

# Weak Convergence in the Prokhorov Metric of Methods for Stochastic Differential Equations

BENOIT CHARBONNEAU<sup>1</sup>, YURIY SVYRYDOV<sup>2</sup>, P. F. TUPPER<sup>3</sup>

<sup>1</sup>Mathematics Department, Duke University, Durham, NC, U.S.A.

<sup>2</sup>Technische Universität München, Munich, Germany.

<sup>3</sup>Department of Mathematics, Simon Fraser University, Burnaby, BC, Canada.

October 19, 2008

## Abstract

We consider the weak convergence of numerical methods for stochastic differential equations (SDEs). Weak convergence is usually expressed in terms of the convergence of expected values of test functions of the trajectories. Here we present an alternative formulation of weak convergence in terms of the well-known Prokhorov metric on spaces of random variables. For a general class of methods, we establish bounds on the rates of convergence in terms of the Prokhorov metric. In doing so, we revisit the original proofs of weak convergence and show explicitly how the bounds on the error depend on the smoothness of the test functions. As an application of our result, we use the Strassen–Dudley theorem to show that the numerical approximation and the true solution to the system of SDEs can be re-embedded in a probability space in such a way that the method converges there in a strong sense. One corollary of this last result is that the method converges in the Wasserstein distance, another metric on spaces of random variables. Another corollary establishes rates of convergence for expected values of test functions assuming only local Lipschitz continuity. We conclude with a review of the existing results for pathwise convergence of weakly converging methods and the corresponding strong results available under re-embedding.

## 1 Introduction

Consider the following system of Ito stochastic differential equations (SDEs)

$$dX = a(X)dt + \sum_{r=1}^q \sigma_r(X)dW_r(t), \quad X(0) = x_0, \quad (1.1)$$

for  $X(t) \in \mathbb{R}^n$ , where the  $W_r(t)$  are independent standard Wiener processes. The simplest numerical method for obtaining approximate solutions to this system is the Euler–Maruyama method: for  $k \geq 0$ , timestep  $\Delta t$ , and  $\Delta_k W_r = W_r((k+1)\Delta t) - W_r(k\Delta t)$ ,

$$X_{k+1} = X_k + a(X_k)\Delta t + \sum_{r=1}^q \sigma_r(X_k)\Delta_k W_r, \quad X_0 = x_0. \quad (1.2)$$

For each  $k$ ,  $X_k$  is an approximation to  $X(k\Delta t)$ . The Euler–Maruyama method converges in the *strong* sense because for each realization of the  $W_r(t)$ , the method gives an approximation to the exact solution of the SDE with that same realization. In particular, as shown in [9, p. 342],

$$(\mathbb{E}(X(T) - X_{T/\Delta t})^2)^{1/2} \leq C\Delta t^{1/2}, \quad (1.3)$$

under certain assumptions on the coefficients  $a$  and  $\sigma$ . In order for such a result to be possible  $X(T)$  and  $X_{T/\Delta t}$  must be defined on the same probability space.

Another way to quantify convergence of a numerical method is to consider the distribution of the random variable generated by the numerical method and see how close it is to the distribution of the true trajectory at the corresponding

point in time. This concept is known as *weak* convergence. The typical way to quantify this is through test functions. For example, for sufficiently smooth functions  $f$  the Euler–Maruyama method (1.2) satisfies

$$|\mathbb{E}f(X_{T/\Delta t}) - \mathbb{E}f(X(T))| \leq C_f \Delta t \quad (1.4)$$

for some constant  $C_f$  depending on  $f$ . See [9, p. 473], or [14] for an important earlier reference.

Strong convergence of a method implies weak convergence, but the converse is not true. For example, let  $N_{rk}$  be independent identically distributed random variables with  $N_{rk} = \pm 1$  with probability  $1/2$ , and set

$$X_{k+1} = X_k + a(X_k)\Delta t + \sum_{r=1}^q \sigma_r(X_k)\Delta t^{1/2}N_{rk}, \quad (1.5)$$

We can define the  $N_{rk}$  in various ways. One possibility is for them to be independent of the Wiener processes driving the SDE (1.1); another is to choose them to be  $N_{rk} = \text{sgn}(\Delta_k W_k)$ . In either case, this method, which we call weak Euler–Maruyama, does not converge strongly to the solution of the SDE. However, (1.4) still holds whatever the relation between the  $N_{rk}$  and the Wiener processes, even if they are defined on different probability spaces.

