



Wasserstein convergence rates for random bit approximations of continuous Markov processes



Stefan Ankirchner^a, Thomas Kruse^b, Mikhail Urusov^{c,*}

^a Institute of Mathematics, University of Jena, Ernst-Abbe-Platz 2, 07745 Jena, Germany

^b Institute of Mathematics, University of Gießen, Arndtstr. 2, 35392 Gießen, Germany

^c Faculty of Mathematics, University of Duisburg-Essen, Thea-Leymann-Str. 9, 45127 Essen, Germany

ARTICLE INFO

Article history:

Received 30 July 2019

Available online 28 August 2020

Submitted by A. Jentzen

Keywords:

One-dimensional Markov process

Speed measure

Markov chain approximation

Numerical scheme

Rate of convergence

Wasserstein distance

ABSTRACT

We determine the convergence speed of a numerical scheme for approximating one-dimensional continuous strong Markov processes. The scheme is based on the construction of certain Markov chains whose laws can be embedded into the process with a sequence of stopping times. Under a mild condition on the process' speed measure we prove that the approximating Markov chains converge at fixed times at the rate of $1/4$ with respect to every p -th Wasserstein distance. For the convergence of paths, we prove any rate strictly smaller than $1/4$. Our results apply, in particular, to processes with irregular behavior such as solutions of SDEs with irregular coefficients and processes with sticky points.

© 2020 Elsevier Inc. All rights reserved.

0. Introduction

In this article we analyze the convergence speed in every p -th Wasserstein distance of a numerical scheme, developed in [6], that allows to approximate the law of any one-dimensional regular continuous strong Markov process (in the sense of Section VII.3 in [36] or Section V.7 in [37]). In the following we refer to the latter processes as *general diffusions*.

The set of general diffusions includes any one-dimensional stochastic process that can be described as a strong or weak solution of a time-homogeneous stochastic differential equation (SDE) with possibly irregular coefficients. There are, however, many general diffusions that cannot be characterized in terms of an SDE. This is, in particular, true for diffusions with sticky features, where a sticky point is located in the interior of the state space. A related interesting phenomenon is slow (or sticky) reflection. In this case the sticky point is located at the boundary of the state space. Recent years have witnessed an increased interest in diffusions with sticky features, see [7], [9], [11], [14], [17], [18], [20], [23], [28], [29] and references therein.

* Corresponding author.

E-mail addresses: s.ankirchner@uni-jena.de (S. Ankirchner), thomas.kruse@math.uni-giessen.de (T. Kruse), mikhail.urusov@uni-due.de (M. Urusov).

Newly, diffusions with slow reflection were applied in [10] to provide bounds (via sticky couplings) for the distance between two multidimensional diffusions with different drifts. Diffusions with slow reflection also attracted interest in economic theory, where such processes characterize optimal continuation values in dynamic principal-agent problems (see, e.g., [39] and [35]). We emphasize that the numerical schemes studied in this paper are able to approximate general diffusions with irregularities such as, e.g., the processes mentioned above.

In order to explain our results, let $Y = (Y_t)_{t \in [0, \infty)}$ be a general diffusion in natural scale with speed measure m . Assume for the rest of the introduction that the state space of Y is equal to the whole real line. In addition, let $(\xi_k)_{k \in \mathbb{N}}$ be an iid sequence of random variables, on a probability space with a measure P , satisfying $P(\xi_k = \pm 1) = \frac{1}{2}$. Let $\bar{h} \in (0, \infty)$ and let $a_h: \mathbb{R} \rightarrow [0, \infty)$, $h \in (0, \bar{h})$, be a family of functions. Given some starting point $y \in \mathbb{R}$, for $h \in (0, \bar{h})$ we denote by $X^h = (X_{kh}^h)_{k \in \mathbb{N}_0}$ the Markov chain defined by

$$X_0^h = y \quad \text{and} \quad X_{(k+1)h}^h = X_{kh}^h + a_h(X_{kh}^h)\xi_{k+1}, \quad \text{for } k \in \mathbb{N}_0. \quad (1)$$

We extend X^h to a continuous-time process by linearly interpolating two neighboring sequence elements; more precisely, we define

$$X_t^h = X_{\lfloor t/h \rfloor h}^h + (t/h - \lfloor t/h \rfloor)(X_{(\lfloor t/h \rfloor + 1)h}^h - X_{\lfloor t/h \rfloor h}^h), \quad t \in [0, \infty). \quad (2)$$

In what follows, the term “Markov chain” is used also for the continuous-time process X^h because it is always clear from the context whether we speak about the discrete-time Markov chain (1) or about the linearly-interpolated continuous-time process X^h of (2). The functional limit theorem in [6] says that if for all compact sets $K \subset \mathbb{R}$ the functions $(a_h)_{h \in (0, \bar{h})}$ satisfy

$$\sup_{y \in K} \left| \frac{1}{2} \int_{(y-a_h(y), y+a_h(y))} (a_h(y) - |u-y|) m(du) - h \right| \in o(h), \quad (3)$$

then the associated Markov chains $(X^h)_{h \in (0, \bar{h})}$ converge in distribution to Y , as $h \rightarrow 0$. To motivate the expression on the left-hand side of (3), we remark that, for $y \in \mathbb{R}$ and $a > 0$, it holds

$$E_y[H_{y-a, y+a}(Y)] = \frac{1}{2} \int_{(y-a, y+a)} (a - |u-y|) m(du) \quad (4)$$

(see Remark 1.2 in [6]), where $H_{y-a, y+a}(Y) = \inf\{t \geq 0 : Y_t \notin (y-a, y+a)\}$ ($\inf \emptyset := \infty$), and E_y denotes the expectation under the measure P_y under which the Markov process Y starts in y : $P_y(Y_0 = y) = 1$. Moreover, it is discussed in [6] that, for any speed measure m , there exist families $(a_h)_{h \in (0, \bar{h})}$ satisfying (3), i.e., any general diffusion can be approximated in law in this way. Since random samples of X^h , $h \in (0, \bar{h})$, can be efficiently generated on a computer, this result opens the door for numerical approximations of the distribution of Y via Monte Carlo simulations of X^h with h chosen small enough.

Notice that we use symmetric Bernoulli random variables ξ_k , $k \in \mathbb{N}$, in (1) rather than say Gaussian ones. There are several reasons for that. Firstly, this is especially convenient in the case when the state space of Y is a bounded subinterval in \mathbb{R} because, by a suitable choice of the functions a_h , $h \in (0, \bar{h})$, we can guarantee that the approximations X^h never leave the state space of Y (see the precise description of the setting in Section 1). Since normal distributions have unbounded support, this would be impossible in the case of Gaussian ξ_k regardless of the choice of a_h . Secondly, there are technical arguments in [6] that do not go through with Gaussian ξ_k , i.e., the functional limit theorem with such a general scope (capable of approximating *any* general diffusion) exists by now for symmetric Bernoulli ξ_k but not for Gaussian ones.

Finally, on certain machines (like field programmable gate arrays) it is more efficient to use *random bits* (as opposed to *random numbers*), see [8] and [16] for more detail. The algorithms based on (1) require only to generate random bits and hence can be very efficiently implemented on such machines. We, therefore, sometimes refer to the processes $X^h = (X_t^h)_{t \in [0, \infty)}$, $h \in (0, \bar{h})$, as *random bit approximations* of Y or as *random bit Markov chains*.

In the present paper we address the question of how *fast* the Markov chains $(X^h)_{h \in (0, \bar{h})}$ converge to Y . In order to obtain our results we impose a stronger assumption than (3). Essentially, we assume that there exists $\lambda \in (0, \infty)$ such that the functions $(a_h)_{h \in (0, \bar{h})}$ satisfy

$$\sup_{y \in \mathbb{R}} \left| \frac{1}{2} \int_{(y-a_h(y), y+a_h(y))} (a_h(y) - |u-y|) m(du) - h \right| \in \mathcal{O}(h^{1+\lambda}), \quad (5)$$

which is an assumption on the approximation scheme, and we discuss that there always exist approximation schemes satisfying (5). In addition, we assume that the Cauchy distribution has a bounded density with respect to the speed measure of Y , i.e., that there exist $k_1 \in (0, \infty)$ and $k_2 \in \{0, 1\}$ such that

$$m(dx) \geq \frac{2}{k_1(1+k_2x^2)} dx \quad (6)$$

(see Remark 1.1 for why we include a binary variable k_2 in (6)). It is worth noting that (6) does not exclude sticky features mentioned above, as the latter are modeled via atoms in the speed measure. Our first convergence result, Theorem 1.4 below, states, for every finite time $T \in (0, \infty)$ and every $p \in [1, \infty)$, that if (5) and (6) are satisfied, then the p -th Wasserstein distance between the law of X_T^h and the law of Y_T converges to zero at least at the rate $\min(\frac{1}{4}, \frac{\lambda}{2})$, i.e., the p -th Wasserstein distance between X_T^h and Y_T is bounded by a constant times $h^{\min(\frac{1}{4}, \frac{\lambda}{2})}$. Under the same assumptions the p -th Wasserstein distance between the laws of the paths of Y and X^h , up to T , converge at least at the rate $\min(\frac{1}{4}, \frac{\lambda}{2}) - \varepsilon$, where ε is a positive real arbitrarily close to zero (see Theorem 1.6 below). We remark that the convergence results apply also to general diffusions whose state spaces are intervals, but not necessarily the whole real line. In such cases Conditions (5) and (6) have to be slightly modified (cf. Condition (A λ) and Condition (C) below).

To provide more details for Condition (5), we note that for all $y \in \mathbb{R}$, $h \in (0, \bar{h})$ there always exists $\hat{a}_h(y) \in (0, \infty)$ such that

$$\frac{1}{2} \int_{(y-\hat{a}_h(y), y+\hat{a}_h(y))} (\hat{a}_h(y) - |u-y|) m(du) = h. \quad (7)$$

In particular, the family $(\hat{a}_h)_{h \in (0, \bar{h})}$ satisfies (5) for all $\lambda \in (0, \infty)$ and, therefore, the path distribution of *any* general diffusion satisfying (6) can be approximated by random bit Markov chains at any Wasserstein rate strictly smaller than $1/4$. In practice, the solutions $(\hat{a}_h)_{h \in (0, \bar{h})}$ cannot be determined in closed form but need to be approximated numerically themselves. Therefore, results for the scheme $(\hat{a}_h)_{h \in (0, \bar{h})}$ of (7) should be complemented by a *perturbation analysis* for (7). This analysis is carried out in the present article and leads to the relaxed version (5) of (7). Theorem 1.6 thus ensures that if Equation (7) is only solved with a precision of order $\mathcal{O}(h^{3/2})$, then the associated Markov chains $(X^h)_{h \in (0, \bar{h})}$ still converge at any rate arbitrarily close to $1/4$ (and, for the convergence of time marginals, Theorem 1.4 provides the rate $1/4$). A further reduction of the precision entails a smaller rate.

Concerning Condition (6), we note that this assumption is not required for the functional limit theorem in [6]. As we measure the convergence speed here with respect to Wasserstein distances of all orders $p \in [1, \infty)$, we necessarily have to ensure that Y and X^h , $h \in (0, \bar{h})$, admit finite moments of all orders. This is achieved

under (6) in Section 3 below. In fact, (6) is nearly a minimal assumption required for our results (see Section 7).

The proofs of our results rely on embeddings of the random bit Markov chains into the diffusion Y with sequences of stopping times. More precisely, for every $h \in (0, \bar{h})$ we define the sequence of stopping times $(\tau_k^h)_{k \in \mathbb{N}_0}$ by the formulas $\tau_0^h = 0$, $\tau_{k+1}^h = \inf\{t \geq \tau_k^h : Y_t \notin (Y_{\tau_k^h} - a_h(Y_{\tau_k^h}), Y_{\tau_k^h} + a_h(Y_{\tau_k^h}))\}$, $k \in \mathbb{N}_0$, and observe that the sequence $(Y_{\tau_k^h})_{k \in \mathbb{N}_0}$ has the same law as the sequence $(X_{kh}^h)_{k \in \mathbb{N}_0}$. Condition (5) allows to conclude that $\tau_{\lfloor T/h \rfloor}^h$ converges to T with respect to every L^p -norm, with $p \in [1, \infty)$, at a rate of at least $\min\{\frac{1}{2}, \lambda\}$. At this point the latter statement can be informally explained by (4). With Condition (6) and a representation of Y as a time-changed Brownian motion we can estimate the L^p -distance between Y_T and $Y_{\tau_{\lfloor T/h \rfloor}^h}$ against the distance between $\tau_{\lfloor T/h \rfloor}^h$ and T , and conclude that $Y_{\tau_{\lfloor T/h \rfloor}^h}$ converges to Y_T in L^p at a rate of at least $\min\{\frac{1}{4}, \frac{\lambda}{2}\}$. We stress that our analysis is optimal. Indeed, we show by means of an example that the rate of L^p -convergence for $p \in [4, \infty)$ cannot be improved beyond $1/4$. As the pair $(Y_T, Y_{\tau_{\lfloor T/h \rfloor}^h})$ constitutes a coupling between the laws of Y_T and X_{kh}^h , we arrive at an estimate for the convergence rate with respect to the p -th Wasserstein distance. A similar procedure can be used for proving the convergence rate for the path distributions.

The results in the present article generalize the results of our article [5] in several directions. In [5] we do not allow for *all* general diffusions, but consider only diffusions that solve an SDE of the form

$$dY_t = \eta(Y_t) dW_t. \quad (8)$$

It is known that (8) has a unique in law weak solution under the Engelbert-Schmidt condition that η is a non-vanishing (possibly irregular) Borel function such that $1/\eta^2$ is locally integrable (see [12] or Theorem 5.5.7 in [24]). This is a special case of the setting considered in the present article with $m(dx) = \frac{2}{\eta^2(x)} dx$. The article [5] does not provide a perturbation analysis, i.e., it only analyzes the situation, where Equation (7) can be solved *exactly* and not only with a precision of the order $\mathcal{O}(h^{1+\lambda})$. Moreover, [5] considers the convergence of the embedded Markov chains with respect to the L^2 -norm only and for the time marginals only. This neither allows to draw conclusions about the convergence of the time marginals in the Wasserstein distance of order $p > 2$ nor to say anything about the convergence speed for the path distributions. It is shown in [5] that the scheme has the rate $1/4$ for approximating the distribution of Y at single points in time under the assumption of global boundedness of $|\eta|$ and $1/|\eta|$. In the present article we replace the boundedness assumption by the weaker assumption (6). In the context of SDEs, (6) imposes a linear growth condition on η (but no regularity condition). Note, however, that (6) allows to go beyond SDEs, e.g., it includes diffusions with sticky features.

As noted in [1], the Wasserstein distance is an appropriate measure for the distance between the path distribution of a diffusion and its Markov chain approximation. In [1] every p -th Wasserstein distance between the path distribution of a Lipschitz continuous and uniformly elliptic SDE and its Euler approximation is shown to converge to zero with rate $\frac{2}{3} - \varepsilon$ for every positive real ε . As discussed above, the approximations of our algorithm also converge with respect to every p -th Wasserstein distance; we prove convergence of our algorithm in Wasserstein distances for a larger class of diffusions in natural scale but at a slower rate.

For classical results on the approximation of SDEs with Lipschitz coefficients via the Euler scheme we refer to the books [25] and [34] and to references therein. The Euler scheme is known to converge also for some SDEs with non-Lipschitz coefficients, and results on the convergence speed of the weak and strong approximation error exist; see, e.g., [26] and [33] and the references therein for results on the weak and strong convergence for the Euler-type approximations of SDEs with discontinuous coefficients (note, however, that the Euler scheme may fail to converge in the numerically weak sense even in the case of continuous coefficients, see [21]). [27] and [15] establish weak convergence rate of certain Euler-type schemes for diffusions with Hölder coefficients, and the rate is a half of the Hölder exponent. In contrast to our scheme, however, the Euler scheme is defined only for SDEs (not for *all* general diffusions) and may fail to

converge even in the stochastically weak sense when the SDE coefficients are irregular (see Section 5.4 in [5]).

