left(\sup_{s \in [0, T]} X_s^{n+1} - \sup_{s \in [0, T]} X_s^n \right) - U_S^{n, n+1} \rightarrow 0 \quad \text{in probability.}$$

By Theorem 1.5 and Lemma A.2, one has

$$(Y, S, U^{n, n+1}) \Rightarrow (Y, S, U)$$

and the function $[0, T] \times \mathbb{D}(\mathbb{R}) \rightarrow \mathbb{R}, (s, u) \mapsto u_s$ is continuous in $\mathbb{P}_{S, U}$ -almost all (s, u) since U is almost surely continuous in S by Lemma 3.3. Consequently,

$$(Y, U_S^{n, n+1}) \Rightarrow (Y, U_S)$$

and, hence,

$$\varepsilon_n^{-1/2} \left(\sup_{s \in [0, T]} X_s^{n+1} - \sup_{s \in [0, T]} X_s^n \right) \xrightarrow{\text{stably}} U_S.$$

The rest follows as in the proof of Lemma 4.4. \square

4.3. *Proof of Theorem 1.11. 1st step.* Denote by $\mathcal{E} = (\mathcal{E}_t)_{t \in [0, T]}$ the stochastic exponential of $(\int_0^t a'(X_{s-}) dY_s)_{t \in [0, T]}$. In particular, \mathcal{E} does not hit zero with probability one; see, for instance, [18], Theorem 1.4.61. In the first step, we show that $\mathbb{E}[U_s U_t | Y] = \Upsilon^2 \phi_{s, t}(Y)$, where

$$\begin{aligned} \phi_{s, t}(Y) &= \sigma^4 \mathcal{E}_s \mathcal{E}_t \int_0^s \frac{(aa')(X_{u-})^2}{\mathcal{E}_{u-}^2} du \\ (4.5) \quad &+ \sigma^2 \lim_{\delta \downarrow 0} \mathcal{E}_s \mathcal{E}_t \sum_{\substack{u \in (0, s]: \\ |\Delta Y_u| \geq \delta}} \frac{(aa')(X_{u-})^2 \Delta Y_u^2}{(1 + a'(X_{u-}) \Delta Y_u)^2 \mathcal{E}_{u-}^2} \end{aligned}$$

and the limit is taken in ucp.

We define $\bar{L} = (\bar{L}_t)_{t \in [0, T]}$ by

$$\bar{L}_t = \sigma^2 \Upsilon B_t + \lim_{\delta \downarrow 0} \sum_{\substack{s \in (0, t]: \\ |\Delta Y_s| \geq \delta}} \frac{1}{1 + a'(X_{s-}) \Delta Y_s} \Delta L_s$$

and note that the process is well-defined since the denominator does not attain the value zero by assumption. Using the product rule and independence of W and B , it is straight forward to verify that

$$\left(\mathcal{E}_t \int_0^t \frac{(aa')(X_{s-})}{\mathcal{E}_{s-}} d\bar{L}_s \right)_{t \in [0, T]}$$

solves the stochastic integral equation (1.11) and by strong uniqueness of the solution equals U , almost surely. We write

$$U_t = \underbrace{\sigma^2 \Upsilon \mathcal{E}_t \int_0^t \frac{(aa')(X_{s-})}{\mathcal{E}_{s-}} dB_s}_{=: Z_t} + \lim_{\delta \downarrow 0} \underbrace{\mathcal{E}_t \sum_{\substack{s \in (0, t]: \\ |\Delta Y_s| \geq \delta}} \frac{(aa')(X_{s-})}{(1 + a'(X_{s-}) \Delta Y_s) \mathcal{E}_{s-}} \Delta L_s}_{Z_t^{(\delta)}}$$

and note that given Y the processes Z and $Z^{(\delta)}$ are independent and have expectation zero. Further, for $0 \leq s \leq t \leq T$ one has

$$\mathbb{E}[Z_s Z_t | Y] = \mathcal{E}_s \mathcal{E}_t \int_0^s \frac{(aa')(X_{u-})^2}{\mathcal{E}_{u-}^2} du$$

and

$$\mathbb{E}[Z_s^{(\delta)} Z_t^{(\delta)} | Y] = \mathcal{E}_s \mathcal{E}_t \sum_{\substack{u \in (0, s]: \\ |\Delta Y_u| \geq \delta}} \frac{(aa')(X_{u-})^2 \Delta Y_u^2}{(1 + a'(X_{u-}) \Delta Y_u)^2 \mathcal{E}_{u-}^2} \mathbb{E}[\sigma_u^2].$$

One easily computes that $\mathbb{E}[\sigma_u^2] = \sigma^2 \Upsilon^2$. Altogether, it follows the wanted statement.

2nd step. Let $A = (A_1, \dots, A_d) : \mathbb{D}(\mathbb{R}) \rightarrow \mathbb{R}^d$ be a linear map of integral type meaning that there are finite signed measures μ_1, \dots, μ_d on $[0, T]$ with

$$A_j x = \int_0^T x_s d\mu_j(s).$$

Then by conditional Fubini and step one,

$$\begin{aligned} & \text{Var}[\nabla f(AX) \cdot AU] \\ &= \sum_{i, j=1}^d \mathbb{E}[\partial_i f(AX) A_i U \partial_j f(AX) A_j U] \\ &= \sum_{i, j=1}^d \mathbb{E} \left[\partial_i f(AX) \partial_j f(AX) \mathbb{E} \left[\int_{[0, T]^2} U_u U_v d\mu_i \otimes \mu_j(u, v) \mid Y \right] \right] \\ &= \Upsilon^2 \sum_{i, j=1}^d \mathbb{E} \left[\partial_i f(AX) \partial_j f(AX) \int_{[0, T]^2} \phi_{u, v}(Y) d\mu_i \otimes \mu_j(u, v) \right]. \end{aligned}$$

3rd step. The supremum dependent case follows by noticing that step one remains valid when choosing $s = t = S$ since S is $\sigma(Y)$ -measurable.