In this paper, we present a different formulation of weak convergence of numerical method in terms of the Prokhorov metric [2]. For any two random elements of  $\mathbb{R}^n$  the Prokhorov metric gives a quantitative measurement of how far apart their distributions are. Its importance in probability theory [3] is that convergence in the Prokhorov metric is equivalent to convergence in distribution (or weak convergence, as it is sometimes known). Before we discuss our results we first review some of the definitions and facts of convergence in distribution in metric spaces. See Billingsley’s book [2] for details.

Consider a metric space  $S$  with metric  $d$ , such as  $\mathbb{R}^n$  with the Euclidean metric. We say that a sequence of random elements  $X_n$  in  $S$  *converges in distribution* to  $X$  in  $S$  if for all bounded continuous  $f : S \rightarrow \mathbb{R}$

$$\mathbb{E}f(X_n) \rightarrow \mathbb{E}f(X), \quad (1.6)$$

as  $n \rightarrow \infty$ . An equivalent definition of convergence in distribution of  $X_n$  to  $X$  is that for all Borel sets  $A$  of  $S$  with  $\mathbb{P}(X \in \partial A) = 0$  we have

$$\mathbb{P}(X_n \in A) \rightarrow \mathbb{P}(X \in A) \quad (1.7)$$

as  $n \rightarrow \infty$ . (Here  $\partial A$  is the boundary of the set  $A$ .) The assumption of boundedness on  $f$  may seem excessive, but in the presence of uniform bounds on the moments of  $X_n$  and  $X$ , convergence in distribution implies that (1.6) holds for more general continuous  $f$  [4, p. 86].

It is not obvious from either of the above definitions of weak convergence how to measure the speed with which a sequence  $X_n$  converges in distribution to  $X$ , since the rate at which limits (1.6) and (1.7) occur depends on  $f$  and  $A$  respectively. The Prokhorov metric is one way to define the distance between the distributions of two random elements, and thus allows us to quantify convergence in distribution. For any two random elements  $X$  and  $Y$  of  $S$  let  $\rho(X, Y)$  be the Prokhorov distance between them (see Section 2 for the definition). This distance is zero if and only if  $X$  and  $Y$  have the same distribution, that is if  $\mathbb{P}(X \in A) = \mathbb{P}(Y \in A)$  for all Borel sets  $A$ . Moreover, if  $X_n$  is a sequence of random elements in a separable metric space  $S$ ,  $\rho(X_n, X) \rightarrow 0$  if and only if  $X_n$  converges in distribution to  $X$ . Thus we say that the Prokhorov metric metrizes convergence in distribution.

In our case, we view the solution of the system of SDEs at time  $T$  as random vector (a random element of  $\mathbb{R}^n$ ), and likewise for the numerical solution at time  $T$ . Then we ask how the Prokhorov distance between  $X(T)$  and  $X_{T/\Delta t}$  depends on  $\Delta t$ . Our main result in Section 3 shows that the usual definition of weak convergence in terms of test functions implies convergence in the Prokhorov metric, and we provide a bound on the rate. One important component of our proof is determining how exactly the constant  $C_f$  in (1.4) depends on  $f$  in the usual proofs of weak convergence [11, 9].