The random bit Markov chains can be embedded into the general diffusion Y with a sequence of stopping times, and hence the Markov chains can be interpreted as exact appearances of Y along a stochastic time grid. There are a few related schemes in the literature employing random time grids to approximate solutions of one-dimensional SDEs. In [13] the authors construct Bernoulli random walks on finite grids in the space, create random walks on certain associated random time grids, and compute the rate with which these random walks converge in L^1 . A similar scheme is introduced in [32] for approximating the CIR process. In contrast to [13] the scheme of [32] is exact along a certain sequence of stopping times. A further scheme, exact on a time grid with exponentially distributed steps, and applicable to SDEs with discontinuous coefficients is suggested in [31]. The results in [13], [31] and [32] apply to SDEs with drift, whereas the present article puts a focus on arbitrary general diffusions in natural scale and hence has a different scope.

To summarize the discussion in the introduction, we can say that in the literature there are plenty of results that study approximations of solutions to SDEs. There are, however, many general diffusions that cannot be written as solutions to SDEs. This is, in particular, the case for general diffusions with sticky features and, in such a situation, reference [6] as well as the present paper come into play. While reference [6] deals with approximations in the sense of weak convergence, in the present paper we aim at approximating general diffusions in a stronger sense¹ and prove the rate of $1/4$. Essentially, the price for obtaining these stronger results is Condition (6), which is, as discussed, not required in [6] but is nearly a minimal assumption for the results of the present paper to hold. Also, the techniques used in the proofs of this paper differ from those in [6]: while in [6] we proceed through various moment estimates for the embedding stopping times, in the present paper we explicitly work with the random time change involved in the construction of a general diffusion. Finally, it is worth noting that the rate of $1/4$ cannot be considered as too slow, as in *almost all*² references above involving approximation methods for SDEs with some rate, the rate degenerates as the diffusion coefficient loses regularity. On the contrary, in our setting the diffusion coefficient of an SDE is allowed to be arbitrarily irregular (just a Borel function); the algorithm even allows to consider general diffusions with sticky points or the like, which cannot be written as solutions to SDEs; our scheme adapts in the way that preserves the rate of $1/4$ in any case.

The paper is organized as follows. In Section 1 we formally describe the processes we approximate, the approximation schemes used for this aim and formulate and discuss our main results. Section 2 studies some useful properties of the approximation schemes. Section 3 establishes moment bounds for the general diffusion and for the approximations. In Section 4 we describe the embedding of the approximating random bit Markov chains into the general diffusion and present (optimal) L^p -convergence rates for the involved stopping times. After the preparations done in Sections 2, 3 and 4, we prove our main results in Sections 5 and 6. Finally, in Section 7 we present examples showing that Condition (6) is essential for our results.

1. Approximation schemes and main results

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, (P_y)_{y \in I}, (Y_t)_{t \geq 0})$ be a one-dimensional continuous strong Markov process in the sense of Section VII.3 in [36]. We refer to this class of processes as *general diffusions* in the sequel. We assume that the state space is an open, half-open or closed interval $I \subseteq \mathbb{R}$. We denote by $I^\circ = (l, r)$ the interior of I , where $-\infty \leq l < r \leq \infty$, and we set $\bar{I} = [l, r]$. Recall that by the definition we have $P_y[Y_0 = y] = 1$ for

¹ Notice that the convergence in every p -th Wasserstein distance implies the convergence of expectations for all continuous path functionals with polynomial growth, whereas the weak convergence is the convergence of expectations for all bounded continuous path functionals.

² The exceptions are [5] and [13]. We recall that the paper [13] discusses different results from ours for an algorithm that has a different scope. As for [5], we refer to the paragraph containing (8), where we explain in detail that the present paper substantially generalizes [5] in many directions.

all $y \in I$. We further assume that Y is regular. This means that for every $y \in I^\circ$ and $x \in I$ we have that $P_y[H_x(Y) < \infty] > 0$, where $H_x(Y) = \inf\{t \geq 0 : Y_t = x\}$ (with the usual convention $\inf \emptyset = \infty$). If there is no ambiguity, we simply write H_x in place of $H_x(Y)$. Moreover, for $a < b$ in \bar{I} we denote by $H_{a,b} = H_{a,b}(Y)$ the first exit time of Y from (a, b) , i.e. $H_{a,b} = H_a \wedge H_b$. We suppose that the diffusion Y is in natural scale. If Y is not in natural scale, then there exists a strictly increasing continuous function $s : I \rightarrow \mathbb{R}$, the so-called scale function, such that $s(Y_t)$, $t \geq 0$, is in natural scale. Let m be the speed measure of the Markov process Y (see VII.3.7 and VII.3.10 in [36]). Recall that for all $a < b$ in I° we have

$$0 < m([a, b]) < \infty. \quad (9)$$

Finally, we always assume that if a boundary point is accessible, then it is absorbing. We refer to Remark 1.11 on how to drop this assumption and to also allow for reflecting boundaries.

Let $\bar{h} \in (0, 1)$ and suppose that for every $h \in (0, \bar{h})$ we are given a measurable function $a_h : \bar{I} \rightarrow [0, \infty)$ such that $a_h(l) = a_h(r) = 0$ and for all $y \in I^\circ$ we have $y \pm a_h(y) \in I$. We refer to each function a_h as a *scale factor*. We next construct a sequence of Markov chains associated to the family of scale factors $(a_h)_{h \in (0, \bar{h})}$. To this end fix a starting point $y \in I^\circ$ of Y . Let $(\xi_k)_{k \in \mathbb{N}}$ be an iid sequence of random variables, on a probability space with a measure P , satisfying $P(\xi_k = \pm 1) = \frac{1}{2}$. We denote by $(X_{kh}^h)_{k \in \mathbb{N}_0}$ the Markov chain defined by

$$X_0^h = y \quad \text{and} \quad X_{(k+1)h}^h = X_{kh}^h + a_h(X_{kh}^h)\xi_{k+1}, \quad \text{for } k \in \mathbb{N}_0. \quad (10)$$

We extend $(X_{kh}^h)_{k \in \mathbb{N}_0}$ to a continuous-time process $X^h = (X_t^h)_{t \in [0, \infty)}$ by linear interpolation, i.e., for all $t \in [0, \infty)$, we set

$$X_t^h = X_{\lfloor t/h \rfloor h}^h + (t/h - \lfloor t/h \rfloor)(X_{\lfloor t/h \rfloor h + 1}^h - X_{\lfloor t/h \rfloor h}^h). \quad (11)$$

The approximation schemes we consider in this paper are the families of processes $(X^h)_{h \in (0, \bar{h})}$ (which are encoded by the families $(a_h)_{h \in (0, \bar{h})}$ of the scale factors). To highlight the dependence of X^h on the starting point $y \in I^\circ$ we also sometimes write $X^{h,y}$.

To formulate our main results we need the additional assumption that the Cauchy distribution has a bounded density with respect to the speed measure. More precisely, we suppose that the following condition is satisfied.

Condition (C). There exist constants $k_1 \in (0, \infty)$, $k_2 \in \{0, 1\}$ such that on I° we have $m(dx) \geq \frac{2}{k_1(1+k_2x^2)} dx$, i.e., for all $A \in \mathcal{B}(I^\circ)$ we have $m(A) \geq \int_A \frac{2}{k_1(1+k_2x^2)} dx$.

Remark 1.1. In the formulation of Condition (C) we allow for two degrees of freedom $k_1 \in (0, \infty)$, $k_2 \in \{0, 1\}$. This might seem superfluous. Indeed, if Condition (C) is satisfied with $k_2 = 0$, then it also holds with $k_2 = 1$. In the results that follow, however, we get sharper bounds if Condition (C) is satisfied with $k_2 = 0$. In particular, the constants appearing in Theorem 1.4 and Theorem 1.6 do not depend on the initial value y of the diffusion Y in the case $k_2 = 0$.

Example 1.2 (*Driftless SDE with possibly irregular diffusion coefficient*). A particular case of our setting is the case, where Y is a solution to the driftless SDE

$$dY_t = \eta(Y_t) dW_t, \quad (12)$$

where $\eta : I^\circ \rightarrow \mathbb{R}$ is a Borel function satisfying the Engelbert-Schmidt conditions

$$\eta(x) \neq 0 \quad \forall x \in I^\circ, \quad (13)$$

$$\eta^{-2} \in L^1_{\text{loc}}(I^\circ) \quad (14)$$

($L^1_{\text{loc}}(I^\circ)$ denotes the set of Borel functions locally integrable on I°). Under (13)–(14) SDE (12) has a unique in law weak solution (see [12] or Theorem 5.5.7 in [24]). In this case the speed measure of Y on I° is given by the formula

$$m(dx) = \frac{2}{\eta^2(x)} dx. \quad (15)$$

We refer to Section 2 in [6] for further details on the application of the approximation scheme to SDEs. If m is given by (15), Condition (C) means that there exist constants $k_1 \in (0, \infty)$, $k_2 \in \{0, 1\}$ such that for all $x \in I^\circ$ it holds that $\eta(x) \leq \sqrt{k_1(1 + k_2 x^2)}$, i.e., the diffusion coefficient η is locally bounded and whenever the state space is unbounded η is of at most linear growth. If $k_2 = 0$, then under Condition (C) the diffusion Y does not move faster than a Brownian motion scaled by $\sqrt{k_1}$.

To formulate our second main assumption we need to introduce some notation. We introduce an auxiliary subset of I° . To this end, if $l > -\infty$, we define, for all $h \in (0, \bar{h})$,

$$l_h = l + \inf \left\{ a \in \left(0, \frac{r-l}{2} \right] : a < \infty \text{ and } \frac{1}{2} \int_{(l, l+2a)} (a - |u - (l+a)|) m(du) \geq h \right\}, \quad (16)$$

where we use the convention $\inf \emptyset = \frac{r-l}{2}$. If $l = -\infty$, we set $l_h = -\infty$. Similarly, if $r < \infty$, then we define, for all $h \in (0, \bar{h})$,

$$r_h = r - \inf \left\{ a \in \left(0, \frac{r-l}{2} \right] : a < \infty \text{ and } \frac{1}{2} \int_{(r-2a, r)} (a - |u - (r-a)|) m(du) \geq h \right\} \quad (17)$$

with the same convention $\inf \emptyset = \frac{r-l}{2}$. If $r = \infty$, we set $r_h = \infty$. In any case, it holds $l_h \leq r_h$ and, moreover, $l_h \leq \frac{l+r}{2} \leq r_h$ whenever l and r are finite. The auxiliary subset is defined by

$$I_h = (l_h, r_h) \cup \{y \in I^\circ : y \pm a_h(y) \in I^\circ\}. \quad (18)$$

With the help of Feller's test for explosions (see, e.g., Theorem 23.12 in [22] or Lemma 2.1 in [2] or Theorem 3.3 in [3]) one can show that l is inaccessible if and only if $l_h = l$ for all $h \in (0, \bar{h})$ (for details, see Remark 4.2 in [6]). Similarly, r is inaccessible if and only if $r_h = r$ for all $h \in (0, \bar{h})$. In particular, if both l and r are inaccessible, then it holds for all $h \in (0, \bar{h})$ that $I_h = I = I^\circ$. If l or r are accessible it holds that $l_h \searrow l$ or $r_h \nearrow r$, respectively, as $h \searrow 0$.

The second assumption we need for our main results is the following condition on the scale factors.

Condition (A λ). There exist $\lambda \in (0, \infty)$, $K \in [0, \infty)$ such that for all $h \in (0, \bar{h})$ it holds

$$\sup_{y \in I_h} \left| \frac{1}{2} \int_{(y-a_h(y), y+a_h(y))} (a_h(y) - |u - y|) m(du) - h \right| \leq K h^{1+\lambda} \quad (19)$$

For every $\lambda \in (0, \infty)$ Condition (A λ) implies Condition (A) in [6]. We refer to the introduction in [6] for an interpretation of the left-hand side of (19) as an upper bound for the one-step temporal error of the approximation scheme.

Throughout the paper we make a convention that if Condition (A λ) is satisfied for some $\lambda \in (0, \infty)$, then \bar{h} is chosen small enough such that for all $h \in (0, \bar{h})$ it holds

$$\sup_{y \in I_h} \left| \frac{1}{2} \int_{(y-a_h(y), y+a_h(y))} (a_h(y) - |u-y|) m(du) - h \right| \leq \gamma h \quad (20)$$

with some $\gamma \in [0, 1)$. In fact, in many auxiliary results that follow (see Sections 2 and 3) we assume only the weaker Condition (20). The stronger Condition (A λ) is generally needed to get rates of convergence.

Before formulating our main results we present an example of an approximation scheme, called EMCEL in [6], which is well-defined for every general diffusion Y and satisfies Condition (A λ) for all $\lambda \in (0, \infty)$.

Example 1.3 (EMCEL approximations). Let $h \in (0, \bar{h})$. The EMCEL(h) scale factor \hat{a}_h is defined by $\hat{a}_h(l) = \hat{a}_h(r) = 0$ and, for all $y \in I^\circ$,

$$\hat{a}_h(y) = \sup \left\{ a \geq 0 : y \pm a \in I \text{ and } \frac{1}{2} \int_{(y-a, y+a)} (a - |z-y|) m(dz) \leq h \right\}. \quad (21)$$

The associated process defined in (10)–(11) is denoted by \hat{X}^h and referred to as Embeddable Markov Chain with Expected time Lag h (we write shortly $\hat{X}^h \in \text{EMCEL}(h)$).³ Expression (21) can be alternatively described as follows. For $y \in (l_h, r_h)$, $\hat{a}_h(y)$ is a unique positive root of the equation (in a)

$$\frac{1}{2} \int_{(y-a, y+a)} (a - |z-y|) m(dz) = h, \quad (22)$$

while, for $y \in (l, l_h]$ (resp., $y \in [r_h, r)$), $\hat{a}_h(y)$ is chosen to satisfy

$$y - \hat{a}_h(y) = l \quad (\text{resp., } y + \hat{a}_h(y) = r). \quad (23)$$

It follows from (23) that the set I_h of (18) corresponding to the EMCEL(h) scale factor \hat{a}_h (and, naturally, denoted by \hat{I}_h) is simply

$$\hat{I}_h = (l_h, r_h).$$

This yields that the left-hand side of (19) vanishes for the EMCEL approximations and, therefore, for such a scheme Condition (A λ) is satisfied for all $\lambda \in (0, \infty)$.

We first formulate our main result concerning the Wasserstein approximation rate for time marginals of Y . To this end we briefly recall the definitions of Wasserstein distances. For $p \in [1, \infty)$ let $\mathcal{M}_p(\mathbb{R})$ denote the set of all probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with finite p -th moment, i.e., probability measures μ satisfying $\int |x|^p \mu(dx) < \infty$. The p -th Wasserstein distance between two probability measures $\mu, \nu \in \mathcal{M}_p(\mathbb{R})$ is defined by

³ The whole family $(\hat{X}^h)_{h \in (0, \bar{h})}$ is referred to as the *EMCEL approximation scheme*. Alternatively, we simply say *EMCEL approximations*.

$$\mathcal{W}_p(\mu, \nu) = \inf \|\xi - \zeta\|_{L^p},$$

where the infimum is taken over all *couplings* between μ and ν , that is, over all random vectors (ξ, ζ) with marginals μ and ν (i.e., $\xi \sim \mu$ and $\zeta \sim \nu$).