APPENDIX

A.1. Stable and weak convergence. We briefly introduce the concept of stable convergence first appearing in Rényi [29].

DEFINITION A.1. Let \mathcal{F}^0 denote a sub- σ -field of \mathcal{F} . A sequence $(Z_n)_{n \in \mathbb{N}}$ of \mathcal{F}^0 -measurable random variables taking values in a Polish space E converges stably with respect to \mathcal{F}^0 to an E -valued \mathcal{F} -measurable random variable Z , if for every $A \in \mathcal{F}^0$ and continuous and bounded function $f : E \rightarrow \mathbb{R}$

$$\lim_{n \rightarrow \infty} \mathbb{E}[\mathbb{1}_A f(Z_n)] = \mathbb{E}[\mathbb{1}_A f(Z)].$$

We briefly write $Z^n \xrightarrow{\text{stably}} Z$.

Stable convergence admits various equivalent definitions.

THEOREM A.2. Let (Z_n) and Z be \mathcal{F}^0 -measurable, respectively, \mathcal{F} -measurable, random variables taking values in a Polish space E . The following statements are equivalent:

1. $Z_n \xrightarrow{\text{stably}} Z$ with respect to \mathcal{F}^0 ,
2. for all bounded \mathcal{F}^0 -measurable random variables U and all bounded and continuous functions $f : E \rightarrow \mathbb{R}$ one has

$$(A.1) \quad \lim_{n \rightarrow \infty} \mathbb{E}[Uf(Z_n)] = \mathbb{E}[Uf(Z)].$$

If $\mathcal{F}^0 = \sigma(Y)$ for a random variable Y taking values in a Polish space E' , then stable convergence is equivalent to weak convergence

$$(A.2) \quad (Y, Z_n) \Rightarrow (Y, Z) \quad \text{in } E \times E'.$$

PROOF. The first equivalence is an immediate consequence of the fact that the set of \mathcal{F}^0 -measurable random variables U for which

$$\lim_{n \rightarrow \infty} \mathbb{E}[Uf(Z_n)] = \mathbb{E}[Uf(Z)]$$

is true is linear and closed with respect to L^1 -norm. Further, (A.2) implies $Z_n \xrightarrow{\text{stably}} Z$ since the L^1 -closure of random variables $g(Y)$ with $g : E' \rightarrow \mathbb{R}$ bounded and continuous contains all indicators $\mathbb{1}_A$ with $A \in \mathcal{F}^0$. Conversely, assuming $Z_n \xrightarrow{\text{stably}} Z$, the sequence of random variables $((Y, Z_n) : n \in \mathbb{N})$ is tight in the product topology and for any $g : E' \rightarrow \mathbb{R}$ bounded and continuous one has

$\mathbb{E}[g(Y)f(Z_n)] \rightarrow \mathbb{E}[g(Y)f(Z)]$ which implies that $(Y, Z_n) \Rightarrow (Y, Z)$. The last statement is proved in complete analogy with the proof of the corresponding statement for weak convergence. \square

As the latter theorem shows, stable and weak convergence are intimately connected and we will make use of results of Jacod and Protter [17] on weak convergence for stochastic differential equations. For the statement, we need the concept of uniform tightness.

DEFINITION A.3. Let $(\mathcal{F}_t)_{t \in [0, T]}$ be a filtration and $(Z^n : n \in \mathbb{N})$ be a sequence of càdlàg (\mathcal{F}_t) -semimartingales. For $\delta > 0$ we represent each semimartingale uniquely in the form

$$Z_t^n = Z_0^n + A_t^{n, \delta} + M_t^{n, \delta} + \sum_{s \leq t} \Delta Z_s^n \mathbb{1}_{\{|\Delta Z_s^n| > \delta\}} \quad \text{for } t \in [0, T],$$

where $A^{n, \delta} = (A_t^{n, \delta})_{t \in [0, T]}$ is a càdlàg predictable process of finite variation and $M = (M_t^{n, \delta})_{t \in [0, T]}$ is a càdlàg local martingale, both processes starting in zero. We say that $(Z^n : n \in \mathbb{N})$ is *uniformly tight*, if the sequence,

$$\langle M^{n, \delta}, M^{n, \delta} \rangle_T + \int_0^T |dA_s^{n, \delta}| + \sum_{0 \leq s \leq T} |\Delta Z_s^{n, i}| \mathbb{1}_{\{|\Delta Z_s^{n, i}| > \delta\}}$$

is tight. The definition does not depend on the particular choice of δ . Multivariate processes are called uniformly tight if each component is uniformly tight.

We cite [17], Theorem 2.3, which is a consequence of [23].

THEOREM A.4. Let Z, Z^1, Z^2, \dots be càdlàg one-dimensional semimartingales and H be a càdlàg one-dimensional adapted process. If:

- (i) $(Z^n : n \in \mathbb{N})$ is uniformly tight and
- (ii) $((H, Z^n) : n \in \mathbb{N}) \Rightarrow (H, Z)$ in $D(\mathbb{R}^2)$,

then

$$\left(H, Z^n, \int_0^\cdot H_{s-} dZ_s^n : n \in \mathbb{N} \right) \Rightarrow \left(H, Z, \int_0^\cdot H_{s-} dZ_s \right) \quad \text{in } D(\mathbb{R}^3).$$

We state a consequence of [23], Theorem 8.2.