In Section 4 we show one consequence of our main result concerning re-embedding trajectories of the SDEs and of the numerical method in a new probability space. Two random vectors (such as  $X(T)$  and  $X_{T/\Delta t}$ ) may either not be close to each other on a realization-by-realization basis, or may be defined on completely different probability spaces. However, it is possible to define new random vectors  $Y$  and  $Z$  jointly on a new probability space such that  $Y$  has the same distribution as  $X(T)$  and  $Z$  has the same distribution as  $X_{T/\Delta t}$ . This construction is called a *re-embedding* of  $X(T)$  and  $X_{T/\Delta t}$  in a new probability space. After re-embedding the random vectors may be close together in a strong

sense and we can look at how quantities like  $\mathbb{E}|Y - Z|$  or  $\mathbb{P}(|Y - Z| > \alpha)$  behave as  $\Delta t$  varies. The Strassen–Dudley theorem says that if two random variables are close in the Prokhorov metric, then there is a re-embedding of them into another probability space for which they are close in probability. A bound on some higher moment of  $Y$  and  $Z$  then gives that  $\mathbb{E}|Y - Z|$  is small. Using our bound in the Prokhorov metric and the Strassen–Dudley theorem, we show that a method with the usual weak convergence of order  $p$  converges strongly after re-embedding with order  $\frac{p}{2p+3} - \varepsilon$  for any  $\varepsilon > 0$ . This is equivalent to proving a rate of convergence in the Wasserstein distance (see Section 4 for a definition). We also use re-embedding to establish rates for the convergence of expectations of test functions requiring only local Lipschitz continuity and polynomial growth.

Finally, in Section 5, we discuss the corresponding result for weak convergence of entire numerical trajectories on  $[0, T]$  to exact trajectories of the original system. Convergence in distribution follows directly from a result of Stroock and Varadhan, which we review. However, to the best of our knowledge there is no bound available for this rate of convergence for general weakly convergent methods. (Though see [15, 8] for some results for the strong Euler–Maruyama method.) Applying Skorohod’s theorem gives the corresponding strong convergence result for the trajectories embedded in another probability space, though again without a rate.

## 2 Metrics on Spaces of Random Elements

Consider a metric space  $(S, d)$  with metric  $d$ . A random element of  $S$  is a measurable function  $X : \Omega \rightarrow S$  where  $(\Omega, \mathcal{F}, \mathbb{P})$  is some probability space. For example, if  $S$  is  $\mathbb{R}^n$  with the metric  $d(x, y) = |x - y|$ , then random elements  $X$  are called random vectors. Even if two random elements  $X$  and  $Y$  of a metric space  $S$  are not close on a realization-by-realization basis, we may still wish to compare their distributions. So we define a metric on the space of random elements of  $S$ . Note that there are two distinct metrics involved:  $d$  which is a metric on the original space  $S$  and another which is a metric on the space of random elements of  $S$ . In this section, we first define the well-known Prokhorov metric  $\rho$ , which is defined for any underlying metric space. Then we introduce the metrics  $\beta_l$  for non-negative integers  $l$ , when the underlying metric space is  $\mathbb{R}^n$ . The latter are similar to the metric  $\beta$  of Fortet and Mourier [7]; see [3, Sec. 11.3].

For a set  $A \subset S$  we define  $A^\varepsilon$ ,  $\varepsilon > 0$ , the set of all points within distance  $\varepsilon$  of  $A$  by

$$A^\varepsilon = \{x \in S \mid \inf_{y \in A} d(x, y) \leq \varepsilon\}. \quad (2.1)$$

The Prokhorov metric is defined as follows.

DEFINITION 1 For random variables  $X$  and  $Y$  in  $S$

$$\rho(X, Y) := \inf\{\varepsilon \mid \mathbb{P}(X \in A) \leq \mathbb{P}(Y \in A^\varepsilon) + \varepsilon, \text{ for all } A \text{ closed}\}.$$

If we identify random elements of  $S$  that have the same distribution, then  $\rho$  is a metric on the set of random elements [3, p. 394]. If  $(S, d)$  is separable (as are all examples in this paper) random elements  $X_n$  converge in distribution to  $X$  if and only if  $\rho(X_n, X) \rightarrow 0$  [3, p. 395]. Note that  $\rho(X, Y) \leq 1$  always.

Here is the Strassen–Dudley Theorem as proven in [3, 12], used later in this section and in Section 4.