Theorem 1.4. Suppose that Condition (C) is satisfied and that there exists $\lambda \in (0, \infty)$ such that Condition (A λ) holds. Let $p \in [1, \infty)$ and $T \in (0, \infty)$. Then there exists a constant $C(p, T) \in [0, \infty)$ such that for all $y \in I^\circ$ and $h \in (0, \bar{h})$ it holds $Y_T \in L^p(P_y)$, $X_{h\lfloor T/h \rfloor}^{h,y} \in L^p(P)$, $X_T^{h,y} \in L^p(P)$,

$$\mathcal{W}_p(P \circ (X_{h\lfloor T/h \rfloor}^{h,y})^{-1}, P_y \circ (Y_T)^{-1}) \leq C(p, T)(1 + k_2|y|)h^{\min\{\frac{1}{4}, \frac{\lambda}{2}\}} \quad (24)$$

and

$$\mathcal{W}_p(P \circ (X_T^{h,y})^{-1}, P_y \circ (Y_T)^{-1}) \leq C(p, T)(1 + k_2|y|)h^{\min\{\frac{1}{4}, \frac{\lambda}{2}\}}. \quad (25)$$

Remark 1.5. In the case where Y does not move faster than a scaled Brownian motion (i.e., $k_2 = 0$), the convergence in Theorem 1.4 is uniform in the starting point $y \in I^\circ$ (the same applies to Theorem 1.6 below).

The proof of Theorem 1.4 is provided in Section 5. Next, we describe the rate at which the law of $X^{h,y}$ converges to the one of Y with respect to the p -th Wasserstein distance on the path space $C([0, T], I)$ endowed with the sup norm. In what follows, we use the notation

$$\|x\|_C = \sup_{t \in [0, T]} |x(t)|$$

for the sup norm on $C([0, T], I)$. Let $\mathcal{M}_p(C([0, T], I))$ denote the set of all probability measures μ on $C([0, T], I)$ (equipped with the Borel σ -field) with finite p -th moment, that is, $\int \|x\|_C^p \mu(dx) < \infty$. The p -th Wasserstein distance between $\mu, \nu \in \mathcal{M}_p(C([0, T], I))$ is defined by

$$\mathcal{W}_p(\mu, \nu) = \inf \left\| \sup_{t \in [0, T]} |\xi_t - \zeta_t| \right\|_{L^p},$$

where the infimum is taken over all couplings between μ and ν , i.e., over all random elements (ξ, ζ) taking values in $C([0, T], I^2)$ with marginals μ and ν (that is, $\xi \sim \mu$ and $\zeta \sim \nu$).

Theorem 1.6. Suppose that Condition (C) is satisfied and that there exists $\lambda \in (0, \infty)$ such that Condition (A λ) holds. Let $p \in [1, \infty)$, $T \in (0, \infty)$ and $\varepsilon \in (0, \min\{\frac{1}{4}, \frac{\lambda}{2}\})$. Then there exists a constant $C(p, \varepsilon, T) \in [0, \infty)$ such that for all $y \in I^\circ$ and $h \in (0, \bar{h})$ it holds that

$$P_y \circ (Y)^{-1} \text{ and } P \circ (X^{h,y})^{-1} \text{ are in } \mathcal{M}_p(C([0, T], I)) \quad (26)$$

and

$$\mathcal{W}_p(P \circ (X^{h,y})^{-1}, P_y \circ (Y)^{-1}) \leq C(p, \varepsilon, T)(1 + k_2|y|)h^{\min\{\frac{1}{4}, \frac{\lambda}{2}\} - \varepsilon}.$$

The proof of Theorem 1.6 is provided in Section 6.

In practice one is often interested in approximating $E_y[f(Y_T)]$ or $E_y[F(Y)]$ for functions $f: I \rightarrow \mathbb{R}$ and path functionals $F: C([0, T], I) \rightarrow \mathbb{R}$. The next result presents the respective rates.

Theorem 1.7. Suppose that Condition (C) is satisfied and that there exists $\lambda \in (0, \infty)$ such that Condition (A λ) holds. Let $T \in (0, \infty)$ and let $F: C([0, T], I) \rightarrow \mathbb{R}$ be a locally Lipschitz continuous path functional with polynomially growing Lipschitz constant, i.e., there exist $L, \alpha \in [0, \infty)$ such that for all $x_1, x_2 \in C([0, T], I)$ it holds

$$|F(x_1) - F(x_2)| \leq L \{1 + (\|x_1\|_C \vee \|x_2\|_C)^\alpha\} \|x_1 - x_2\|_C. \quad (27)$$

Then for every $\varepsilon \in (0, \min\{\frac{1}{4}, \frac{\lambda}{2}\})$ there exist a constant $C(\alpha, \varepsilon, T) \in [0, \infty)$ such that for all $h \in (0, \bar{h})$ and $y \in I^\circ$ it holds

$$\left| E \left[F(X_t^{h,y}; t \in [0, T]) \right] - E_y \left[F(Y_t; t \in [0, T]) \right] \right| \leq LC(\alpha, \varepsilon, T)(1 + |y|^\alpha)(1 + k_2|y|)h^{\min\{\frac{1}{4}, \frac{\lambda}{2}\} - \varepsilon}, \quad (28)$$

where we use the convention $0^0 := 1$ (such an expression can appear in the case $\alpha = 0$).

Remark 1.8. Note that both expectations in (28) are well-defined due to (27) (which implies polynomial growth of F) and existence of all polynomial moments (see (26)). Furthermore, we stress that the special case of a globally Lipschitz path functional F is explicitly included in (27) by the possibility $\alpha = 0$ and results in a better dependence on y on the right-hand side of (28).

Theorem 1.7 is essentially deduced from Theorem 1.6, but we also need several other technical results. Therefore, we postpone the proof till Section 6.

In the same vein, we get the rate $\min\{\frac{1}{4}, \frac{\lambda}{2}\}$ when approximating $E_y[f(Y_T)]$ for a locally Lipschitz continuous function $f: I \rightarrow \mathbb{R}$ with polynomially growing Lipschitz constant. A precise formulation is similar to that of Theorem 1.7. In the proof one needs to apply Theorem 1.4 in place of Theorem 1.6.

As we see in all previous statements, the number $\lambda \in (0, \infty)$ in Condition (A λ) is essential for the convergence rate. It is, therefore, reasonable to expect that the convergence in every p -th Wasserstein distance (not necessarily with a rate) holds only under Condition (C) and the condition that the left-hand side in (19) is $o(h)$, as $h \rightarrow 0$. This is indeed true but does not follow from the previous results. We formulate this as a separate theorem, which is also proved in Section 6.

Theorem 1.9. Assume that Condition (C) holds and that

$$\sup_{y \in I_h} \left| \frac{1}{2} \int_{(y-a_h(y), y+a_h(y))} (a_h(y) - |u - y|) m(du) - h \right| \in o(h), \quad h \rightarrow 0. \quad (29)$$

Let $T \in (0, \infty)$, $y \in I^\circ$ and $p \in [1, \infty)$. Then we have (26) and, moreover,

$$\mathcal{W}_p(P \circ (X^{h,y})^{-1}, P_y \circ (Y)^{-1}) \rightarrow 0, \quad h \rightarrow 0, \quad (30)$$

or, equivalently, for all continuous path functionals $F: C([0, T], I) \rightarrow \mathbb{R}$ satisfying

$$|F(x)| \leq L(1 + \|x\|_C^p), \quad x \in C([0, T], I)$$

($L \in [0, \infty)$ is an arbitrary constant), it holds

$$E \left[F(X_t^{h,y}; t \in [0, T]) \right] \rightarrow E_y \left[F(Y_t; t \in [0, T]) \right], \quad h \rightarrow 0. \quad (31)$$

Remark 1.10 (*Applicability of the main results*). (i) [Theoretical view] The assumptions in our main results, Theorems 1.4 and 1.6, are Condition (C) and Condition (A λ). Condition (C) is an assumption on the general diffusion Y we want to approximate, and it cannot be dropped (we present examples in Section 7). Condition (A λ) is an assumption on the approximation scheme. The main results thus provide convergence rates of the Markov chain approximation schemes (10)–(11) defined by scale factors satisfying Condition (A λ), with some $\lambda \in (0, \infty)$, to general diffusions satisfying Condition (C).

Given a task to approximate some general diffusion Y we usually have a choice of an approximation scheme, and we always can choose a scheme that gives the best rates in Theorems 1.4 and 1.6 (namely, rate $\frac{1}{4}$ in Theorem 1.4 and rate $\frac{1}{4}-$ in Theorem 1.6) whenever Y satisfies Condition (C). Indeed, we can take the EMCEL approximation scheme (see Example 1.3), which is well-defined for every general diffusion Y and satisfies Condition (A λ) for all $\lambda \in (0, \infty)$.

(ii) [Towards a practical implementation] In general, Equation (22) defining the EMCEL scale factors can rarely be solved in a closed form. Therefore, in practice we often need to use scale factors $(a_h)_{h \in (0, \bar{h})}$ that are approximations of the EMCEL scale factors $(\hat{a}_h)_{h \in (0, \bar{h})}$, and for this latter case the main results tell us with which precision to solve (22) in order still to achieve the best rates. Namely, Condition (A λ) with $\lambda = \frac{1}{2}$ still guarantees rate $\frac{1}{4}$ in Theorem 1.4 and rate $\frac{1}{4}-$ in Theorem 1.6. As an informal summary, we get the following recipe. Solving (22) with a precision of order $\mathcal{O}(h^{3/2})$ provides the best rates in the main results. A further reduction of the precision in solving (22) entails smaller rates.

Also recall that our results are presented for a general diffusion Y in natural scale. If Y is not in natural scale, we should apply our results to $s(Y)$, where s is the scale function of Y . For instance, if Y is driven by an SDE with drift, s is represented by a certain integral that involves the SDE coefficients. This integral can rarely be computed in a closed form, but the procedure in this case is to write down Equation (22) for $s(Y)$ and solve it with the precision of order $\mathcal{O}(h^{3/2})$ (in other words, there is no need to approximate s in an intermediate step; but approximating the solution to (22) is necessary and can be, of course, a challenging problem).

On the other hand, our results should not be understood in the way that, in all cases, we insist on solving (22) with the precision of order $\mathcal{O}(h^{3/2})$. In “nice” situations (e.g., SDEs with “sufficiently regular” coefficients) one may fall back on simpler schemes, e.g., the Euler one, and check if its scale factors satisfy Condition (A λ) with $\lambda = \frac{1}{2}$. We suggest to solve (22) with the precision of order $\mathcal{O}(h^{3/2})$ only in “irregular” cases, which, however, can be an involved task. To deal with this problem in general, it is helpful to study properties of the EMCEL scheme, which we do in [4]. In particular, [4] contains a certain ODE for the EMCEL scale factors \hat{a}_h , $h \in (0, \bar{h})$, which holds in full generality, i.e., for *all* possible speed measures m .

Finally, it is worth mentioning that the EMCEL scheme is explicitly described in a number of SDE examples in Section 5 of [5] and for sticky Brownian motion in Section 7 of [6]. In addition, in Section 8 of [6] a Brownian motion slowed down on the Cantor set is successfully approximated in closed form by means of approximate solutions to (22) (the latter case is an example, where there is no chance to have an explicit solution to (22), but we can still write an explicit approximation scheme).

Remark 1.11 (*Reflecting boundaries*). We described the approximation scheme under the assumption that if a boundary point is accessible, then it is absorbing. The case of reflecting boundaries can be reduced to this situation. In more detail, we can include both instantaneously reflecting and slowly reflecting (sticky) boundaries as follows. If Z is a general diffusion in natural scale with reflecting boundaries, then there exists a general diffusion Y in natural scale with inaccessible and/or absorbing boundaries and a Lipschitz continuous function f with Lipschitz constant 1 such that $Z \stackrel{d}{=} f(Y)$ (see Section 6 of [6] for details; in the special case, where $[l, \infty)$ is the state space of Z and $l \in \mathbb{R}$ a reflecting boundary, we can find Y being a certain general diffusion on \mathbb{R} and take $f(y) = l + |y - l|$, $y \in \mathbb{R}$). Let X^h , $h \in (0, \bar{h})$, be the processes

of (10)–(11) approximating Y . To approximate Z we define $Z^h = f(X^h)$, $h \in (0, \bar{h})$. As f has Lipschitz constant 1, it holds for all $h \in (0, \bar{h})$ and $p \in [1, \infty)$ that

$$\mathcal{W}_p(P \circ (Z^h)^{-1}, P_y \circ (Z)^{-1}) \leq \mathcal{W}_p(P \circ (X^h)^{-1}, P_y \circ (Y)^{-1}),$$

hence, the rates of Theorems 1.4 and 1.6 are preserved also for the convergence of $(Z^h)_{h \in (0, \bar{h})}$ to Z in every p -th Wasserstein distance.

2. Properties of the approximation schemes

In this section we collect some useful properties of approximation schemes satisfying Condition (20).

Lemma 2.1. *Assume that (20) is satisfied. Then for all $y \in I^\circ$ we have*

$$\lim_{h \rightarrow 0} a_h(y) = 0. \quad (32)$$

In particular, (32) holds under Condition (A λ) with an arbitrary $\lambda \in (0, \infty)$.

Proof. We first notice that any $y \in I^\circ$ belongs to I_h whenever $h \in (0, \bar{h})$ is sufficiently small. Hence, (20) implies that for all $y \in I^\circ$

$$\lim_{h \rightarrow 0} \int_{(y-a_h(y), y+a_h(y))} (a_h(y) - |u-y|) m(du) = 0.$$

Let $y \in I^\circ$ and assume that $\limsup_{h \rightarrow 0} a_h(y) > 0$. Then there exists $\varepsilon > 0$ and a sequence $(h_n)_{n \in \mathbb{N}} \subset (0, \bar{h})$ such that $\lim_{n \rightarrow \infty} h_n = 0$ and $a_{h_n}(y) \geq \varepsilon$ for all $n \in \mathbb{N}$. Then it holds for all $n \in \mathbb{N}$ that

$$\int_{(y-a_{h_n}(y), y+a_{h_n}(y))} (a_{h_n}(y) - |u-y|) m(du) \geq \int_{(y-\varepsilon, y+\varepsilon)} (\varepsilon - |u-y|) m(du) > 0.$$

The obtained contradiction completes the proof. \square

We fix an arbitrary $y_0 \in I^\circ$. In what follows we make a convention that if (20) is satisfied, then \bar{h} is chosen small enough to guarantee

$$\sup_{h \in (0, \bar{h})} a_h(y_0) < \infty. \quad (33)$$

Indeed, due to Lemma 2.1, under (20) we can always achieve (33) by reducing \bar{h} if necessary.

Proposition 2.2. *Assume that (20) is satisfied. Then it holds for all $h \in (0, \bar{h})$ and $y_1, y_2 \in I_h$ with $y_1 < y_2$ that*

$$y_1 - y_2 - \frac{2\gamma a_h(y_2)}{1-\gamma} \leq a_h(y_2) - a_h(y_1) \leq y_2 - y_1 + \frac{2\gamma a_h(y_1)}{1-\gamma}. \quad (34)$$

Moreover, there exists $C \in [0, \infty)$ such that for all $h \in (0, \bar{h})$ and $y \in I^\circ$ we have

$$a_h(y) \leq C + |y|. \quad (35)$$

In particular, both statements hold under Condition (A λ) with an arbitrary $\lambda \in (0, \infty)$.