THEOREM A.5. Let H, Z, Z^1, Z^2, \dots be as in the previous theorem. Further, let Y be an adapted càdlàg semimartingale. We define $U^n := (U_t^n)_{t \in [0, T]}$ and $U := (U_t)_{t \in [0, T]}$ by

$$U_t^n = Z_t^n + \int_0^t U_{s-}^n H_{s-} dY_s, \quad U_t = Z_t + \int_0^t U_{s-} H_{s-} dY_s \quad \text{for } t \in [0, T].$$

If

$$\left(Z^n, \int_0^\cdot H_{s-} dY_s \right) \Rightarrow \left(Z, \int_0^\cdot H_{s-} dY_s \right) \quad \text{in } D(\mathbb{R}^2),$$

then

$$\left(Z^n, \int_0^\cdot H_{s-} dY_s, U^n \right) \Rightarrow \left(Z, \int_0^\cdot H_{s-} dY_s, U \right) \quad \text{in } D(\mathbb{R}^3).$$

The definition of uniform tightness and the two theorems above have natural extension to the multivariate setting and we refer the reader to [23] for more details. Further results about stable convergence of stochastic process can be found in [15] and [18].

A helpful lemma in the treatment of weak convergence is the following.

LEMMA A.6. *Let A, A^1, A^2, \dots be processes with trajectories in $\mathbb{D}(\mathbb{R}^d)$.*

1. *Suppose that for every $m \in \mathbb{N}$, $A^m, A^{1,m}, A^{2,m}, \dots$ are processes with trajectories in $\mathbb{D}(\mathbb{R}^d)$ such that:*

- (a) $\forall \delta > 0: \lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(\sup_{t \in [0, T]} |A_t^{n,m} - A_t^n| > \delta) = 0,$
- (b) $\lim_{m \rightarrow \infty} \mathbb{P}(\sup_{t \in [0, T]} |A_t^m - A_t| > \delta) = 0.$

Provided that one has convergence $A^{n,m} \Rightarrow A^m$ for every $m \in \mathbb{N}$, it is also true that

$$A^n \Rightarrow A.$$

2. *Suppose that B^1, B^2, \dots are processes with trajectories in $\mathbb{D}(\mathbb{R}^d)$ such that for all $\delta > 0$*

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\sup_{t \in [0, T]} |B_t^n - A_t^n| > \delta \right) = 0.$$

Then one has weak convergence $A^n \Rightarrow A$ if and only if $B^n \Rightarrow A$.

PROOF. To prove weak convergence on $\mathbb{D}(\mathbb{R}^d)$ it suffices to consider bounded and continuous test functions $f: \mathbb{D}(\mathbb{R}^d) \rightarrow \mathbb{R}$ that are additionally Lipschitz continuous with respect to supremum norm. Using this, it is elementary to verify the first statement. Further, the second statement is an immediate consequence of the first one. \square

REMARK A.7. In general, we call approximations $A^m, A^{1,m}, A^{2,m}, \dots$ with properties (a) and (b) of part one of the lemma *good approximations* for A, A^1, A^2, \dots . Further, approximations B^1, B^2, \dots as in part two will be called *asymptotically equivalent in ucp* to A^1, A^2, \dots .

A.2. Auxiliary estimates. We will make use of the following analogue of Lemma A.6 for tightness.

LEMMA A.8. *Let $(A_n)_{n \in \mathbb{N}}$ and, for every $m \in \mathbb{N}$, $(A_n^{(m)})_{n \in \mathbb{N}}$ be sequences of L^2 -integrable random variables. If*

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{E}[|A_n - A_n^{(m)}|^2] = 0$$

and, for every $m \in \mathbb{N}$, the sequence $(A_n^{(m)})_{n \in \mathbb{N}}$ is uniformly \mathbb{L}^2 -integrable, then also the sequence $(A_n)_{n \in \mathbb{N}}$ is uniformly \mathbb{L}^2 -integrable. In particular, if there is a sequence $(B_n)_{n \in \mathbb{N}}$ of uniformly \mathbb{L}^2 -integrable random variables with

$$\lim_{n \rightarrow \infty} \mathbb{E}[|B_n - A_n|^2] = 0,$$

then $(A_n)_{n \in \mathbb{N}}$ is uniformly \mathbb{L}^2 -integrable.

PROOF. For $\eta > 0$ and $n, m \in \mathbb{N}$, one has

$$\begin{aligned} \mathbb{E}[|A_n|^2 \mathbb{1}_{\{|A_n| \geq \eta\}}] &\leq 2\mathbb{E}[|A_n - A_n^{(m)}|^2] + 2\mathbb{E}[|A_n^{(m)}|^2 \mathbb{1}_{\{|A_n| \geq \eta\}}] \\ &\leq 2\mathbb{E}[|A_n - A_n^{(m)}|^2] + 2\mathbb{E}[|A_n^{(m)}|^2 \mathbb{1}_{\{|A_n^{(m)}| \geq \eta/2\}}] \\ &\quad + 2\mathbb{E}[|A_n^{(m)}|^2 \mathbb{1}_{\{|A_n^{(m)}| < \eta/2, |A_n - A_n^{(m)}| \geq \eta/2\}}] \\ &\leq 2\mathbb{E}[|A_n - A_n^{(m)}|^2] + 2\mathbb{E}[|A_n^{(m)}|^2 \mathbb{1}_{\{|A_n^{(m)}| \geq \eta/2\}}] \\ &\quad + \frac{\eta^2}{2} \mathbb{P}(|A_n - A_n^{(m)}| \geq \eta/2) \\ &\leq 4\mathbb{E}[|A_n - A_n^{(m)}|^2] + 2\mathbb{E}[|A_n^{(m)}|^2 \mathbb{1}_{\{|A_n^{(m)}| \geq \eta/2\}}], \end{aligned}$$