THEOREM 2.1 ([2, p. 73]) Let  $(S, d)$  be a separable metric space. If  $X$  and  $\bar{X}$  are random elements of  $S$  with  $\rho(X, \bar{X}) < \alpha$ , then there are random elements  $Y$  and  $Z$  of  $S$  defined on a common probability space such that  $Y$  has the same distribution as  $X$ ,  $Z$  has the same distribution as  $\bar{X}$  and

$$\mathbb{P}(|Y - Z| > \alpha) < \alpha.$$

□

We now define a class of metrics  $\beta_l$  on random vectors, that is, random elements of the metric space  $(\mathbb{R}^n, |\cdot|)$ . Let  $f : \mathbb{R}^k \rightarrow \mathbb{R}$ . Let  $\alpha$  be a vector of length  $k$  with non-negative integer components. Let  $|\alpha| := \sum_i \alpha_i$  and

$$D^\alpha f := \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \dots \partial x_n^{\alpha_n}}.$$

If we wish to emphasize the argument of  $f$  in our notation we use  $D_x^\alpha$  instead of  $D^\alpha$ . For  $l \geq 0$  and  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  let

$$\|f\|_l := \sum_{|\alpha| \leq l} \sup_{x \in \mathbb{R}^n} |D^\alpha f(x)|. \quad (2.2)$$

DEFINITION 2 For random vectors  $X$  and  $Y$  in  $\mathbb{R}^n$  and for  $l \geq 0$  we let

$$\beta_l(X, Y) = \sup_{\|f\|_l \leq 1} |\mathbb{E}f(X) - \mathbb{E}f(Y)|$$

It is straightforward to check that  $\beta_l$  is a metric on the space of random variables.

The following theorem is the main result in this section, and allows us to show in the next section that solutions generated by weak numerical methods converge in the Prokhorov metric.

THEOREM 2.2 For each  $l \geq 0$  there is a constant  $C > 0$  such that for any random vectors  $X, Y$  in  $\mathbb{R}^n$ ,

$$\rho(X, Y) \leq C\beta_l(X, Y)^{1/(1+l)}.$$

*Proof.* Here we closely follow [3, p. 396]. Consider any closed set  $K$  in  $\mathbb{R}^n$  and  $\varepsilon \in (0, 1]$ . Let  $K^\varepsilon$  be defined as in Equation (2.1). The lemma following this theorem shows that there is a smooth function  $f$  and a constant  $C$  such that depends on  $n$  but not on  $\varepsilon$  or  $K$  such that

$$\mathbf{1}_K(x) \leq f(x) \leq \mathbf{1}_{K^\varepsilon}(x) \quad \text{and} \quad \|f\|_l \leq C\varepsilon^{-l}.$$

Without loss of generality we assume that  $C \geq 1$ . We now use the function  $f$  to establish the required bound. For any random variables  $X$  and  $Y$

$$\begin{aligned} \mathbb{P}(Y \in K) &\leq \mathbb{E}f(Y) \\ &\leq \mathbb{E}f(X) + \|f\|_l \beta_l(X, Y) \\ &\leq \mathbb{P}(X \in K^\varepsilon) + C\varepsilon^{-l} \beta_l(X, Y). \end{aligned}$$

So for any  $\varepsilon \in (0, 1]$

$$\rho(X, Y) \leq \max(\varepsilon, C\varepsilon^{-l} \beta_l(X, Y)). \quad (2.3)$$

Now if  $\beta_l(X, Y) > 1$ , since  $C \geq 1$  and  $\rho(X, Y) \leq 1$  the result is immediately true. So we assume that  $\beta_l(X, Y) \leq 1$  and choose  $\varepsilon$  so that  $\beta_l(X, Y) = \varepsilon^{l+1}$ . Then  $\varepsilon \leq 1$  and (2.3) gives us  $\rho(X, Y) \leq \varepsilon \max(1, C)$ . So  $\rho(X, Y) \leq C\beta_l(X, Y)^{1/(1+l)}$  as required.  $\square$