Proof. For the first part of the proof fix $h \in (0, \bar{h})$ and $y_1, y_2 \in I_h$ with $y_1 < y_2$. We only prove the second inequality in (34). The first one follows from similar considerations. Assume that $a_h(y_2) - a_h(y_1) \geq y_2 - y_1$ (else there is nothing to show). Note that (20) ensures that $a_h(y_1) > 0$. It holds that

$$\begin{aligned}
 & \frac{1}{2} \int_{(y_1 - a_h(y_1), y_1 + a_h(y_1))} (a_h(y_1) - |u - y_1|) m(du) \left(1 + \frac{a_h(y_2) - y_2 - a_h(y_1) + y_1}{a_h(y_1)} \right) \\
 &= \frac{1}{2} \int_{(y_1 - a_h(y_1), y_1 + a_h(y_1))} (a_h(y_1) - |u - y_1|) m(du) \\
 & \quad + \frac{a_h(y_2) - y_2 - a_h(y_1) + y_1}{2a_h(y_1)} \int_{(y_1 - a_h(y_1), y_1 + a_h(y_1))} (a_h(y_1) - |u - y_1|) m(du) \\
 &\leq \frac{1}{2} \int_{(y_1 - a_h(y_1), y_1 + a_h(y_1))} (a_h(y_1) - |u - y_1|) m(du) \\
 & \quad + \frac{1}{2} (a_h(y_2) - y_2 - a_h(y_1) + y_1) m((y_1 - a_h(y_1), y_1 + a_h(y_1))) \\
 &= \frac{1}{2} \int_{(y_1 - a_h(y_1), y_1 + a_h(y_1))} (a_h(y_2) - |u - y_1| - y_2 + y_1) m(du).
 \end{aligned} \tag{36}$$

Combining this with the fact that $\forall u \in \mathbb{R}: |u - y_2| \leq |u - y_1| + |y_1 - y_2| = |u - y_1| + y_2 - y_1$ implies that

$$\begin{aligned}
 & \frac{1}{2} \int_{(y_1 - a_h(y_1), y_1 + a_h(y_1))} (a_h(y_1) - |u - y_1|) m(du) \left(1 + \frac{a_h(y_2) - y_2 - a_h(y_1) + y_1}{a_h(y_1)} \right) \\
 & \leq \frac{1}{2} \int_{(y_1 - a_h(y_1), y_1 + a_h(y_1))} (a_h(y_2) - |u - y_2|) m(du).
 \end{aligned} \tag{37}$$

The assumption that $a_h(y_2) - a_h(y_1) \geq y_2 - y_1$ ensures that $(y_1 - a_h(y_1), y_1 + a_h(y_1)) \subseteq (y_2 - a_h(y_2), y_2 + a_h(y_2))$. Moreover it holds that $a_h(y_2) - |u - y_2| \geq 0$ for all $u \in (y_2 - a_h(y_2), y_2 + a_h(y_2))$. Combining this with (37) shows that

$$\begin{aligned}
 & \frac{1}{2} \int_{(y_1 - a_h(y_1), y_1 + a_h(y_1))} (a_h(y_1) - |u - y_1|) m(du) \left(1 + \frac{a_h(y_2) - y_2 - a_h(y_1) + y_1}{a_h(y_1)} \right) \\
 & \leq \frac{1}{2} \int_{(y_2 - a_h(y_2), y_2 + a_h(y_2))} (a_h(y_2) - |u - y_2|) m(du).
 \end{aligned} \tag{38}$$

It follows from (20) that

$$\frac{1}{2} \int_{(y_2 - a_h(y_2), y_2 + a_h(y_2))} (a_h(y_2) - |u - y_2|) m(du) \leq (1 + \gamma)h \tag{39}$$

and that

$$\frac{1}{2} \int_{(y_1 - a_h(y_1), y_1 + a_h(y_1))} (a_h(y_1) - |u - y_1|) m(du) \geq (1 - \gamma)h. \tag{40}$$

Hence, it holds that

$$(1 - \gamma)h \left(1 + \frac{a_h(y_2) - y_2 - a_h(y_1) + y_1}{a_h(y_1)} \right) \leq (1 + \gamma)h \quad (41)$$

This implies that

$$a_h(y_2) - a_h(y_1) \leq y_2 - y_1 + \frac{2\gamma a_h(y_1)}{1 - \gamma}. \quad (42)$$

It remains to prove Inequality (35). Applying (33) together with the second inequality in (34) with $y_1 = y_0$ and $y_2 > y_0$ as well as the first inequality in (34) with $y_2 = y_0$ and $y_1 < y_0$, we infer that there exists $\tilde{C} \in [0, \infty)$ such that, for all pairs (h, y) with $h \in (0, \bar{h})$, $y \in I_h$, it holds $a_h(y) \leq \tilde{C} + |y|$. Furthermore, for all pairs (h, y) with $h \in (0, \bar{h})$, $y \in I^\circ \setminus I_h$, we have $y + a_h(y) = r$ or $y - a_h(y) = l$, which means that $y \mapsto a_h(y)$ is affine with slope ± 1 and intercept r or $-l$. This completes the proof. \square

Proposition 2.2 provides the functional bound $C + |y|$ (independent of h) for all functions a_h , $h \in (0, \bar{h})$, satisfying (20). For any fixed $y \in I^\circ$, we know from Lemma 2.1 that $a_h(y) \rightarrow 0$ as $h \rightarrow 0$ (again under (20)). A natural question is, therefore, to find a functional bound for a_h that depends on h and vanishes as $h \rightarrow 0$. Under Condition (C) in addition to (20), such a bound is suggested in the next result.

Proposition 2.3. *Suppose that Condition (C) is satisfied and that (20) holds true. Then there exists a constant $C \in (0, \infty)$ such that for all $h \in (0, \bar{h})$ and $y \in I^\circ$ it holds that*

$$a_h(y) \leq C(1 + k_2|y|)\sqrt{h}. \quad (43)$$

In particular, (43) holds under Condition (C) and Condition (A λ) with an arbitrary $\lambda \in (0, \infty)$.

Proof. By (20), it holds for $h \in (0, \bar{h})$ and $y \in I_h$ that

$$\frac{1}{2} \int_{(y-a_h(y), y+a_h(y))} (a_h(y) - |u - y|) m(du) \leq (\gamma + 1)h. \quad (44)$$

It follows from (16), (17) and (18) that for $h \in (0, \bar{h})$ and $y \in I^\circ \setminus I_h$

$$\frac{1}{2} \int_{(y-a_h(y), y+a_h(y))} (a_h(y) - |u - y|) m(du) \leq h.$$

In particular, (44) holds, in fact, for all $h \in (0, \bar{h})$ and $y \in I^\circ$. Now Condition (C) implies for all $h \in (0, \bar{h})$ and $y \in I^\circ$

$$\begin{aligned} (\gamma + 1)h &\geq \int_{y-a_h(y)}^{y+a_h(y)} \frac{a_h(y) - |u - y|}{k_1(1 + k_2u^2)} du = a_h^2(y) \int_{-1}^1 \frac{1 - |z|}{k_1(1 + k_2(y + za_h(y))^2)} dz \\ &\geq \frac{a_h^2(y)}{k_1(1 + k_2 \max_{z \in [-1, 1]} (y + za_h(y))^2)} \int_{-1}^1 (1 - |z|) dz \\ &= \frac{a_h^2(y)}{k_1(1 + k_2(|y| + a_h(y))^2)}. \end{aligned} \quad (45)$$

Combining this with Proposition 2.2 ensures there exists $C_1 \in (0, \infty)$ such that for all $h \in (0, \bar{h})$ and $y \in I^\circ$ it holds that

$$a_h(y) \leq \sqrt{(\gamma + 1)k_1(1 + k_2(2|y| + C_1)^2)}\sqrt{h}. \quad (46)$$

This proves (43). \square

3. Condition (C) implies finite moments

We first establish moment bounds and a kind of temporal regularity for the general diffusion under Condition (C).

Theorem 3.1. *Assume that Condition (C) is satisfied and let $p \in [1, \infty)$ and $T \in (0, \infty)$. Then there exists $C \in (0, \infty)$ such that for all $y \in I^\circ$ and $s < t$ in $[0, T]$ it holds*

$$\left\| \sup_{u \in [s, t]} |Y_u - Y_s| \right\|_{L^p(P_y)} \leq C(1 + k_2|y|)\sqrt{t - s}. \quad (47)$$

In particular, for all $p \in [1, \infty)$ and $T \in (0, \infty)$ there exists $D \in (0, \infty)$ such that for all $y \in I^\circ$ we have

$$\left\| \sup_{u \in [0, T]} |Y_u - y| \right\|_{L^p(P_y)} \leq D(1 + k_2|y|). \quad (48)$$

Proof. Throughout the proof we assume without loss of generality that $p \in [2, \infty)$ (the statement for $p \in [1, 2)$ follows via monotonicity of L^p norms). We fix $y \in I^\circ$ and denote by C_i , $i \in \mathbb{N}$, constants that do not depend on y . We use the fact that Y can be written as a time-changed Brownian motion. To be more precise, by extending the probability space (if necessary), we can find a Brownian motion W starting in y under P_y such that

$$Y_t = W_{\gamma(t)}, \quad \text{for all } t \in [0, \infty), \quad (49)$$

with the random time change $\gamma: [0, \infty) \rightarrow [0, \infty)$ being the right inverse of A , i.e.,

$$\gamma(t) = \inf\{s \in [0, \infty): A(s) > t\}, \quad t \in [0, \infty), \quad (50)$$

where the increasing process $A: [0, \infty) \rightarrow [0, \infty]$ is given by the formula

$$A(t) = \frac{1}{2} \int_I L_t^x m(dx), \quad t \in [0, \infty), \quad (51)$$

and (L_t^x) , $t \in [0, \infty)$, $x \in \mathbb{R}$, denotes the local time of W (see Theorem V.47.1 in [37] and notice that our speed measure is twice as large as the speed measure in [37]). Furthermore, A is strictly increasing and continuous on $[0, H_{l,r}(W)] \cap [0, \infty)$ and infinite on $(H_{l,r}(W), \infty)$; γ is continuous on $[0, \infty)$, strictly increasing on $[0, H_{l,r}(Y)]$ (notice that $H_{l,r}(Y) = A(H_{l,r}(W))$) and, on the event $\{H_{l,r}(Y) < \infty\}$, γ is constant on $[H_{l,r}(Y), \infty)$ (recall that accessible boundaries are assumed to be absorbing). It follows from Condition (C) that for all $s < t$ in $[0, H_{l,r}(W)]$ it holds

$$A(t) - A(s) = \frac{1}{2} \int_I (L_t^x - L_s^x) m(dx) \geq \int_{\mathbb{R}} 1_I(x) \frac{(L_t^x - L_s^x)}{k_1(1 + k_2x^2)} dx = \int_s^t \frac{1}{k_1(1 + k_2W_u^2)} du, \quad (52)$$

where the last equality follows from the occupation time formula. Let $f \in C^1([0, \infty), [0, \infty))$ be the function given by $f(t) = \int_0^t k_1(1 + k_2 W_{\gamma(s)}^2) ds$, $t \in [0, \infty)$. Then it follows with the change of variables in Stieltjes integrals for all $t \in [0, H_{l,r}(W)]$ that

$$\int_0^{A(t)} k_1(1 + k_2 W_{\gamma(s)}^2) ds = f(A(t)) = \int_0^t f'(A(s)) dA(s) = \int_0^t k_1(1 + k_2 W_s^2) dA(s).$$

This and (52) imply for all $s < t$ in $[0, H_{l,r}(Y)]$ that

$$\gamma(t) - \gamma(s) \leq \int_{\gamma(s)}^{\gamma(t)} k_1(1 + k_2 W_u^2) dA(u) = \int_s^t k_1(1 + k_2 W_{\gamma(u)}^2) du = \int_s^t k_1(1 + k_2 Y_u^2) du$$

(notice that $A(\gamma(t)) = t$ because $t \leq H_{l,r}(Y)$). As γ is constant on $[H_{l,r}(Y), \infty)$, the latter estimate

$$\gamma(t) - \gamma(s) \leq \int_s^t k_1(1 + k_2 Y_u^2) du \quad (53)$$

holds, in fact, for all $s < t$ in $[0, \infty)$.

Next, let $\hat{\tau}$ be an $(\mathcal{F}_t)_{t \in [0, \infty)}$ -stopping time. Then it holds for all $t \in [0, \infty)$ that

$$E_y[\sup_{s \in [0, t]} |Y_s^{\hat{\tau}} - y|^p] = E_y[\sup_{s \in [0, t \wedge \hat{\tau}]} |Y_s - y|^p] = E_y[\sup_{s \in [0, \gamma(t \wedge \hat{\tau})]} |W_s - y|^p]. \quad (54)$$

This and the Burkholder-Davis-Gundy inequality show that there exists $C_1 \in (0, \infty)$ (independent of $\hat{\tau}$ and y) such that for all $t \in [0, \infty)$ it holds

$$E_y[\sup_{s \in [0, t]} |Y_s^{\hat{\tau}} - y|^p] = E_y[\sup_{s \in [0, \gamma(t \wedge \hat{\tau})]} |W_s - y|^p] \leq C_1 E_y[\gamma(t \wedge \hat{\tau})^{\frac{p}{2}}]. \quad (55)$$

Then it follows with (53) for all $t \in [0, \infty)$ that

$$\begin{aligned} E_y[\sup_{s \in [0, t]} |Y_s^{\hat{\tau}} - y|^p] &\leq C_1 E_y \left[\left(\int_0^{t \wedge \hat{\tau}} k_1(1 + k_2 Y_s^2) ds \right)^{\frac{p}{2}} \right] \\ &\leq C_1 k_1^{\frac{p}{2}} E_y \left[\left(t + 2k_2 y^2 t + 2k_2 \int_0^t (Y_s^{\hat{\tau}} - y)^2 ds \right)^{\frac{p}{2}} \right] \\ &\leq 2^{\frac{p}{2}-1} C_1 k_1^{\frac{p}{2}} \left((t + 2k_2 y^2 t)^{\frac{p}{2}} + (2k_2)^{\frac{p}{2}} E_y \left[\left(\int_0^t (Y_s^{\hat{\tau}} - y)^2 ds \right)^{\frac{p}{2}} \right] \right). \end{aligned} \quad (56)$$

Combining this with Jensen's inequality proves for all $t \in [0, T]$ that

$$\begin{aligned}
E_y \left[\sup_{u \in [0, t]} |Y_u^{\hat{\tau}} - y|^p \right] &\leq 2^{\frac{p}{2}-1} C_1 k_1^{\frac{p}{2}} \left((t + 2k_2 y^2 t)^{\frac{p}{2}} + (2k_2)^{\frac{p}{2}} t^{\frac{p}{2}-1} \int_0^t E_y \left[|Y_s^{\hat{\tau}} - y|^p \right] ds \right) \\
&\leq 2^{\frac{p}{2}-1} C_1 k_1^{\frac{p}{2}} \left((T + 2k_2 y^2 T)^{\frac{p}{2}} + (2k_2)^{\frac{p}{2}} T^{\frac{p}{2}-1} \int_0^t E_y \left[\sup_{u \in [0, s]} |Y_u^{\hat{\tau}} - y|^p ds \right] \right).
\end{aligned} \tag{57}$$

We choose $\hat{\tau} = \hat{\tau}_n = \inf\{s \in [0, \infty) : |Y_s - y| \geq n\}$, $n \in \mathbb{N}$. Then it holds for all $n \in \mathbb{N}$, $t \in [0, T]$ that $\int_0^t E_y[\sup_{u \in [0, s]} |Y_u^{\hat{\tau}_n} - y|^p] ds \leq n^p T < \infty$. Applying Gronwall's inequality to (57) shows for all $n \in \mathbb{N}$ and $t \in [0, T]$ that

$$E_y \left[\sup_{s \in [0, t]} |Y_s^{\hat{\tau}_n} - y|^p \right] \leq C_1 k_1^{\frac{p}{2}} 2^{\frac{p}{2}-1} (T + 2k_2 y^2 T)^{\frac{p}{2}} e^{C_1 k_1^{\frac{p}{2}} 2^{\frac{p}{2}-1} (2k_2)^{\frac{p}{2}} T^{\frac{p}{2}-1} t}. \tag{58}$$

Since $\hat{\tau}_n \nearrow \infty$ as $n \rightarrow \infty$, it follows with monotone convergence that

$$E_y \left[\sup_{s \in [0, T]} |Y_s - y|^p \right] = \lim_{n \rightarrow \infty} E_y \left[\sup_{s \in [0, T]} |Y_s^{\hat{\tau}_n} - y|^p \right] \leq C_1 k_1^{\frac{p}{2}} 2^{\frac{p}{2}-1} (T + 2k_2 y^2 T)^{\frac{p}{2}} e^{C_1 k_1^{\frac{p}{2}} 2^{\frac{p}{2}-1} (2k_2)^{\frac{p}{2}} T^{\frac{p}{2}}}. \tag{59}$$

To summarize, this proves that there exists $C_2 \in (0, \infty)$ (independent of $y \in I^\circ$) such that it holds

$$\left\| \sup_{s \in [0, T]} |Y_s - y| \right\|_{L^p(P_y)} \leq C_2 (1 + k_2 |y|). \tag{60}$$

Now note that for all $s < t$ in $[0, \infty)$ it holds that

$$E_y \left[\sup_{u \in [s, t]} |Y_u - Y_s|^p \right] = E_y \left[\sup_{u \in [\gamma(s), \gamma(t)]} |W_u - W_{\gamma(s)}|^p \right]. \tag{61}$$

This and the Burkholder-Davis-Gundy inequality show that there exists $C_3 \in (0, \infty)$ such that for all $s < t$ in $[0, \infty)$ it holds

$$E_y \left[\sup_{u \in [s, t]} |Y_u - Y_s|^p \right] \leq C_3 E_y \left[|\gamma(t) - \gamma(s)|^{\frac{p}{2}} \right]. \tag{62}$$

Then it follows with (53) for all $s < t$ in $[0, \infty)$ that

$$\begin{aligned}
E_y \left[\sup_{u \in [s, t]} |Y_u - Y_s|^p \right] &\leq C_3 E_y \left[\left(\int_s^t k_1 (1 + k_2 Y_u^2) du \right)^{\frac{p}{2}} \right] \\
&\leq C_3 k_1^{\frac{p}{2}} E_y \left[\left(1 + 2k_2 y^2 + 2k_2 \sup_{u \in [s, t]} |Y_u - y|^2 \right)^{\frac{p}{2}} \right] (t - s)^{\frac{p}{2}}.
\end{aligned} \tag{63}$$

Together with (60), this proves (47). Finally, (48) immediately follows from (47). \square

The next result establishes uniform in h moment bounds for the approximations.