where we used Chebychew's inequality in the last step. Let now $\varepsilon > 0$. By assumption, we can choose m sufficiently large such that for all large n , say for $n \geq n_0$, $4\mathbb{E}[|A_n - A_n^{(m)}|^2] \leq \varepsilon/2$. Further, by the uniform L^2 -integrability of $(A_n^{(m)})_{n \in \mathbb{N}}$ we can choose η large to ensure that for all $n \in \mathbb{N}$, $2\mathbb{E}[|A_n^{(m)}|^2 \mathbb{1}_{\{|A_n^{(m)}| \geq \eta/2\}}] \leq \varepsilon/2$ so that $\mathbb{E}[|A_n|^2 \mathbb{1}_{\{|A_n| \geq \eta\}}] \leq \varepsilon$ for $n \geq n_0$. For $n = 1, \dots, n_0 - 1$ this estimate remains true for a sufficiently enlarged η , since finitely many L^2 -integrable random variables are always uniformly L^2 -integrable. \square

LEMMA A.9. *Let A_1, A_2, \dots be real random variables and let $(\varepsilon_k)_{k \in \mathbb{N}}$ satisfy (ML1) (see Section 1.2), and $L(\delta)$ and $n_k(\delta)$ be as in (1.13). Suppose that:*

1. $\text{Var}(\varepsilon_{k-1}^{-1/2} A_k) \rightarrow \zeta$ and
2. $(\varepsilon_{k-1}^{-1/2} A_k : k \in \mathbb{N})$ is L^2 -uniformly integrable.

Denote by $(A_{k,j} : k, j \in \mathbb{N})$ independent random variables with $\mathcal{L}(A_{k,j}) = \mathcal{L}(A_k)$. The random variables $(\widehat{S}_\delta : \delta \in (0, 1))$ given by

$$\widehat{S}_\delta := \sum_{k=1}^{L(\delta)} \frac{1}{n_k(\delta)} \sum_{j=1}^{n_k(\delta)} A_{k,j}$$

satisfy

$$\delta^{-1}(\widehat{S}_\delta - \mathbb{E}[\widehat{S}_\delta]) \Rightarrow \mathcal{N}(0, \zeta).$$

PROOF. Without loss of generality, we can and will assume that the random variables A_1, A_2, \dots have zero mean.

1st step. We first show that the variance of \widehat{S}_δ converges. One has

$$\text{Var}(\widehat{S}_\delta) = \sum_{k=1}^{L(\delta)} \frac{1}{n_k(\delta)} \text{Var}(A_k) = \sum_{k=1}^{L(\delta)} \underbrace{\left[\frac{\delta^2}{L(\delta)\varepsilon_{k-1}} \right]}_{=: a_{k,\delta}} \varepsilon_{k-1} \frac{\text{Var}(A_k)}{\varepsilon_{k-1}}.$$

It is elementary to verify that $\sum_{k=1}^{L(\delta)} (a_{k,\delta} \delta^{-2} - L(\delta)^{-1}) \rightarrow 0$ as $\delta \downarrow 0$. By the boundedness of $(\text{Var}(A_k)/\varepsilon_{k-1})_{k \in \mathbb{N}}$ one has

$$\left| \delta^{-2} \text{Var}(\widehat{S}_\delta) - \frac{1}{L(\delta)} \sum_{k=1}^{L(\delta)} \frac{\text{Var}(A_k)}{\varepsilon_{k-1}} \right| \rightarrow 0$$

and we get that $\lim_{\delta \downarrow 0} \text{Var}(\delta^{-1} \widehat{S}_\delta) = \zeta$ since the Césaro mean of a convergent sequence converges to its limit.

2nd step. In view of the Lindeberg condition (see, e.g., [20], Theorem 5.12), it suffices to verify that for arbitrarily fixed $\kappa > 0$ one has

$$\Sigma(\delta) := \sum_{k=1}^{L(\delta)} \sum_{j=1}^{n_k(\delta)} \mathbb{E} \left[\left(\frac{A_{k,j}}{\delta n_k(\delta)} \right)^2 \mathbb{1}_{\{|A_{k,j}/(\delta n_k(\delta))| > \kappa\}} \right] \rightarrow 0 \quad \text{as } \delta \downarrow 0.$$

We estimate

$$\Sigma(\delta) \leq \delta^{-2} \sum_{k=1}^{L(\delta)} \frac{\varepsilon_{k-1}}{n_k(\delta)} \mathbb{E} \left[\frac{A_k^2}{\varepsilon_{k-1}} \mathbb{1}_{\{|A_k|/\sqrt{\varepsilon_{k-1}} > \kappa \delta n_k^{(\delta)}/\sqrt{\varepsilon_{k-1}}\}} \right]$$

and note that for $k = 1, \dots, L(\delta)$

$$\varepsilon_{k-1} \geq \varepsilon_{L(\delta)-1} = T M^{-L(\delta)+1} \geq T \delta^2,$$

where we used that $\alpha \geq 1/2$ in the previous step. Hence, for these k , one has $\delta n_k^{(\delta)}/\sqrt{\varepsilon_{k-1}} \geq \delta^{-1} \sqrt{\varepsilon_{k-1}} L(\delta) \geq \sqrt{T} L(\delta)$. Consequently,

$$\Sigma(\delta) \leq \delta^{-2} \sum_{k=1}^{L(\delta)} \frac{\varepsilon_{k-1}}{n_k(\delta)} \mathbb{E} \left[\frac{A_k^2}{\varepsilon_{k-1}} \mathbb{1}_{\{|A_k|/\sqrt{\varepsilon_{k-1}} > \kappa \sqrt{T} L(\delta)\}} \right].$$