LEMMA 2.1 For each closed set  $K \subset \mathbb{R}^n$  there is a parametrized family of functions  $f_\varepsilon(x)$  for  $\varepsilon \in (0, 1]$  such that

$$\mathbf{1}_K(x) \leq f_\varepsilon(x) \leq \mathbf{1}_{K^\varepsilon}(x), \quad (2.4)$$

and there is a constant  $C$  depending on  $n$  but not on  $\varepsilon$ ,  $K$ , or  $l$  such that

$$\|f_\varepsilon\|_l \leq C\varepsilon^{-l}. \quad (2.5)$$

*Proof.* We use the method of mollifiers; see, for example [6, p. 629]. Define  $\eta: \mathbb{R}^n \rightarrow \mathbb{R}$  by

$$\eta(x) := \begin{cases} D \exp\left(\frac{1}{\|x\|^2 - 1}\right) & \text{if } \|x\| < 1 \\ 0 & \text{if } \|x\| \geq 1, \end{cases}$$

where  $D > 0$  is selected so that  $\int_{\mathbb{R}^n} \eta(x) dx = 1$ . The mollifier  $\eta \in C^\infty$  is positive with support in the unit ball about the origin. Define

$$\eta_\varepsilon(x) := \frac{1}{\varepsilon^n} \eta\left(\frac{x}{\varepsilon}\right).$$

This function is in  $C^\infty$ , has support on the ball of radius  $\varepsilon$  about the origin, and also has integral 1.

Let  $K'$  be the closure of  $K^{\varepsilon/2}$  and let

$$f_\varepsilon(x) := \int_{\mathbb{R}^n} \eta_{\varepsilon/2}(y) \mathbf{1}_{K'}(x-y) dy.$$

The function  $f_\varepsilon$  is 1 on  $K$ , 0 on  $\mathbb{R}^n \setminus K^\varepsilon$ , and between zero and one elsewhere. So  $f_\varepsilon$  satisfies the condition of Equation (2.4).

In [6, p. 630] it is shown that

$$D^\alpha f_\varepsilon(x) = \int_{\mathbb{R}^n} D_x^\alpha \eta_{\varepsilon/2}(x-y) \mathbf{1}_{K'}(y) dy.$$

So

$$\begin{aligned} |D^\alpha f_\varepsilon(x)| &\leq \int_{\mathbb{R}^n} |D_x^\alpha \eta_{\varepsilon/2}(x-y)| dy = \int_{\mathbb{R}^n} (\varepsilon/2)^{-n} |D_x^\alpha \eta(2(x-y)/\varepsilon)| dy \\ &= (\varepsilon/2)^{-|\alpha|} \int_{\mathbb{R}^n} |D_z^\alpha \eta(z)| dz, \end{aligned}$$

where we have used the change of variables  $z = 2(y-x)/\varepsilon$ ,  $D_z^\alpha = (-\varepsilon/2)^{|\alpha|} D_x^\alpha$ . The integral in the last expression is finite and does not depend on  $K$  or  $\varepsilon$ . Summing over all  $\alpha$  with  $|\alpha| \leq l$  gives us

$$\|f\|_l \leq \sum_{|\alpha| \leq l} C_\alpha \varepsilon^{-|\alpha|} \leq C \varepsilon^{-l}$$

for some constants  $C_\alpha, C$ , for all  $\varepsilon \in (0, 1]$ . □

For completeness we include the following theorem which together with Theorem 2.2 shows that the metrics  $\rho$  and  $\beta_l$  induce the same topology on the space of random elements of  $\mathbb{R}^n$ . Thus, as for  $\rho$ ,  $\beta_l(X_n, X) \rightarrow 0$  if and only if  $X_n$  converges to  $X$  in distribution. This result is analogous to Theorem 11.6.5 in [3].

**THEOREM 2.3** For all  $l \geq 1$ , and random  $X$  and  $\bar{X}$  in  $\mathbb{R}^n$ , the metrics  $\rho$  and  $\beta_l$  satisfy

$$\beta_l(X, \bar{X}) \leq 2\rho(X, \bar{X}).$$

*Proof.* Let  $\rho(X, \bar{X}) = \varepsilon$ . Using Theorem 2.1, let  $Y$  and  $Z$  be random vectors on the same probability space with the same distributions as  $X$  and  $\bar{X}$  respectively such that  $\mathbb{P}(|Y - Z| > \varepsilon) < \varepsilon$ . Then