Theorem 3.2. Suppose that Condition (C) is satisfied and that (20) holds true. Let $p \in [1, \infty)$ and $T \in (0, \infty)$. Then there exists $C \in (0, \infty)$ such that for all $y \in I^\circ$ it holds that

$$\sup_{h \in (0, \bar{h})} \left\| \sup_{s \in [0, T]} |X_s^{h,y} - y| \right\|_{L^p(P)} \leq \sup_{h \in (0, \bar{h})} \left\| \max_{k \in \{0, \dots, \lceil T/h \rceil\}} |X_{kh}^{h,y} - y| \right\|_{L^p(P)} \leq C(1 + k_2|y|). \quad (64)$$

In particular, (64) holds under Condition (C) and Condition (A λ) with an arbitrary $\lambda \in (0, \infty)$.

Proof. Throughout the proof we assume without loss of generality that $p \in [2, \infty)$. We fix $y \in I^\circ$ and denote by C_i , $i \in \mathbb{N}$, constants that do not depend on y . For all $h \in (0, \bar{h})$ let $M_n^{h,y} = \max_{k \in \{0, \dots, n\}} |X_{kh}^{h,y} - y|$, $n \in \mathbb{N}_0$. The first inequality in (64) holds by the construction of the continuous-time process $X^{h,y}$. The fact that for every $h \in (0, \bar{h})$ the process $(X_{kh}^{h,y})_{k \in \mathbb{N}_0}$ is a martingale, the Burkholder-Davis-Gundy inequality and the definition of $(X_{kh}^{h,y})_{k \in \mathbb{N}_0}$ ensure that there exists $C_1 \in (0, \infty)$ (independent of y) such that for all $h \in (0, \bar{h})$ and $n \in \mathbb{N}$ it holds that

$$E[|M_n^{h,y}|^p] \leq C_1 E \left[\left(\sum_{k=0}^{n-1} (X_{(k+1)h}^{h,y} - X_{kh}^{h,y})^2 \right)^{\frac{p}{2}} \right] = C_1 E \left[\left(\sum_{k=0}^{n-1} (a_h(X_{kh}^{h,y}))^2 \right)^{\frac{p}{2}} \right]. \quad (65)$$

This together with Proposition 2.3 shows that there exists a constant $C_2 \in (0, \infty)$ such that for all $h \in (0, \bar{h})$ and $n \in \mathbb{N}$ it holds

$$\begin{aligned} E[|M_n^{h,y}|^p] &\leq C_2 E \left[\left(\sum_{k=0}^{n-1} h(1 + k_2(X_{kh}^{h,y})^2) \right)^{\frac{p}{2}} \right] \\ &= C_2 E \left[\left(hn + k_2h \sum_{k=0}^{n-1} (X_{kh}^{h,y})^2 \right)^{\frac{p}{2}} \right] \\ &\leq C_2 E \left[\left((1 + 2k_2y^2)hn + 2k_2h \sum_{k=0}^{n-1} (X_{kh}^{h,y} - y)^2 \right)^{\frac{p}{2}} \right] \\ &\leq C_2 2^{\frac{p}{2}-1} \left[((1 + 2k_2y^2)hn)^{\frac{p}{2}} + (2k_2h)^{\frac{p}{2}} E \left[\left(\sum_{k=0}^{n-1} (X_{kh}^{h,y} - y)^2 \right)^{\frac{p}{2}} \right] \right]. \end{aligned} \quad (66)$$

Combining this with Jensen's inequality proves for all $h \in (0, \bar{h})$, $n \in \{1, \dots, \lceil T/h \rceil\}$ that

$$\begin{aligned} E[|M_n^{h,y}|^p] &\leq C_2 2^{\frac{p}{2}-1} \left[((1 + 2k_2y^2)(T+1))^{\frac{p}{2}} + (2k_2h)^{\frac{p}{2}} n^{\frac{p}{2}-1} E \left[\sum_{k=0}^{n-1} |X_{kh}^{h,y} - y|^p \right] \right] \\ &\leq C_2 2^{\frac{p}{2}-1} \left[((1 + 2k_2y^2)(T+1))^{\frac{p}{2}} + (2k_2)^{\frac{p}{2}} (T+1)^{\frac{p}{2}-1} h \sum_{k=0}^{n-1} E[|M_k^{h,y}|^p] \right]. \end{aligned} \quad (67)$$

It follows from Gronwall's inequality that for all $h \in (0, \bar{h})$, $n \in \{1, \dots, \lceil T/h \rceil\}$ it holds

$$E[|M_n^{h,y}|^p] \leq C_2 2^{\frac{p}{2}-1} ((1 + 2k_2y^2)(T+1))^{\frac{p}{2}} \left(1 + C_2 2^{\frac{p}{2}-1} (2k_2)^{\frac{p}{2}} (T+1)^{\frac{p}{2}-1} h \right)^{n-1}. \quad (68)$$

This implies that

$$\sup_{h \in (0, \bar{h})} E \left[\max_{k \in \{0, \dots, \lceil T/h \rceil\}} |X_{kh}^{h,y} - y|^p \right] \leq C_2 2^{\frac{p}{2}-1} ((1 + 2k_2 y^2)(T + 1))^{\frac{p}{2}} \left[\sup_{h \in (0, \bar{h})} \left(1 + C_2 2^{\frac{p}{2}-1} (2k_2)^{\frac{p}{2}} (T + 1)^{\frac{p}{2}-1} h \right)^{\frac{T}{h}} \right] < \infty, \quad (69)$$

which completes the proof. \square

4. Embedding the approximations into the general diffusion

The proofs of Theorems 1.4 and 1.6 rely on embeddings of the Markov chains $(X_{kh}^{h,y})_{k \in \mathbb{N}_0}$ into the diffusion Y with sequences of stopping times. In Section 3 in [6] (see, in particular, Proposition 3.1 therein) we construct for all $h \in (0, \bar{h})$ a sequence of stopping times $(\tau_k^h)_{k \in \mathbb{N}_0}$ such that for all $h \in (0, \bar{h})$ and $y \in I^\circ$ we have

$$\text{Law}_{P_y} \left(Y_{\tau_k^h}; k \in \mathbb{N}_0 \right) = \text{Law}_P \left(X_{kh}^{h,y}; k \in \mathbb{N}_0 \right). \quad (70)$$

In the case, where both boundary points are inaccessible, the construction of the stopping times is straightforward: for all $h \in (0, \bar{h})$ set $\tau_0^h = 0$ and then recursively define τ_{k+1}^h as the first time $t \geq \tau_k^h$ such that $|Y_t - Y_{\tau_k^h}| = a_h(Y_{\tau_k^h})$. In the case of accessible (then absorbing) boundary points the stopping times are deterministically extended on reaching an absorbing boundary whenever $Y_{\tau_k^h} \notin I_h$. For more details, see Section 3 in [6].

Remark 1.2 in [6] shows that for every $y \in I^\circ$ and $a \in [0, \infty)$ such that $y \pm a \in I$ the expected time it takes Y started in y to leave the interval $(y - a, y + a)$ is equal to $\frac{1}{2} \int_{(y-a, y+a)} (a - |u - y|) m(du)$. Therefore, Condition (A λ) ensures, roughly speaking, for all $k \in \mathbb{N}_0$ that the expected time lag $\tau_{k+1}^h - \tau_k^h$ deviates from h by an order of $h^{1+\lambda}$. The next result concludes that the global temporal error $|\tau_k^h - kh|$ is of order $h^{\min\{\frac{1}{2}, \lambda\}}$ under Condition (A λ). It is worth noting that, in contrast to many preceding results, we do not need Condition (C) here.

Proposition 4.1. *Let $p \in [1, \infty)$, $T \in (0, \infty)$, $\lambda \in (0, \infty)$ and assume that Condition (A λ) is satisfied. Then there exists $C \in (0, \infty)$ such that for all $y \in I^\circ$ and $h \in (0, \bar{h})$ it holds that*

$$\left\| \sup_{k \in \{1, \dots, \lceil T/h \rceil\}} |\tau_k^h - kh| \right\|_{L^p(P_y)} \leq C h^{\min\{\frac{1}{2}, \lambda\}}. \quad (71)$$

Remark 4.2. Note that the order of convergence of the right-hand side of (71) does not become better than \sqrt{h} for $\lambda \in [\frac{1}{2}, \infty)$. We refer the reader to Equation (98) in the proof of Proposition 5.3 below to see that \sqrt{h} is the optimal order that, in general, cannot be improved even if $\frac{1}{2} \int_{(y-a_h(y), y+a_h(y))} (a_h(y) - |u - y|) m(du) = h$ for all $y \in I^\circ$, $h \in (0, \bar{h})$, i.e., Condition (A λ) is satisfied for all $\lambda \in (0, \infty)$.

Proof. We introduce the time lags between two consecutive stopping times $\rho_k^h = \tau_k^h - \tau_{k-1}^h$, $h \in (0, \bar{h})$, $k \in \mathbb{N}$. It follows from Proposition 5.2 and Remark 1.2 in [6] that there exists a constant $C_1 \in (0, \infty)$ such that for all $y \in I^\circ$ and $h \in (0, \bar{h})$ it holds that

$$\begin{aligned}
& \left\| \sup_{k \in \{1, \dots, \lfloor T/h \rfloor\}} \left| \tau_k^h - \sum_{n=1}^k E[\rho_n^h | \mathcal{F}_{\tau_{n-1}^h}] \right| \right\|_{L^p(P_y)} \\
& \leq C_1 \sqrt{T} \left(\sqrt{h} + \frac{\sup_{z \in I} (E_z[H_{z-a_h(z), z+a_h(z)}(Y)])}{\sqrt{h}} \right) \\
& = C_1 \sqrt{T} \left(\sqrt{h} + \frac{\frac{1}{2} \sup_{z \in I} \left(\int_{(z-a_h(z), z+a_h(z))} (a_h(z) - |u-z|) m(du) \right)}{\sqrt{h}} \right).
\end{aligned} \tag{72}$$

It follows from the definition of the sets I_h , $h \in (0, \bar{h})$, that for all $h \in (0, \bar{h})$ it holds

$$\frac{1}{2} \sup_{z \in I \setminus I_h} \int_{(z-a_h(z), z+a_h(z))} (a_h(z) - |u-z|) m(du) \leq h.$$

Condition (A λ) implies for all $h \in (0, \bar{h})$

$$\frac{1}{2} \sup_{z \in I_h} \int_{(z-a_h(z), z+a_h(z))} (a_h(z) - |u-z|) m(du) \leq Kh^{1+\lambda} + h \leq (K+1)h.$$

Hence, we obtain that for all $y \in I^\circ$ and $h \in (0, \bar{h})$

$$\left\| \sup_{k \in \{1, \dots, \lfloor T/h \rfloor\}} \left| \tau_k^h - \sum_{n=1}^k E[\rho_n^h | \mathcal{F}_{\tau_{n-1}^h}] \right| \right\|_{L^p(P_y)} \leq C_1 \sqrt{T} (K+2) \sqrt{h}. \tag{73}$$

Moreover, Proposition 5.3, Remark 1.2 in [6] and Condition (A λ) imply that for all $y \in I^\circ$ and $h \in (0, \bar{h})$ it holds that

$$\begin{aligned}
& \left\| \sup_{k \in \{1, \dots, \lfloor T/h \rfloor\}} \left| \left(\sum_{n=1}^k E[\rho_n^h | \mathcal{F}_{\tau_{n-1}^h}] \right) - kh \right| \right\|_{L^p(P_y)} \leq \frac{T}{h} \sup_{z \in I_h} |E_z[H_{z-a_h(z), z+a_h(z)}(Y)] - h| \\
& = \frac{T}{h} \sup_{z \in I_h} \left| \frac{1}{2} \int_{(z-a_h(z), z+a_h(z))} (a_h(z) - |u-z|) m(du) - h \right| \leq TKh^\lambda.
\end{aligned} \tag{74}$$

Combining (73) and (74) proves that for all $y \in I^\circ$ and $h \in (0, \bar{h})$ it holds that

$$\begin{aligned}
& \left\| \sup_{k \in \{1, \dots, \lfloor T/h \rfloor\}} |\tau_k^h - kh| \right\|_{L^p(P_y)} \leq TKh^\lambda + C_1 \sqrt{T} (K+2) \sqrt{h} \\
& \leq \left(TK + C_1 \sqrt{T} (K+2) \right) h^{\min\{\frac{1}{2}, \lambda\}}.
\end{aligned}$$

This concludes the proof. \square

In addition to the moment estimates for general diffusions presented in Section 3 we need a moment estimate up to the stopping times $\tau_{\lfloor T/h \rfloor}^h$ for $T \in (0, \infty)$.

Proposition 4.3. Assume that Condition (C) holds and that (20) is satisfied. Let $p \in [1, \infty)$ and $T \in (0, \infty)$. Then there exists $C \in (0, \infty)$ such that for all $y \in I^\circ$ it holds

$$\sup_{h \in (0, \bar{h})} \left\| \sup_{s \in [0, \tau_{\lfloor T/h \rfloor}^h]} |Y_s| \right\|_{L^p(P_y)} \leq C(1 + |y|). \quad (75)$$

In particular, (75) holds under Condition (C) and Condition (A λ) with an arbitrary $\lambda \in (0, \infty)$.