By uniform L^2 -integrability of $(A_k/\sqrt{\varepsilon_{k-1}})_{k \in \mathbb{N}}$ and the fact that $L(\delta) \rightarrow \infty$, we get that

$$\mathbb{E} \left[\frac{A_k^2}{\varepsilon_{k-1}} \mathbb{1}_{\{|A_k|/\sqrt{\varepsilon_{k-1}} > \kappa \sqrt{T} L(\delta)\}} \right] \leq a(\delta) \quad \text{for } k = 1, \dots, L(\delta),$$

with $(a_\delta)_{\delta \in (0,1)}$ being positive reals with $\lim_{\delta \downarrow 0} a_\delta = 0$. Hence, $\Sigma(\delta) \leq a_\delta \delta^{-2} \sum_{k=1}^{L(\delta)} \frac{\varepsilon_{k-1}}{n_k^{(\delta)}}$ and we remark that the analysis of step one yields equally well that $\delta^{-2} \sum_{k=1}^{L(\delta)} \frac{\varepsilon_{k-1}}{n_k^{(\delta)}}$ converges to a finite limit. \square

A.3. Estimates for Lévy-driven SDEs. Let $Y = (Y_t)_{t \in [0,T]}$ denote a square integrable Lévy process with triplet (b, σ^2, ν) .

LEMMA A.10. *Let (ε_n) and (h_n) be positive decreasing sequences such that*

$$\sup_{n \in \mathbb{N}} \nu(B(0, h_n)^c) \varepsilon_n < \infty.$$

One has

$$(A.3) \quad \varepsilon_n \left(\int_{B(0, h_n)^c} x \nu(dx) \right)^2 \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Further, if the limit $\lim_{n \rightarrow \infty} \nu(B(0, h_n)^c) \varepsilon_n =: \theta$ exists and is strictly positive, then $\lim_{n \rightarrow \infty} h_n/\sqrt{\varepsilon_n} = 0$. If additionally $\int x^2 \log^2(1 + 1/x) \nu(dx) < \infty$, then

$$(A.4) \quad \begin{aligned} \lim_{n \rightarrow \infty} \int_{B(0, h_n)} x^2 \nu(dx) \log^2 \left(1 + \frac{1}{\varepsilon_n} \right) &= 0 \quad \text{and} \\ \lim_{n \rightarrow \infty} \frac{h_n^2}{\varepsilon_n} \log^2 \left(1 + \frac{1}{\varepsilon_n} \right) &= 0. \end{aligned}$$

PROOF. One has for fixed $h > 0$ for all $n \in \mathbb{N}$ that

$$\begin{aligned} \varepsilon_n \left(\int_{B(0, h_n)^c} x \nu(dx) \right)^2 \\ \leq 2\varepsilon_n \left(\int_{B(0, h)^c} x \nu(dx) \right)^2 + 2\varepsilon_n \left(\int_{B(0, h) \setminus B(0, h_n)} x \nu(dx) \right)^2. \end{aligned}$$

The first term on the right-hand side tends to zero since ε_n tends to zero. Further, the Cauchy–Schwarz inequality yields for the second term

$$\varepsilon_n \left(\int_{B(0, h) \setminus B(0, h_n)} x \nu(dx) \right)^2 \leq \varepsilon_n \nu(B(0, h_n)^c) \int_{B(0, h)} x^2 \nu(dx).$$

By assumption, $(\varepsilon_n \nu(B(0, h_n)^c))$ is uniformly bounded and by choosing h arbitrarily small we can make the integral as small as we wish. This proves (A.3).

We assume that $\lim_{n \rightarrow \infty} \nu(B(0, h_n))^c \varepsilon_n =: \theta > 0$. The second statement follows by noting that

$$\theta^2 \frac{h_n^2}{\varepsilon_n} \sim \varepsilon_n h_n^2 \nu(B(0, h_n))^c \leq \varepsilon_n \left(\int_{B(0, h_n)^c} x \nu(dx) \right)^2 \rightarrow 0.$$

The first estimate in (A.4) follows from

$$\int_{B(0, h_n)} x^2 \nu(dx) \leq \underbrace{\int_{B(0, h_n)} x^2 \log^2 \left(1 + \frac{1}{x} \right) \nu(dx)}_{\rightarrow 0} (\log(1 + 1/h_n))^{-2}$$

and recalling that $h_n/\sqrt{\varepsilon_n} \rightarrow 0$. The second estimate in (A.4) follows in complete analogy to the proof of (A.3). \square

LEMMA A.11. *Let $p \geq 2$ and suppose that $\mathbb{E}[|Y_T|^p] < \infty$. Then there exists a finite constant κ such that for every predictable process H one has*

$$\mathbb{E} \left[\sup_{t \in [0, T]} \left| \int_0^t H_s \, dY_s \right|^p \right] \leq \kappa \int_0^T \mathbb{E}[|H_s|^p] \, ds.$$

If $p = 2$, one can choose $\kappa = 2b^2T + 8(\sigma^2 + \int x^2 \nu(dx))$.

PROOF. The proof is standard; see, for instance, [28], Theorem V.66. The explicit constant in the $p = 2$ case can be deduced with Doob’s L^2 -inequality and the Cauchy–Schwarz inequality. \square

LEMMA A.12. *Irrespective of the choice of the parameters (ε_n) and (h_n) , one has*

$$\sup_{n \in \mathbb{N}} \mathbb{E} \left[\sup_{t \in [0, T]} |X_t^n|^2 \right] < \infty.$$

The proof of the lemma is standard and can be found, for instance, in [22], Lemma 8.