$$\begin{aligned} \beta_l(X, \bar{X}) &\leq \sup_{\|f\|_l \leq 1} \mathbb{E}|f(Y) - f(Z)| \\ &= \sup_{\|f\|_l \leq 1} \left\{ \mathbb{E}[\mathbf{1}_{|Y-Z| > \varepsilon} |f(Y) - f(Z)|] + \mathbb{E}[\mathbf{1}_{|Y-Z| \leq \varepsilon} |f(Y) - f(Z)|] \right\} \\ &\leq \sup_{\|f\|_l \leq 1} \left\{ \mathbb{E}[\mathbf{1}_{|Y-Z| > \varepsilon}] 2 \sup_x |f(x)| + \varepsilon \max_{y,z} \frac{|f(y) - f(z)|}{\|y - z\|} \right\} \\ &\leq \sup_{\|f\|_l \leq 1} \left\{ 2\varepsilon \sup_x |f(x)| + \varepsilon \sum_i \sup_x |D^i(f(x))| \right\} \\ &\leq 2\varepsilon, \end{aligned}$$

as required. □

### 3 Convergence of Numerical Methods

Here we prove our result on the convergence in the Prokhorov metric of numerical approximations to exact solutions of SDEs. We consider the system of Ito SDEs

$$dX = a(X)dt + \sum_{r=1}^q \sigma_r(X) dW_r(t), \tag{3.1}$$

where  $X(t) \in \mathbb{R}^n$ ,  $a: \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $\sigma_r: \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$  for all  $r$ . The  $W_r$ ,  $r = 1, \dots, q$  are mutually independent standard Wiener processes. We set the initial condition to be  $X(0) = x_0 \in \mathbb{R}^n$ .

To prove our convergence theorem we build on a weak convergence result from [11]. This result is expressed for a rather general method for the system (3.1):

$$X_{k+1} = X_k + \bar{\delta}(X_k, \Delta t; \xi_k), \quad (3.2)$$

with  $X_0 = x_0$ . Here  $\bar{\delta}$  is vector-valued function and  $\xi_k$ ,  $k = 0, 1, \dots$  is a sequence of independent random vectors. Usually we suppress the  $\xi_k$  from the notation and view  $\bar{\delta}(X_k, \Delta t)$  as a random vector. We denote its  $i$ th component by  $\bar{\delta}_i(X_k, \Delta t)$ . Here  $X_k$  is intended to be an approximation to  $X(k\Delta t)$ . In the following we use  $\delta$  to denote the increment of the true solution over a time interval: for the solution  $X$  to Equation (3.1) with  $X(0) = x$ , set

$$\delta(x, \Delta t) = X(\Delta t) - X(0).$$

Thus  $\delta(x, \Delta t)$ , like  $\bar{\delta}(x, \Delta t)$ , is a random vector. The  $i$ th component of  $\delta(x, \Delta t)$  is denoted  $\delta_i(x, \Delta t)$ .

Theorem 3.1 below gives a rate of convergence of  $\mathbb{E}f(X_k)$  to  $\mathbb{E}f(X(k\Delta t))$  in which the dependence of the constant on  $f$  is given. This result is an corollary of the result of [11, p. 100] or [9, p. 473] in which the dependence of the constant on  $f$  is not made explicit. Here, by making stronger assumptions on the coefficients  $a$  and  $\sigma_r$ , we show that the constant is linear in  $\|f\|_{2p+2}$  where  $p$  is the order of the method. (See Section 2 for a definition of  $\|\cdot\|_{2p+2}$ .)

**THEOREM 3.1** Let  $T > 0$  be fixed. Suppose that

- (a) the coefficients  $a$  and  $\sigma_r$  of the system of SDEs (3.1) have globally Lipschitz derivatives up to and including order  $2p+2$ ;
- (b) there is some scalar function  $K(x)$  with at most polynomial growth as  $x \rightarrow \infty$  such that

$$\left| \mathbb{E} \left( \prod_{j=1}^s \delta_j(x, \Delta t) - \prod_{j=1}^s \bar{\delta}_j(x, \Delta t) \right) \right| \leq K(x) \Delta t^{p+1},$$

for  $s = 1, \dots, 2p+1$  and

$$\mathbb{E} \prod_{j=1}^{2p+2} |\bar{\delta}_j(x, \Delta t)| \leq K(x) \Delta t^{p+1};$$