Proof. First note that by construction of the stopping times τ_k^h , $h \in (0, \bar{h})$, $k \in \mathbb{N}_0$, it holds P_y -almost surely for all $y \in I^\circ$ that for all $h \in (0, \bar{h})$ and $k \in \mathbb{N}_0$

$$\sup_{s \in [\tau_k^h, \tau_{k+1}^h]} |Y_s| \leq \max\{|Y_{\tau_k^h} + a_h(Y_{\tau_k^h})|, |Y_{\tau_k^h} - a_h(Y_{\tau_k^h})|\} = |Y_{\tau_k^h}| + a_h(Y_{\tau_k^h}). \quad (76)$$

This together with Proposition 2.2 ensures that there exists $C_1 \in (0, \infty)$ (independent of $y \in I^\circ$) such that for all $h \in (0, \bar{h})$ and $k \in \mathbb{N}_0$ it holds

$$\sup_{s \in [\tau_k^h, \tau_{k+1}^h]} |Y_s| \leq C_1 + 2|Y_{\tau_k^h}|. \quad (77)$$

Combining this with Theorem 3.2 and (70) shows that there exists $C_2 \in (0, \infty)$ such that for all $y \in I^\circ$ it holds

$$\begin{aligned} \sup_{h \in (0, \bar{h})} \left\| \sup_{s \in [0, \tau_{\lfloor T/h \rfloor}^h]} |Y_s| \right\|_{L^p(P_y)} &= \sup_{h \in (0, \bar{h})} \left\| \max_{k \in \{0, \dots, \lfloor T/h \rfloor - 1\}} \sup_{s \in [\tau_k^h, \tau_{k+1}^h]} |Y_s| \right\|_{L^p(P_y)} \\ &\leq C_1 + 2 \sup_{h \in (0, \bar{h})} \left\| \max_{k \in \{0, \dots, \lfloor T/h \rfloor - 1\}} |Y_{\tau_k^h}| \right\|_{L^p(P_y)} \\ &\leq C_1 + 2(C_2(1 + k_2|y|) + |y|), \end{aligned} \quad (78)$$

which completes the proof. \square

5. Rate of convergence for time marginals

In this section, we prove Theorem 1.4. Throughout the section let $(\tau_k^h)_{k \in \mathbb{N}_0}$, $h \in (0, \bar{h})$, be the embedding stopping times introduced in Section 4. A central step in our approach is to establish L^p bounds between the random variables Y_T and $Y_{\tau_{\lfloor T/h \rfloor}^h}$. These are provided in the next result.

Theorem 5.1. Suppose that Condition (C) is satisfied and that there exists $\lambda \in (0, \infty)$ such that Condition (A λ) holds. Let $T \in (0, \infty)$ and $p \in [1, \infty)$. Then there exists a constant $C(p, T) \in (0, \infty)$ such that for all $y \in I^\circ$ and $h \in (0, \bar{h})$ we have

$$\left\| Y_{\tau_{\lfloor T/h \rfloor}^h} - Y_T \right\|_{L^p(P_y)} \leq C(p, T)(1 + k_2|y|)h^{\min\{\frac{1}{4}, \frac{\lambda}{2}\}}. \quad (79)$$

Proof. Throughout the proof we use the notation from the proof of Theorem 3.1, in particular, the fact that Y can be written as a time-changed Brownian motion (recall (49)). We fix $y \in I^\circ$ and denote by $C_i \in (0, \infty)$, $i \in \mathbb{N}$, constants that are independent of y . First note that for all $h \in (0, \bar{h})$ it holds that

$$E_y[|Y_{\tau_{\lfloor T/h \rfloor}^h} - Y_T|^p] = E_y[|W_{\gamma(\tau_{\lfloor T/h \rfloor}^h)} - W_{\gamma(T)}|^p]. \quad (80)$$

This and the Burkholder-Davis-Gundy inequality show that there exists $C_1 \in (0, \infty)$ such that for all $h \in (0, \bar{h})$ it holds

$$E_y[|Y_{\tau_{\lfloor T/h \rfloor}^h} - Y_T|^p] \leq C_1 E_y[|\gamma(\tau_{\lfloor T/h \rfloor}^h) - \gamma(T)|^{\frac{p}{2}}]. \quad (81)$$

Then it follows with (53) and the Cauchy-Schwarz inequality that there exists $C_2 \in (0, \infty)$ such that for all $h \in (0, \bar{h})$ we have

$$\begin{aligned} E_y[|Y_{\tau_{\lfloor T/h \rfloor}^h} - Y_T|^p] &\leq C_2 E_y \left[\left(\int_{\tau_{\lfloor T/h \rfloor}^h \wedge T}^{\tau_{\lfloor T/h \rfloor}^h \vee T} (1 + k_2 Y_u^2) du \right)^{\frac{p}{2}} \right] \\ &\leq C_2 E_y \left[(1 + k_2 \sup_{u \in [0, \tau_{\lfloor T/h \rfloor}^h \vee T]} Y_u^2)^{\frac{p}{2}} |\tau_{\lfloor T/h \rfloor}^h - T|^{\frac{p}{2}} \right] \\ &\leq C_2 \sqrt{E_y \left[\left(1 + k_2 \sup_{u \in [0, \tau_{\lfloor T/h \rfloor}^h \vee T]} Y_u^2 \right)^p \right] E_y \left[|\tau_{\lfloor T/h \rfloor}^h - T|^p \right]}. \end{aligned} \quad (82)$$

The triangle inequality ensures for all $h \in (0, \bar{h})$ that

$$\begin{aligned} &\|Y_{\tau_{\lfloor T/h \rfloor}^h} - Y_T\|_{L^p(P_y)} \\ &\leq C_2^{\frac{1}{p}} \sqrt{\left(1 + k_2 \left\| \sup_{u \in [0, \tau_{\lfloor T/h \rfloor}^h \vee T]} Y_u^2 \right\|_{L^p(P_y)} \right) \left(\left\| \tau_{\lfloor T/h \rfloor}^h - h \left\lfloor \frac{T}{h} \right\rfloor \right\|_{L^p(P_y)} + T - h \left\lfloor \frac{T}{h} \right\rfloor \right)} \\ &\leq C_2^{\frac{1}{p}} \sqrt{\left(1 + k_2 \left\| \sup_{u \in [0, \tau_{\lfloor T/h \rfloor}^h \vee T]} |Y_u| \right\|_{L^{2p}(P_y)}^2 \right) \left(\left\| \tau_{\lfloor T/h \rfloor}^h - h \left\lfloor \frac{T}{h} \right\rfloor \right\|_{L^p(P_y)} + h \right)}. \end{aligned} \quad (83)$$

Theorems 3.1 and 4.3 ensure that there exists $C_3 \in (0, \infty)$ such that

$$\sup_{h \in (0, \bar{h})} \left\| \sup_{u \in [0, \tau_{\lfloor T/h \rfloor}^h \vee T]} |Y_u| \right\|_{L^{2p}(P_y)} \leq C_3(1 + |y|). \quad (84)$$

Moreover, Proposition 4.1 implies that there exists $C_4 \in (0, \infty)$ such that for all $h \in (0, \bar{h})$ it holds that

$$\left\| \tau_{\lfloor T/h \rfloor}^h - h \left\lfloor \frac{T}{h} \right\rfloor \right\|_{L^p(P_y)} \leq C_4 h^{\min\{\frac{1}{2}, \lambda\}}. \quad (85)$$

Combining (83), (84) and (85) entails that there exists $C_5 \in (0, \infty)$ such that

$$\|Y_{\tau_{\lfloor T/h \rfloor}^h} - Y_T\|_{L^p(P_y)} \leq C_5(1 + k_2|y|)h^{\min\{\frac{1}{4}, \frac{\lambda}{2}\}}. \quad (86)$$

This completes the proof. \square

Proof of Theorem 1.4. Throughout the proof let $p \in [1, \infty)$ and $T \in (0, \infty)$. By Theorems 3.1 and 3.2, the random variables Y_T , $X_{h\lfloor T/h \rfloor}^{h,y}$ and $X_T^{h,y}$, $h \in (0, \bar{h})$, admit finite p -th moments. For every $h \in (0, \bar{h})$ we define the random variable

$$Y_T^h = Y_{\tau_{\lfloor T/h \rfloor}^h} + (T/h - \lfloor T/h \rfloor)(Y_{\tau_{\lfloor T/h \rfloor+1}^h} - Y_{\tau_{\lfloor T/h \rfloor}^h}) \quad (87)$$

and notice that, by (70) and (11), for every $h \in (0, \bar{h})$ and $y \in I^\circ$ it holds

$$\text{Law}_{P_y}(Y_{\tau_{\lfloor T/h \rfloor}^h}) = \text{Law}_P(X_{h\lfloor T/h \rfloor}^{h,y}) \quad \text{and} \quad \text{Law}_{P_y}(Y_T^h) = \text{Law}_P(X_T^{h,y}).$$

In particular, the random vector $(Y_{\tau_{\lfloor T/h \rfloor}^h}, Y_T)$ (resp. (Y_T^h, Y_T)) is a coupling between the laws of $X_{h\lfloor T/h \rfloor}^{h,y}$ and Y_T (resp. $X_T^{h,y}$ and Y_T). Therefore, Statement (24) of Theorem 1.4 follows directly from Theorem 5.1. Next, applying (87), Proposition 2.3, Theorem 3.2 and (70), we obtain that there exist constants $C_1 \in (0, \infty)$ and $C_2 \in (0, \infty)$ such that for all $y \in I^\circ$ and $h \in (0, \bar{h})$ it holds

$$\begin{aligned} \|Y_T^h - Y_{\tau_{\lfloor T/h \rfloor}^h}\|_{L^p(P_y)} &\leq \|Y_{\tau_{\lfloor T/h \rfloor+1}^h} - Y_{\tau_{\lfloor T/h \rfloor}^h}\|_{L^p(P_y)} = \|a_h(Y_{\tau_{\lfloor T/h \rfloor}^h})\|_{L^p(P_y)} \\ &\leq C_1 \sqrt{h} \|1 + k_2 |Y_{\tau_{\lfloor T/h \rfloor}^h}|\|_{L^p(P_y)} \leq C_2 (1 + k_2 |y|) \sqrt{h}. \end{aligned}$$

Together with Theorem 5.1, this establishes (25) and concludes the proof. \square

5.1. Optimality of the rate $1/4$

In this subsection we consider the EMCEL scheme $a_h = \hat{a}_h$ (see Example 1.3) in the situation, where both boundaries are inaccessible and hence $I_h = I^\circ$ for all $h \in (0, \bar{h})$. Recall that in this case scale factors a_h are chosen in such a way that for all $y \in I^\circ$, $h \in (0, \bar{h})$ it holds

$$\frac{1}{2} \int_{(y-a_h(y), y+a_h(y))} (a_h(y) - |u-y|) m(du) = h. \quad (88)$$

In particular, Condition (A λ) is satisfied for all $\lambda \in (0, \infty)$. In this situation Theorem 5.1 ensures that $Y_{\tau_{\lfloor T/h \rfloor}^h}$ converges to Y_T in L^p at a rate of at least $1/4$. We close this section by showing that the convergence rate $1/4$ is the best rate that can be obtained for this L^p convergence. Indeed, for such a nice Markov process as Brownian motion, the approximation scheme does not converge faster. To prove this, we need the following auxiliary result.

Lemma 5.2. *Let W be a Brownian motion starting in 0, τ a square integrable stopping time. Then it holds that*

$$E[W_\tau^4] \geq \frac{1}{4} E[\tau^2]. \quad (89)$$

Proof. First let τ be bounded. Then it follows with Itô's formula that

$$E[\tau W_\tau^2] = E \left[\int_0^\tau (2s W_s) dW_s + \int_0^\tau (W_s^2 + s) ds \right] \geq E \left[\int_0^\tau s ds \right] = \frac{1}{2} E[\tau^2]. \quad (90)$$

This together with Young's inequality implies that

$$\frac{1}{2} E[\tau^2] \leq E[\tau W_\tau^2] \leq E \left[\frac{1}{4} \tau^2 + W_\tau^4 \right]. \quad (91)$$

This proves (89) for bounded stopping times. Next, let τ be square integrable. Then it holds for all $n \in \mathbb{N}$ that

$$E[W_{\tau \wedge n}^4] \geq \frac{1}{4} E[(\tau \wedge n)^2]. \quad (92)$$

By monotone convergence the right-hand side converges to $\frac{1}{4} E[\tau^2]$ as $n \rightarrow \infty$. The Burkholder-Davis-Gundy inequality ensures that $\sup_{s \in [0, \tau]} W_s^4$ is integrable and hence dominated convergence ensures that

$$\lim_{n \rightarrow \infty} E[W_{\tau \wedge n}^4] = E[W_\tau^4], \quad (93)$$

which completes the proof. \square

We are now in a position to prove that our scheme applied to W does not converge faster in L^p , $p \geq 4$, than at the rate $1/4$.

Proposition 5.3. *Assume that $I = \mathbb{R}$ and that the speed measure m satisfies $m(dx) = 2dx$. In other words, the general diffusion we are considering is a Brownian motion W on \mathbb{R} . Let scale factors satisfy $a_h(y) = \sqrt{h}$ for all $h \in (0, 1)$, $y \in \mathbb{R}$. Then (88) is satisfied and for all $p \in [4, \infty)$, $y \in \mathbb{R}$, $T \in (0, \infty)$ and $h \in (0, 1 \wedge T)$ we have*

$$\|W_{\tau_{[T/h]}^h} - W_T\|_{L^p(P_y)} \geq \|W_{\tau_{[T/h]}^h} - W_T\|_{L^4(P_y)} \geq \left(\frac{(T-h) \text{Var}(H_{-1,1}(W))}{4} \right)^{\frac{1}{4}} h^{\frac{1}{4}}. \quad (94)$$

Proof. The claim that (88) is satisfied follows directly from the assumption that $a_h(y) = \sqrt{h}$ for all $h \in (0, 1)$, $y \in \mathbb{R}$. To prove (94) we can without loss of generality assume that $y = 0$ since m is space invariant. For the remainder of the proof we fix $T \in (0, \infty)$ and $h \in (0, 1 \wedge T)$. By the construction in the beginning of Section 4 the stopping times satisfy for all $k \in \mathbb{N}_0$ that $\tau_0^h = 0$ and $\tau_{k+1}^h = \inf\{t \geq \tau_k^h : |W_t - W_{\tau_k^h}| = \sqrt{h}\}$. It follows, e.g., from Proposition 4.1, that the stopping time $\tau_{[T/h]}^h$ is square integrable. Then (a conditional version of) Lemma 5.2 and the fact that $E[\tau_{[T/h]}^h] = h[T/h]$ show that

$$\begin{aligned} E[|W_{\tau_{[T/h]}^h} - W_T|^4] &= E[|W_{\tau_{[T/h]}^h \vee T}^h - W_{\tau_{[T/h]}^h \wedge T}^h|^4] \geq \frac{1}{4} E[(\tau_{[T/h]}^h \vee T - \tau_{[T/h]}^h \wedge T)^2] \\ &= \frac{1}{4} E[(\tau_{[T/h]}^h - T)^2] \geq \frac{1}{4} E \left[\left(\tau_{[T/h]}^h - h \left\lfloor \frac{T}{h} \right\rfloor \right)^2 \right]. \end{aligned} \quad (95)$$

Moreover, it holds that $E[\tau_{k+1}^h - \tau_k^h | \mathcal{F}_{\tau_k^h}] = h$ for all $k \in \mathbb{N}_0$. In particular, the increments $\tau_{k+1}^h - \tau_k^h$, $k \in \mathbb{N}_0$, are uncorrelated. Hence we have

$$\begin{aligned} E \left[\left(\tau_{[T/h]}^h - h \left\lfloor \frac{T}{h} \right\rfloor \right)^2 \right] &= \text{Var}(\tau_{[T/h]}^h) = \text{Var} \left(\sum_{k=0}^{\lfloor T/h \rfloor - 1} (\tau_{k+1}^h - \tau_k^h) \right) \\ &= \sum_{k=0}^{\lfloor T/h \rfloor - 1} \text{Var}(\tau_{k+1}^h - \tau_k^h). \end{aligned} \quad (96)$$