A.4. Perturbation estimates for SDEs. In this section, we collect perturbation estimates for solutions of stochastic differential equations. For $n, m \in \mathbb{N}$, we denote by $Z^n, \bar{Z}^n, Z^{n,m}$ and $\bar{Z}^{n,m}$ càdlàg semimartingales and by Y a square integrable Lévy process all with respect to the same filtration. Further, let $H^n, H^{n,m}$ and H be càglàd adapted processes. We represent Y as in (1.2) and consider as approximations the processes $Y^m = (Y_t^m)_{t \in [0, T]}$ given by

$$Y_t^m = bt + \sigma W_t + \lim_{\delta \downarrow 0} \int_{(0, t] \times (V_m \setminus B(0, \delta))} x \, d\bar{\Pi}(s, x),$$

where V_1, V_2, \dots denote an increasing sequence of Borel sets with $\bigcup_{m \in \mathbb{N}} V_m = \mathbb{R} \setminus \{0\}$.

In the first part of the subsection, we derive perturbation estimates for the processes $\mathcal{U}^{n,m} = (\mathcal{U}_t^{n,m})_{t \in [0, T]}$ and $\bar{\mathcal{U}}^{n,m} = (\bar{\mathcal{U}}_t^{n,m})_{t \in [0, T]}$ given as solutions to

$$\mathcal{U}_t^{n,m} = \int_0^t \mathcal{U}_{s-}^{n,m} H_s^{n,m} dY_s^m + \mathcal{Z}_t^{n,m}$$

and

$$\bar{\mathcal{U}}_t^{n,m} = \int_0^t \bar{\mathcal{U}}_{s-}^{n,m} H_s^{n,m} dY_s + \bar{\mathcal{Z}}_t^{n,m}.$$

LEMMA A.13. *Suppose that*

$$(A.5) \quad \sup_{t \in [0, T]} |H_t^{n,m}| \quad \text{and} \quad \mathbb{E} \left[\sup_{t \in [0, T]} |\mathcal{Z}_t^{n,m}|^2 \right]$$

are uniformly bounded over all $n, m \in \mathbb{N}$. Then

$$\sup_{n, m \in \mathbb{N}} \mathbb{E} \left[\sup_{t \in [0, T]} |\mathcal{U}_t^{n,m}|^2 \right] < \infty.$$

PROOF. Suppose that the expressions in (A.5) are bounded by κ_1 , denote by \mathcal{T} a stopping time and define $z_{\mathcal{T}}(t) = \mathbb{E}[\sup_{s \in [0, t \wedge \mathcal{T}]} |\mathcal{U}_s^{n,m}|^2]$ for $t \in [0, T]$. By Lemma A.11, there exists a finite constant κ_2 such that

$$\begin{aligned} z_{\mathcal{T}}(t) &\leq 2\kappa_2 \int_0^t \mathbb{E}[\mathbb{1}_{\{s \leq \mathcal{T}\}} |\bar{\mathcal{U}}_{s-}^{n,m}|^2 |H_s^{n,m}|^2] ds + 2\mathbb{E} \left[\sup_{s \in [0, t]} |\mathcal{Z}_s^{n,m}|^2 \right] \\ &\leq 2\kappa_2 \kappa_1^2 \int_0^t z_{\mathcal{T}}(s) ds + 2\kappa_1. \end{aligned}$$

We replace \mathcal{T} by a localising sequence $(\mathcal{T}_k)_{k \in \mathbb{N}}$ of stopping times for which each $z_{\mathcal{T}_k}$ is finite and conclude with Gronwall's inequality that $z_{\mathcal{T}_k}$ is uniformly bounded over all $k \in \mathbb{N}$ and $n, m \in \mathbb{N}$. The result follows by monotone convergence. \square

LEMMA A.14. *Suppose that*

$$\sup_{t \in [0, T]} |H_t^{n,m}|$$

is uniformly bounded over all n, m and that $Y^m = Y$ for all $m \in \mathbb{N}$ or

$$\sup_{n, m \in \mathbb{N}} \mathbb{E} \left[\sup_{t \in [0, T]} |\mathcal{Z}_t^{n,m}|^2 \right] < \infty.$$

If additionally

$$(A.6) \quad \lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{E} \left[\sup_{t \in [0, T]} |\mathcal{Z}_t^{n,m} - \bar{\mathcal{Z}}_t^{n,m}|^2 \right] = 0,$$

then,

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{E} \left[\sup_{t \in [0, T]} |\mathcal{U}_t^{n,m} - \bar{\mathcal{U}}_t^{n,m}|^2 \right] \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

PROOF. We rewrite, for $t \in [0, T]$,

$$\begin{aligned} \mathcal{U}_t^{n,m} - \bar{\mathcal{U}}_t^{n,m} &= \int_0^t (\mathcal{U}_{s-}^{n,m} - \bar{\mathcal{U}}_{s-}^{n,m}) H_s^{n,m} dY_s - \int_0^t \mathcal{U}_{s-}^{n,m} H_s^{n,m} d(Y - Y^m)_s \\ &\quad + \mathcal{Z}_t^{n,m} - \bar{\mathcal{Z}}_t^{n,m}. \end{aligned}$$

We fix $n, m \in \mathbb{N}$ and consider $z(t) = \mathbb{E}[\sup_{s \in [0,t]} |\mathcal{U}_s^{n,m} - \bar{\mathcal{U}}_s^{n,m}|^2]$ for $t \in [0, T]$. Further, denote by κ_1 a uniform bound for $\sup_{n,m \in \mathbb{N}} \sup |H_t|^{n,m}$ and, if applicable, for $\sup_{n,m} \mathbb{E}[\sup_{t \in [0,T]} |\mathcal{Z}_t^{n,m}|^2]$. Using that $(a_1 + a_2 + a_3)^2 \leq 3(a_1^2 + a_2^2 + a_3^2)$ ($a_1, a_2, a_3 \in \mathbb{R}$) and Lemma A.11, we get that

$$\begin{aligned} z(t) &\leq 3\kappa_2\kappa_1^2 \int_0^t z(s) ds + 3\mathbb{E}\left[\sup_{s \in [0,t]} \left| \int_0^s \mathcal{U}_{s-}^{n,m} H_u^{n,m} d(Y - Y^m)_u \right|^2\right] \\ &\quad + 3\mathbb{E}\left[\sup_{s \in [0,t]} |\mathcal{Z}_s^{n,m} - \bar{\mathcal{Z}}_s^{n,m}|^2\right] \end{aligned}$$

with κ_2 being uniformly bounded. In view of (A.6), the statement follows with Gronwall's inequality, once we showed that