The assumption that for all $y \in \mathbb{R}$ we have $a_h(y) = \sqrt{h}$ implies that $\tau_{k+1}^h - \tau_k^h \stackrel{d}{=} H_{-\sqrt{h}, \sqrt{h}}(W)$. Therefore we obtain

$$E \left[\left(\tau_{[T/h]}^h - h \left\lfloor \frac{T}{h} \right\rfloor \right)^2 \right] = \left\lfloor \frac{T}{h} \right\rfloor \text{Var} \left(H_{-\sqrt{h}, \sqrt{h}}(W) \right). \quad (97)$$

The self-similarity of the Brownian motion states that the process $(\sqrt{h}W_{\frac{t}{h}})_{t \in [0, \infty)}$ is again a Brownian motion. This implies that $H_{-\sqrt{h}, \sqrt{h}}(W) \stackrel{d}{=} hH_{-1,1}(W)$. Therefore we obtain that

$$E \left[\left(\tau_{\lfloor T/h \rfloor}^h - h \left\lfloor \frac{T}{h} \right\rfloor \right)^2 \right] = h^2 \left\lfloor \frac{T}{h} \right\rfloor \text{Var}(H_{-1,1}(W)) \geq h(T-h) \text{Var}(H_{-1,1}(W)). \quad (98)$$

Combining this with (95) completes the proof. \square

6. Convergence results on the path space

In this section, we prove Theorems 1.6, 1.7 and 1.9. Again let $(\tau_k^h)_{k \in \mathbb{N}_0}$, $h \in (0, \bar{h})$, be the embedding stopping times introduced in Section 4. We define a continuous-time process $Y^h = (Y_t^h)_{t \in [0, \infty)}$ via linear interpolation of $(Y_{\tau_k^h})_{k \in \mathbb{N}_0}$:

$$Y_t^h = Y_{\tau_{\lfloor t/h \rfloor}^h} + (t/h - \lfloor t/h \rfloor)(Y_{\tau_{\lfloor t/h \rfloor + 1}^h} - Y_{\tau_{\lfloor t/h \rfloor}^h}), \quad t \in [0, \infty). \quad (99)$$

Then it follows from (70) that for all $h \in (0, \bar{h})$ and $y \in I^\circ$ we have

$$\text{Law}_{P_y}(Y_t^h; t \in [0, \infty)) = \text{Law}_P(X_t^{h,y}; t \in [0, \infty)) \quad (100)$$

(indeed, cf. (99) with (11)). We proceed similarly as in Section 5 and first establish L^p bounds between the processes Y and Y^h .

Theorem 6.1. *Suppose that Condition (C) is satisfied and that there exists $\lambda \in (0, \infty)$ such that Condition (A λ) holds. Let $p \in [1, \infty)$, $T \in (0, \infty)$ and $\varepsilon \in (0, \min\{\frac{1}{4}, \frac{\lambda}{2}\})$. Then there exists a constant $C(p, \varepsilon, T) \in (0, \infty)$ such that for all $y \in I^\circ$ and $h \in (0, \bar{h})$ we have*

$$\left\| \sup_{t \in [0, T]} |Y_t^h - Y_t| \right\|_{L^p(P_y)} \leq C(p, \varepsilon, T)(1 + k_2|y|)h^{\min\{\frac{1}{4}, \frac{\lambda}{2}\} - \varepsilon}. \quad (101)$$

Proof. To simplify some expressions we treat explicitly only the case $\lambda \in (0, \frac{1}{2}]$ (in general, one should replace all instances of λ in the proof with $\min\{\frac{1}{2}, \lambda\}$). Without loss of generality we assume that $\varepsilon \in (0, \frac{1}{p} \wedge \frac{\lambda}{2})$. Throughout the proof we fix $h \in (0, \bar{h})$ and $y \in I^\circ$. We use the notation $C_i(p, \varepsilon, T) \in (0, \infty)$, $i \in \mathbb{N}$, for constants independent of h and y .

First, observe that the Markov inequality and Proposition 4.1 imply that there exists a constant $C_1(p, T) \in (0, \infty)$ such that

$$P_y[\tau_{\lfloor T/h \rfloor}^h \geq T + 1] \leq P_y \left[\tau_{\lfloor T/h \rfloor}^h - h \left\lfloor \frac{T}{h} \right\rfloor \geq 1 \right] \leq E_y \left[\left| \tau_{\lfloor T/h \rfloor}^h - h \left\lfloor \frac{T}{h} \right\rfloor \right|^p \right] \leq C_1(p, T)h^{\lambda p}. \quad (102)$$

This, the Cauchy-Schwarz inequality, Theorems 3.1 and 3.2 together with (100) prove that there exists a constant $C_3(p, T) \in (0, \infty)$ such that

$$\begin{aligned}
& \left\| 1_{\{\tau_{\lfloor T/h \rfloor}^h \geq T+1\}} \sup_{t \in [0, T]} |Y_t^h - Y_t| \right\|_{L^p(P_y)} \leq (P_y[\tau_{\lfloor T/h \rfloor}^h \geq T+1])^{\frac{1}{2p}} \left\| \sup_{t \in [0, T]} |Y_t^h - Y_t| \right\|_{L^{2p}(P_y)} \\
& \leq C_1(p, T)^{\frac{1}{2p}} h^{\frac{\lambda}{2}} \left(\left\| \sup_{t \in [0, T]} |Y_t^h - y| \right\|_{L^{2p}(P_y)} + \left\| \sup_{t \in [0, T]} |Y_t - y| \right\|_{L^{2p}(P_y)} \right) \\
& \leq C_3(p, T)(1 + k_2|y|)h^{\frac{\lambda}{2}}.
\end{aligned} \tag{103}$$

Next, by the triangle inequality we have that

$$\sup_{t \in [0, T]} |Y_t^h - Y_t| \leq \sup_{t \in [0, T]} |Y_t^h - Y_{\tau_{\lfloor t/h \rfloor}^h}| + \sup_{t \in [0, T]} |Y_{\tau_{\lfloor t/h \rfloor}^h} - Y_t|. \tag{104}$$

We consider the two terms on the right-hand side of Inequality (104) separately on the set $\{\tau_{\lfloor T/h \rfloor}^h < T+1\}$. We start with the latter one. By Theorem 3.1, there exists $C_4(\varepsilon, T) \in (0, \infty)$ such that for all $s, t \in [0, T+1]$ we have

$$E_y[|Y_s - Y_t|^{\frac{1}{\varepsilon}}] \leq C_4(\varepsilon, T)(1 + k_2|y|)^{\frac{1}{\varepsilon}}|s - t|^{\frac{1}{2\varepsilon}}. \tag{105}$$

Together with the fact that $0 \leq \frac{1}{2} - \frac{\varepsilon}{\lambda} < \varepsilon(\frac{1}{2\varepsilon} - 1)$, this allows to apply the Kolmogorov-Chentsov theorem (see, e.g., Theorem I.2.1 in [36]) to obtain that there exists a constant $C_5(\varepsilon, T) \in (0, \infty)$ such that

$$\left\| \sup_{\substack{s, t \in [0, T+1] \\ s \neq t}} \frac{|Y_s - Y_t|}{|s - t|^{\frac{1}{2} - \frac{\varepsilon}{\lambda}}} \right\|_{L^{\frac{1}{\varepsilon}}(P_y)} \leq C_5(\varepsilon, T)(1 + k_2|y|). \tag{106}$$

Moreover, it holds that

$$\sup_{t \in [0, T]} |\tau_{\lfloor t/h \rfloor}^h - t| \leq \sup_{k \in \{0, \dots, \lfloor T/h \rfloor\}} \sup_{t \in [kh, (k+1)h)} |\tau_k^h - t| \leq \sup_{k \in \{0, \dots, \lfloor T/h \rfloor\}} |\tau_k^h - kh| + h. \tag{107}$$

Since $\frac{1}{2} - \frac{\varepsilon}{\lambda} < 1$, we further obtain $\sup_{t \in [0, T]} |\tau_{\lfloor t/h \rfloor}^h - t|^{\frac{1}{2} - \frac{\varepsilon}{\lambda}} \leq \sup_{k \in \{0, \dots, \lfloor T/h \rfloor\}} |\tau_k^h - kh|^{\frac{1}{2} - \frac{\varepsilon}{\lambda}} + h^{\frac{1}{2} - \frac{\varepsilon}{\lambda}}$. This, together with Proposition 4.1, implies that for all $\alpha \in [1, \infty)$ there exists a constant $C_6(\alpha, \varepsilon, T) \in (0, \infty)$ such that

$$\begin{aligned}
\left\| \sup_{t \in [0, T]} |\tau_{\lfloor t/h \rfloor}^h - t|^{\frac{1}{2} - \frac{\varepsilon}{\lambda}} \right\|_{L^\alpha(P_y)} & \leq h^{\frac{1}{2} - \frac{\varepsilon}{\lambda}} + \left\| \sup_{k \in \{0, \dots, \lfloor T/h \rfloor\}} |\tau_k^h - kh|^{\frac{1}{2} - \frac{\varepsilon}{\lambda}} \right\|_{L^\alpha(P_y)} \\
& \leq C_6(\alpha, \varepsilon, T)h^{\frac{\lambda}{2} - \varepsilon}.
\end{aligned} \tag{108}$$

On the set $\{\tau_{\lfloor T/h \rfloor}^h < T+1\}$ it holds that

$$\sup_{t \in [0, T]} |Y_{\tau_{\lfloor t/h \rfloor}^h} - Y_t| \leq \left[\sup_{\substack{s, t \in [0, T+1] \\ s \neq t}} \frac{|Y_s - Y_t|}{|s - t|^{\frac{1}{2} - \frac{\varepsilon}{\lambda}}} \right] \left[\sup_{t \in [0, T]} |\tau_{\lfloor t/h \rfloor}^h - t|^{\frac{1}{2} - \frac{\varepsilon}{\lambda}} \right].$$

This, the Hölder inequality (applied with the conjugates $\frac{1}{p\varepsilon}$ and $\frac{1}{1-p\varepsilon}$), (106) and (108) ensure that there exists a constant $C_7(p, \varepsilon, T) \in (0, \infty)$ such that

$$\begin{aligned} & \left\| 1_{\{\tau_{\lfloor T/h \rfloor}^h < T+1\}} \sup_{t \in [0, T]} |Y_{\tau_{\lfloor t/h \rfloor}^h} - Y_t| \right\|_{L^p(P_y)} \\ & \leq \left\| \sup_{\substack{s, t \in [0, T+1] \\ s \neq t}} \frac{|Y_s - Y_t|}{|s - t|^{\frac{1}{2} - \frac{\varepsilon}{\lambda}}} \right\|_{L^{\frac{p}{1-p\varepsilon}}(P_y)} \left\| \sup_{t \in [0, T]} |\tau_{\lfloor t/h \rfloor}^h - t|^{\frac{1}{2} - \frac{\varepsilon}{\lambda}} \right\|_{L^{\frac{p}{1-p\varepsilon}}(P_y)} \leq C_7(p, \varepsilon, T)(1 + k_2|y|)h^{\frac{\lambda}{2} - \varepsilon}. \quad (109) \end{aligned}$$

Next, we consider the first term on the right-hand side of (104). By definition (99) of Y^h it holds that

$$\sup_{t \in [0, T]} |Y_t^h - Y_{\tau_{\lfloor t/h \rfloor}^h}| \leq \sup_{k \in \{0, \dots, \lfloor T/h \rfloor\}} |Y_{\tau_{k+1}^h} - Y_{\tau_k^h}| = \sup_{k \in \{0, \dots, \lfloor T/h \rfloor\}} |a_h(Y_{\tau_k^h})|. \quad (110)$$

Then Proposition 2.3, Theorem 3.2 and (100) show that there exist $C_8, C_9(p, T) \in (0, \infty)$ such that

$$\begin{aligned} & \left\| 1_{\{\tau_{\lfloor T/h \rfloor}^h < T+1\}} \sup_{t \in [0, T]} |Y_t^h - Y_{\tau_{\lfloor t/h \rfloor}^h}| \right\|_{L^p(P_y)} \leq C_8 h^{\frac{1}{2}} \left\| \sup_{k \in \{0, \dots, \lfloor T/h \rfloor\}} (1 + k_2|Y_{\tau_k^h}|) \right\|_{L^p(P_y)} \\ & \leq C_9(p, T)(1 + k_2|y|)h^{\frac{1}{2}}. \end{aligned} \quad (111)$$

Combining (104) with (109) and (111) shows that there exists $C_{10}(p, \varepsilon, T) \in (0, \infty)$ such that

$$\left\| 1_{\{\tau_{\lfloor T/h \rfloor}^h < T+1\}} \sup_{t \in [0, T]} |Y_t^h - Y_t| \right\|_{L^p(P_y)} \leq C_{10}(p, \varepsilon, T)(1 + k_2|y|)h^{\frac{\lambda}{2} - \varepsilon}. \quad (112)$$

Finally, the triangle inequality, (103) and (112) imply the claim. \square

Proof of Theorem 1.6. Throughout the proof let $p \in [1, \infty)$ and $T \in (0, \infty)$. First note that due to Theorems 3.1 and 3.2 the law of Y and the law of $X^{h,y}$, $h \in (0, \bar{h})$, $y \in I^\circ$, on the path space $C([0, T], I)$ are elements of $\mathcal{M}_p(C([0, T], I))$. It follows from (100) that the random element (Y^h, Y) constitutes a coupling between the laws of $X^{h,y}$ and Y on the path space $C([0, T], I)$. Therefore, Theorem 1.6 follows from Theorem 6.1. \square

Proof of Theorem 1.7. We use the shorthand notations $X^{h,y}$ and Y for the processes $(X_t^{h,y})_{t \in [0, T]}$ and $(Y_t)_{t \in [0, T]}$, respectively. Theorem 4.1 in [38] ensures that for any $h \in (0, \bar{h})$ and $y \in I^\circ$ there exists an optimal coupling (ξ^h, ζ) on a probability space with a measure $P_{h,y}^*$ between the law of $X^{h,y}$ and the law of Y with respect to the \mathcal{W}_2 -distance, i.e., it holds

$$\begin{aligned} \text{Law}_{P_{h,y}^*}(\xi^h; t \in [0, T]) &= \text{Law}_P(X_t^{h,y}; t \in [0, T]), \\ \text{Law}_{P_{h,y}^*}(\zeta; t \in [0, T]) &= \text{Law}_{P_y}(Y_t; t \in [0, T]), \\ \mathcal{W}_2(P \circ (X^{h,y})^{-1}, P_y \circ (Y)^{-1}) &= \sqrt{E_{h,y}^*[\|\xi^h - \zeta\|_C^2]}. \end{aligned}$$

It holds for all $h \in (0, \bar{h})$ and $y \in I^\circ$ that

$$\left| E[F(X_t^{h,y}; t \in [0, T])] - E_y[F(Y_t; t \in [0, T])] \right| = |E_{h,y}^*[F(\xi^h) - F(\zeta)]|.$$

This together with (27), the Cauchy-Schwarz inequality and the triangle inequality in L^2 implies for all $h \in (0, \bar{h})$ and $y \in I^\circ$ that

$$\begin{aligned}
& \left| E \left[F(X_t^{h,y}; t \in [0, T]) \right] - E_y \left[F(Y_t; t \in [0, T]) \right] \right| \leq L E_{h,y}^* \left[\{1 + (\|\xi^h\|_C \vee \|\zeta\|_C)^\alpha\} \|\xi^h - \zeta\|_C \right] \\
& \leq L \sqrt{E_{h,y}^* \left[(1 + (\|\xi^h\|_C \vee \|\zeta\|_C)^\alpha)^2 \right]} \sqrt{E_{h,y}^* [\|\xi^h - \zeta\|_C^2]} \\
& \leq L \left(1 + \sqrt{E_{h,y}^* [(\|\xi^h\|_C \vee \|\zeta\|_C)^{2\alpha}] } \right) \sqrt{E_{h,y}^* [\|\xi^h - \zeta\|_C^2]}.
\end{aligned} \tag{113}$$