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{E}\left[\sup_{t \in [0,T]} \left| \int_0^t \mathcal{U}_{s-}^{n,m} H_s^{n,m} d(Y - Y^m)_s \right|^2\right] = 0.$$

If $Y = Y^m$, this is trivially true. In the remaining case, we can apply Lemma A.13 due to the uniform boundedness of $\mathbb{E}[\sup_{t \in [0,T]} |\mathcal{Z}_t^{n,m}|^2]$ and conclude with Doob's L^2 -inequality and the martingale property of $Y - Y^m$ that

$$\begin{aligned} \mathbb{E}\left[\sup_{t \in [0,T]} \left| \int_0^t \mathcal{U}_{s-}^{n,m} H_s^{n,m} d(Y - Y^m)_s \right|^2\right] &\leq 4 \int_0^T \mathbb{E}[|\mathcal{U}_{s-}^{n,m}|^2 |H_s^{n,m}|^2] d\langle Y - Y^m \rangle_s \\ &\leq 4\kappa_1^2 \kappa_3 T \int_{V_m^c} x^2 \nu(dx) \end{aligned}$$

with κ_3 denoting the constant appearing in Lemma A.13. All constants do not depend on n, m and the latter integral tends to 0 as $m \rightarrow \infty$. \square

We denote by τ_1, τ_2, \dots adapted càdlàg processes with $\tau_n(t) \leq t$ for all $t \in [0, T]$ and focus on perturbation estimates for the processes $\mathcal{U}^n = (\mathcal{U}_t^n)_{t \in [0,T]}$ and $\bar{\mathcal{U}}^n = (\bar{\mathcal{U}}_t^n)_{t \in [0,T]}$ given as solutions to

$$\mathcal{U}_t^n = \int_0^t \mathcal{U}_{\tau_n(s-)}^n H_s^n dY_s + \mathcal{Z}_t^n$$

and

$$\bar{\mathcal{U}}_t^n = \int_0^t \bar{\mathcal{U}}_{\tau_n(s-)}^n H_s^n dY_s + \bar{\mathcal{Z}}_t^n.$$

LEMMA A.15. 1. (Stochastic convergence) If:

- (a) $\tau_n(t) = t$, for $t \in [0, T]$,
- (b) $Z^n - \bar{Z}^n \rightarrow 0$ and $H^n - H \rightarrow 0$ in ucp, as $n \rightarrow \infty$, and
- (c) the sequences $(\sup_{t \in [0, T]} |Z_t^n| : n \in \mathbb{N})$ and $(\sup_{t \in [0, T]} |H_t^n| : n \in \mathbb{N})$ are tight,

then

$$U^n - \bar{U}^n \rightarrow 0 \quad \text{in ucp, as } n \rightarrow \infty.$$

2. (Moment estimates) Let $p \geq 2$. If:

- (a) Y has Lévy measure ν satisfying $\int |x|^p \nu(dx) < \infty$, and
- (b) the expressions

$$\sup_{t \in [0, T]} |H_t^n| \quad \text{and} \quad \mathbb{E} \left[\sup_{t \in [0, T]} |Z_t^n|^p \right]$$

are uniformly bounded over $n \in \mathbb{N}$,

then

$$\sup_{n \in \mathbb{N}} \mathbb{E} \left[\sup_{t \in [0, T]} |U_t^n|^p \right] < \infty.$$

PROOF. (1) Statement 1 follows when combining Theorems 2.5(b) and 2.3(d) in [17].

(2) Since $\int |x|^p \nu(dx) < \infty$ the process (Y_t) has bounded p th moment and the statement can be proved similarly as Lemma A.13 by using Lemma A.11 and Gronwall's inequality. \square

Acknowledgement. We thank two anonymous referees for their valuable comments.

REFERENCES

- [1] ANKIRCHNER, S., DEREICH, S. and IMKELLER, P. (2007). Enlargement of filtrations and continuous Girsanov-type embeddings. In *Séminaire de Probabilités XL. Lecture Notes in Math.* **1899** 389–410. Springer, Berlin. [MR2409018](#)
- [2] APPLEBAUM, D. (2009). *Lévy Processes and Stochastic Calculus*, 2nd ed. *Cambridge Studies in Advanced Mathematics* **116**. Cambridge Univ. Press, Cambridge. [MR2512800](#)
- [3] ASMUSSEN, S. and ROSIŃSKI, J. (2001). Approximations of small jumps of Lévy processes with a view towards simulation. *J. Appl. Probab.* **38** 482–493. [MR1834755](#)
- [4] BEN ALAYA, M. and KEBAIER, A. (2015). Central limit theorem for the multilevel Monte Carlo Euler method. *Ann. Appl. Probab.* **25** 211–234.
- [5] BERTOIN, J. (1996). *Lévy Processes. Cambridge Tracts in Mathematics* **121**. Cambridge Univ. Press, Cambridge. [MR1406564](#)
- [6] BRUTI-LIBERATI, N., NIKITPOULOS-SKLIBOSIOS, C. and PLATEN, E. (2006). First order strong approximations of jump diffusions. *Monte Carlo Methods Appl.* **12** 191–209. [MR2274692](#)