It follows from Theorem 1.6 that there exists $C_1 \in [0, \infty)$ such that for all $h \in (0, \bar{h})$ and $y \in I^\circ$

$$\sqrt{E_{h,y}^* [\|\xi^h - \zeta\|_C^2]} = \mathcal{W}_2(P \circ (X^{h,y})^{-1}, P_y \circ (Y)^{-1}) \leq C_1(1 + k_2|y|)h^{\min\{\frac{1}{4}, \frac{\lambda}{2}\} - \varepsilon}. \tag{114}$$

For the next step first assume that $\alpha > 0$ and let $\beta = \max\{2\alpha, 1\}$. By Jensen's inequality and Theorems 3.1 and 3.2 there exists $C_2 \in [0, \infty)$ such that we have for all $h \in (0, \bar{h})$ and $y \in I^\circ$

$$\begin{aligned}
\sqrt{E_{h,y}^* [(\|\xi^h\|_C \vee \|\zeta\|_C)^{2\alpha}]} &= (E_{h,y}^* [(\|\xi^h\|_C \vee \|\zeta\|_C)^{2\alpha}])^{\frac{\alpha}{2\alpha}} \leq (E_{h,y}^* [(\|\xi^h\|_C \vee \|\zeta\|_C)^\beta])^{\frac{\alpha}{\beta}} \\
&= \|\|\xi^h\|_C \vee \|\zeta\|_C\|_{L^\beta(P_{h,y}^*)}^\alpha \leq \left(\|\|\xi^h\|_C\|_{L^\beta(P_{h,y}^*)} + \|\|\zeta\|_C\|_{L^\beta(P_{h,y}^*)} \right)^\alpha \\
&= \left(\|\|X^h\|_C\|_{L^\beta(P_{h,y}^*)} + \|\|Y\|_C\|_{L^\beta(P_{h,y}^*)} \right)^\alpha \leq C_2(1 + |y|^\alpha).
\end{aligned} \tag{115}$$

Note that (115) clearly also holds in the case $\alpha = 0$. Combining (113), (114) and (115) establishes (28) and completes the proof. \square

Proof of Theorem 1.9. Throughout the proof let $T \in (0, \infty)$, $y \in I^\circ$ and $p \in [1, \infty)$. As in the proof of Theorem 1.6, Statement (26) follows from Theorems 3.1 and 3.2. Let $F: C([0, T], I) \rightarrow \mathbb{R}$ be a continuous path functional satisfying

$$|F(x)| \leq L(1 + \|x\|_C^p), \quad x \in C([0, T], I),$$

with some constant $L \in [0, \infty)$. By Theorem 3.2, we have $\sup_{h \in (0, \bar{h})} E \left[\|X^{h,y}\|_C^{2p} \right] < \infty$, hence

$$\sup_{h \in (0, \bar{h})} E \left[F(X^{h,y})^2 \right] < \infty,$$

which implies the uniform integrability of the family $\{F(X^{h,y})\}_{h \in (0, \bar{h})}$. By the main result in [6], we have the weak convergence $X^{h,y} \xrightarrow{w} Y$ on the path space. The continuity of the functional F yields $F(X^{h,y}) \xrightarrow{w} F(Y)$, which, together with the uniform integrability of $\{F(X^{h,y})\}_{h \in (0, \bar{h})}$, implies (31). Finally, the equivalence between (30) and (31) follows from Theorem 6.9 in [38] (also consult Definition 6.8 there). \square

7. On Condition (C)

In this section we show that Condition (C) is a nearly minimal assumption for our results. To this end fix any starting point $y \in I^\circ$ and note that Y is a P_y -local martingale due to our assumption that the boundaries are inaccessible or absorbing. Let us consider the conditions

$$r = \infty \quad \text{and} \quad \int^\infty x m(dx) < \infty \tag{116}$$

and

$$l = -\infty \quad \text{and} \quad \int_{-\infty}^{\infty} |x| m(dx) < \infty, \quad (117)$$

where the notation $\int_{-\infty}^{\infty} x m(dx) < \infty$ is understood as $\int_z^{\infty} x m(dx) < \infty$ for some (equivalently, for any) $z \in I^\circ$, and the notation $\int_{-\infty}^{\infty} |x| m(dx) < \infty$ is understood in a similar way. The article [30] establishes that Y is a strict P_y -local martingale⁴ if and only if at least one of the conditions (116) and (117) is satisfied. [19] complements this result by showing that

- (a) Y is not a P_y -supermartingale if and only if (117) holds;
- (b) Y is not a P_y -submartingale if and only if (116) holds.

Note that under Condition (C) neither (116) nor (117) is satisfied.

Example 7.1. Suppose that $r = \infty$ and that the speed measure is given by $m(dx) = \frac{2}{\eta^2(x)} dx$ on I° with η satisfying $\liminf_{x \rightarrow \infty} \frac{\eta^2(x)}{x^2 \log^\alpha(x)} > 0$ for some $\alpha \in (1, \infty)$. The fact that

$$\lim_{x \rightarrow \infty} \int_e^x \frac{z}{z^2 (\log(z))^\alpha} dz = \frac{1}{\alpha - 1} \lim_{x \rightarrow \infty} \left(1 - \frac{1}{(\log(x))^{\alpha-1}} \right) = \frac{1}{\alpha - 1}$$

implies that (116) is satisfied. In particular, it follows from [30] that the associated diffusion Y is a strict P_y -local martingale. Hence, there exists $T \in (0, \infty)$ such that $E_y[\sup_{t \in [0, T]} |Y_t|] = \infty$. In particular, this implies that the claim of Theorem 3.1 (and hence that of Theorem 1.6) does not hold true in this example, even for $p = 1$, as the law of Y does not belong to $\mathcal{M}_1(C([0, T], I))$. Moreover, also the result on path-dependent rate in L^p , Theorem 6.1, which a priori does not require the law of Y to belong to $\mathcal{M}_p(C([0, T], I))$ (only the difference $Y^h - Y$ matters), does not hold true, even for $p = 1$, as for any fixed $h \in (0, \bar{h})$ the law of Y^h belongs to $\mathcal{M}_1(C([0, T], I))$ (see (100) and (10)–(11)). The error criterion in the left-hand side of (101) is just too strong for this example regardless of the specific approximation scheme used.

Notice that Condition (C) is only marginally violated in this example.

Example 7.2. Consider any general diffusion Y for which

$$\text{one of (116)–(117) is satisfied, while the other is not.} \quad (118)$$

As in the previous example, it is possible to only marginally violate Condition (C) and to achieve (118). It follows from Statements (a) and (b) above that Y is either a strict P_y -supermartingale or a strict P_y -submartingale.⁵ In particular, there exists $T \in (0, \infty)$ such that $E_y[Y_T] \neq y$. The approximating Markov chain $(X_{kh}^{y,h})_{k \in \mathbb{N}_0}$ is always a discrete-time martingale (see (10) and also notice that each random variable $X_{kh}^{y,h}$ takes finitely many values and, in particular, belongs to L^1). Therefore, we always have $E[X_T^{y,h}] = y$. Thus, in this example we have

$$E[X_T^{y,h}] \not\rightarrow E_y[Y_T] \quad \text{as } h \rightarrow 0,$$

and hence, the claim of Theorem 1.9 does not hold true even for the (non-path-dependent) functional $F(x) = x(T)$. Similarly, the claim of Theorem 1.4 (let alone the ones of Theorems 1.6 and 1.7) are violated as well whenever (118) is satisfied.

⁴ A *strict* local martingale is a local martingale that fails to be a martingale.

⁵ A *strict* supermartingale is a supermartingale that is not a martingale. A *strict* submartingale is understood in a similar way.

Acknowledgments

We thank two anonymous referees for the careful reading of the manuscript and for their comments. Thomas Kruse and Mikhail Urusov acknowledge the support from the *German Research Foundation* through the project 415705084.

References

- [1] A. Alfonsi, B. Jourdain, A. Kohatsu-Higa, Pathwise optimal transport bounds between a one-dimensional diffusion and its Euler scheme, *Ann. Appl. Probab.* 24 (3) (2014) 1049–1080.
- [2] S. Ankirchner, N. Kazi-Tani, M. Klein, T. Kruse, Stopping with expectation constraints: 3 points suffice, *Electron. J. Probab.* 24 (2019) 66.
- [3] S. Ankirchner, M. Klein, T. Kruse, M. Urusov, on a certain local martingale in a general diffusion setting, Preprint, hal-01700656, 2018.
- [4] S. Ankirchner, T. Kruse, W. Löhr, M. Urusov, Properties of the EMCEL scheme for approximating irregular diffusions, Preprint, arXiv:2004.10316, 2020.
- [5] S. Ankirchner, T. Kruse, M. Urusov, Numerical approximation of irregular SDEs via Skorokhod embeddings, *J. Math. Anal. Appl.* 440 (2) (2016) 692–715.
- [6] S. Ankirchner, T. Kruse, M. Urusov, A functional limit theorem for coin tossing Markov chains, *Ann. Inst. Henri Poincaré Probab. Stat.* (2020), in press.
- [7] R.F. Bass, A stochastic differential equation with a sticky point, *Electron. J. Probab.* 19 (o) (2014) 32.
- [8] C. Brugger, C. de Schryver, N. Wehn, S. Omland, M. Hefter, K. Ritter, A. Kostiuik, R. Korn, Mixed precision multi-level Monte Carlo on hybrid computing systems, in: 2014 IEEE Conference on Computational Intelligence for Financial Engineering Economics (CIFEr), March 2014, pp. 215–222.
- [9] B. Can, M. Caglar, Conditional law and occupation times of two-sided sticky Brownian motion, Preprint, arXiv:1910.10213, 2019.
- [10] A. Eberle, R. Zimmer, Sticky couplings of multidimensional diffusions with different drifts, *Ann. Inst. Henri Poincaré Probab. Stat.* 55 (4) (2019) 2370–2394.
- [11] H.-J. Engelbert, G. Peskir, Stochastic differential equations for sticky Brownian motion, *Stochastics* 86 (6) (2014) 993–1021.
- [12] H.J. Engelbert, W. Schmidt, On solutions of one-dimensional stochastic differential equations without drift, *Z. Wahrscheinlichkeitstheor. Verw. Geb.* 68 (3) (1985) 287–314.
- [13] P. Etoré, A. Lejay, A Donsker theorem to simulate one-dimensional processes with measurable coefficients, *ESAIM Probab. Stat.* 11 (2007) 301–326.
- [14] T. Fattler, M. Grothaus, R. Voßhall, Construction and analysis of a sticky reflected distorted Brownian motion, *Ann. Inst. Henri Poincaré Probab. Stat.* 52 (2) (2016) 735–762.
- [15] N. Frikha, On the weak approximation of a skew diffusion by an Euler-type scheme, *Bernoulli* 24 (3) (2018) 1653–1691.
- [16] M.B. Giles, M. Hefter, L. Mayer, K. Ritter, Random bit quadrature and approximation of distributions on Hilbert spaces, *Found. Comput. Math.* 19 (1) (2019) 205–238.
- [17] M. Grothaus, R. Voßhall, Stochastic differential equations with sticky reflection and boundary diffusion, *Electron. J. Probab.* 22 (2017) 7.
- [18] M. Grothaus, R. Voßhall, Strong Feller property of sticky reflected distorted Brownian motion, *J. Theor. Probab.* 31 (2) (2018) 827–852.
- [19] A. Gushchin, M. Urusov, M. Zervos, On the submartingale/supermartingale property of diffusions in natural scale, *Proc. Steklov Inst. Math.* 287 (1) (2014) 122–132.
- [20] H. Hajri, M. Caglar, M. Arnaudon, Application of stochastic flows to the sticky Brownian motion equation, *Electron. Commun. Probab.* 22 (2017) 3.
- [21] M. Hutzenthaler, A. Jentzen, P. Kloeden, Strong and weak divergence in finite time of Euler’s method for stochastic differential equations with non-globally Lipschitz continuous coefficients, *Proc. R. Soc. Lond., Ser. A, Math. Phys. Eng. Sci.* 467 (2130) (2011) 1563–1576.
- [22] O. Kallenberg, *Foundations of Modern Probability*, second edition, Probability and Its Applications (New York), Springer-Verlag, New York, 2002.
- [23] I. Karatzas, A.N. Shiryaev, M. Shkolnikov, On the one-sided Tanaka equation with drift, *Electron. Commun. Probab.* 16 (2011) 664–677.
- [24] I. Karatzas, S.E. Shreve, *Brownian Motion and Stochastic Calculus*, second edition, Graduate Texts in Mathematics, vol. 113, Springer-Verlag, New York, 1991.
- [25] P. Kloeden, E. Platen, *Numerical Solution of Stochastic Differential Equations*, vol. 23, Springer, 1992.
- [26] A. Kohatsu-Higa, A. Lejay, K. Yasuda, Weak rate of convergence of the Euler-Maruyama scheme for stochastic differential equations with non-regular drift, *J. Comput. Appl. Math.* 326 (2017) 138–158.
- [27] V. Konakov, S. Menozzi, Weak error for the Euler scheme approximation of diffusions with non-smooth coefficients, *Electron. J. Probab.* 22 (2017).
- [28] V. Konarovskiyi, Coalescing-fragmentating Wasserstein dynamics: particle approach, Preprint, arXiv:1711.03011v3, 2017.
- [29] V. Konarovskiyi, M. von Renesse, Reversible coalescing-fragmentating Wasserstein dynamics on the real line, Preprint, arXiv:1709.02839v2, 2017.
- [30] S. Kotani, On a condition that one-dimensional diffusion processes are martingales, in: *Seminar on Probability, XXXIX*, in: *Lecture Notes in Math.*, vol. 1874, Springer, Berlin, 2006, pp. 149–156.

- [31] A. Lejay, L. Lenôtre, G. Pichot, An exponential timestepping algorithm for diffusion with discontinuous coefficients, *J. Comput. Phys.* 396 (2019) 888–904.
- [32] G.N. Milstein, J. Schoenmakers, Uniform approximation of the Cox-Ingersoll-Ross process via exact simulation at random times, *Adv. Appl. Probab.* 48 (4) (2016) 1095–1116.
- [33] H.-L. Ngo, D. Taguchi, Strong convergence for the Euler-Maruyama approximation of stochastic differential equations with discontinuous coefficients, *Stat. Probab. Lett.* 125 (2017) 55–63.
- [34] G. Pagès, *Numerical Probability*, Universitext, Springer, Cham, 2018, An introduction with applications to finance.
- [35] T. Piskorski, M.M. Westerfield, Optimal dynamic contracts with moral hazard and costly monitoring, *J. Econ. Theory* 166 (2016) 242–281.
- [36] D. Revuz, M. Yor, *Continuous Martingales and Brownian Motion*, third edition, Grundlehren der Mathematischen Wissenschaften (Fundamental Principles of Mathematical Sciences), vol. 293, Springer-Verlag, Berlin, 1999.
- [37] L.C.G. Rogers, D. Williams, *Diffusions, Markov Processes, and Martingales*, Cambridge Mathematical Library, vol. 2, Cambridge University Press, Cambridge, 2000, Itô calculus, Reprint of the second (1994) edition.
- [38] C. Villani, *Optimal Transport. Old and New*, Grundlehren der Mathematischen Wissenschaften (Fundamental Principles of Mathematical Sciences), vol. 338, Springer-Verlag, Berlin, 2009.
- [39] J.Y. Zhu, Optimal contracts with shirking, *Rev. Econ. Stud.* 80 (2) (2013) 812–839.