- [7] CREUTZIG, J., DEREICH, S., MÜLLER-GRONBACH, T. and RITTER, K. (2009). Infinite-dimensional quadrature and approximation of distributions. *Found. Comput. Math.* **9** 391–429. [MR2519865](#)
- [8] DEREICH, S. (2008). The coding complexity of diffusion processes under supremum norm distortion. *Stochastic Process. Appl.* **118** 917–937. [MR2418250](#)
- [9] DEREICH, S. (2011). Multilevel Monte Carlo algorithms for Lévy-driven SDEs with Gaussian correction. *Ann. Appl. Probab.* **21** 283–311. [MR2759203](#)
- [10] DEREICH, S. and HEIDENREICH, F. (2011). A multilevel Monte Carlo algorithm for Lévy-driven stochastic differential equations. *Stochastic Process. Appl.* **121** 1565–1587. [MR2802466](#)
- [11] DEREICH, S. and LI, S. (2015). Multilevel Monte Carlo implementation for SDEs driven by truncated stable processes. Preprint.
- [12] GILES, M. B. (2008). Multilevel Monte Carlo path simulation. *Oper. Res.* **56** 607–617. [MR2436856](#)
- [13] GLASSERMAN, P. (2004). *Monte Carlo Methods in Financial Engineering: Stochastic Modelling and Applied Probability. Applications of Mathematics (New York)* **53**. Springer, New York. [MR1999614](#)
- [14] HEINRICH, S. (2001). Multilevel Monte Carlo methods. In *Large-Scale Scientific Computing. Lecture Notes in Comput. Sci.* **2179** 58–67. Springer, Berlin.
- [15] JACOD, J. (1997). On continuous conditional Gaussian martingales and stable convergence in law. In *Séminaire de Probabilités, XXXI. Lecture Notes in Math.* **1655** 232–246. Springer, Berlin. [MR1478732](#)
- [16] JACOD, J. and MÉMIN, J. (1981). Weak and strong solutions of stochastic differential equations: Existence and stability. In *Stochastic Integrals (Proc. Sympos., Univ. Durham, Durham, 1980). Lecture Notes in Math.* **851** 169–212. Springer, Berlin. [MR0620991](#)
- [17] JACOD, J. and PROTTER, P. (1998). Asymptotic error distributions for the Euler method for stochastic differential equations. *Ann. Probab.* **26** 267–307. [MR1617049](#)
- [18] JACOD, J. and SHIRYAEV, A. N. (2003). *Limit Theorems for Stochastic Processes*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **288**. Springer, Berlin. [MR1943877](#)
- [19] JEULIN, T. and YOR, M., eds. (1985). *Grossissements de Filtrations: Exemples et Applications. Lecture Notes in Math.* **1118**. Springer, Berlin. [MR0884713](#)
- [20] KALLENBERG, O. (2002). *Foundations of Modern Probability*, 2nd ed. *Probability and Its Applications (New York)*. Springer, New York. [MR1876169](#)
- [21] KLOEDEN, P. E. and PLATEN, E. (1992). *Numerical Solution of Stochastic Differential Equations. Applications of Mathematics (New York)* **23**. Springer, Berlin. [MR1214374](#)
- [22] KOHATSU-HIGA, A. and TANKOV, P. (2010). Jump-adapted discretization schemes for Lévy-driven SDEs. *Stochastic Process. Appl.* **120** 2258–2285. [MR2684745](#)
- [23] KURTZ, T. G. and PROTTER, P. E. (1996). Weak convergence of stochastic integrals and differential equations. In *Probabilistic Models for Nonlinear Partial Differential Equations (Montecatini Terme, 1995). Lecture Notes in Math.* **1627** 1–41. Springer, Berlin. [MR1431298](#)
- [24] LI, S. (2015). Multilevel Monte Carlo for Lévy-driven SDEs. Ph.D. thesis, in preparation.
- [25] MAGHSOODI, Y. (1996). Mean square efficient numerical solution of jump-diffusion stochastic differential equations. *Sankhyā Ser. A* **58** 25–47. [MR1659067](#)
- [26] MORDECKI, E., SZEPESSY, A., TEMPONE, R. and ZOURARIS, G. E. (2008). Adaptive weak approximation of diffusions with jumps. *SIAM J. Numer. Anal.* **46** 1732–1768. [MR2399393](#)
- [27] PLATEN, E. (1982). An approximation method for a class of Itô processes with jump component. *Litovsk. Mat. Sb.* **22** 124–136. [MR0659025](#)

- [28] PROTTER, P. E. (2005). *Stochastic Integration and Differential Equations*, 2nd ed. *Stochastic Modelling and Applied Probability* **21**. Springer, Berlin.
- [29] RÉNYI, A. (1963). On stable sequences of events. *Sankhyā Ser. A* **25** 293–302. [MR0170385](#)
- [30] ROSIŃSKI, J. (2001). Series representations of Lévy processes from the perspective of point processes. In *Lévy Processes* 401–415. Birkhäuser, Boston, MA. [MR1833707](#)
- [31] SATO, K.-I. (1999). *Lévy Processes and Infinitely Divisible Distributions*. *Cambridge Studies in Advanced Mathematics* **68**. Cambridge Univ. Press, Cambridge. [MR1739520](#)

INSTITUTE FOR MATHEMATICAL STATISTICS
WESTFÄLISCHE WILHELMS-UNIVERSITÄT MÜNSTER
ORLÉANS-RING 10
48149 MÜNSTER
GERMANY
E-MAIL: steffen.dereich@wwu.de
sangmeng.li@wwu.de