- (c) for all  $m \geq 1$  the expectations  $\mathbb{E}|X_k|^{2m}$  exist and are uniformly bounded with respect to  $\Delta t$  and  $k = 0, 1, \dots, \lfloor T/\Delta t \rfloor$ .
- (d) the function  $f(x)$  together with its partial derivatives of order up to and including  $2p+2$  are bounded. Then for all  $k\Delta t \in [0, T]$

$$|\mathbb{E}f(X(k\Delta t)) - \mathbb{E}f(X_k)| \leq C \|f\|_{2p+2} \Delta t^p.$$

The constant  $C$  depends on  $x$ ,  $a$ ,  $\sigma_r$  and  $T$  but not on  $f$  and  $\Delta t$ .

*Proof.* We define  $Y(x, t)$  to be  $X(k\Delta t)$  where  $X$  is the solution of (3.1) with initial condition  $X(t) = x$ ,  $t \leq k\Delta t$ . Define the function  $u(x, t)$  by

$$u(x, t) := \mathbb{E}f(Y(x, t)).$$

It follows from the proof of Theorem 2.1 in [11, p. 100] that

$$|\mathbb{E}f(X(k\Delta t)) - \mathbb{E}f(X_k)| \leq A \Delta t^p \max_{t \in [0, k\Delta t]} \|u(\cdot, t)\|_{2p+2}$$

for some constant  $A$  not depending on  $f$ .

In [10, p. 223] it is shown that if the coefficients of  $a$  and  $\sigma_r$  of the system of SDEs (3.1) have globally Lipschitz continuous derivatives up to order  $2p+2$  (condition (a)) then  $Y(t, x)$  has continuous derivatives with respect to  $x$  up to order  $2p+2$ , almost surely.

Let  $\partial_i$  denote differentiation of a function with respect to its  $i$ th argument. Formally, we can differentiate  $u$  with respect to  $x$  to obtain

$$\begin{aligned} \partial_i u &= \sum_a \mathbb{E} [(\partial_a f(Y)) (\partial_i Y_a)], \\ \partial_i \partial_j u &= \sum_{a,b} \mathbb{E} [(\partial_a \partial_b f(Y)) (\partial_i Y_a) (\partial_j Y_b)] + \sum_a \mathbb{E} [(\partial_a f(Y)) (\partial_i \partial_j Y_a)], \end{aligned}$$

and so forth, using the product and chain rules. To justify the formal differentiations, we need only observe that all multi-derivatives of  $f$  up to order  $2p+2$  are bounded, and remark that it follows from [10] that all moments of the derivatives of  $Y$  up to order  $2p+2$  are finite. The exchange of differentiation with expectation then follows in each case by Fubini's theorem [17, p. 222]. Applying the Cauchy–Schwarz inequality to each term gives

$$\sup_{x \in \mathbb{R}^d} |D^\alpha u| \leq \sum_i \left\{ \mathbb{E}(D^{\beta_i} f)^2 \right\}^{1/2} E_{\beta_i},$$

where  $\beta_i$  are a sequence of multi-indices with  $|\beta_i| \leq |\alpha|$  and  $E_{\beta_i}$  are some constants independent of  $f$ . So

$$\sup_{x \in \mathbb{R}^d} |D^\alpha u| \leq F_\alpha \|f\|_{|\alpha|},$$

for some constants  $F_\alpha$  independent of  $f$ . Summing this inequality over all  $\alpha$  with  $|\alpha| \leq 2p+2$  gives us the result.  $\square$  Putting Theorems 2.2 and 3.1 together gives us our conclusion for this section.

**COROLLARY 3.1** Let conditions (a), (b), (c) of Theorem 3.1 be satisfied. Then for some constant  $K$

$$\rho(X(k\Delta t), X_k) \leq K\Delta t^{p/(2p+3)},$$

for all  $k\Delta t \leq T$ .

*Proof.* By Theorem 3.1 and the definition of  $\beta_l$ , we have that

$$\beta_{2p+2}(X(k\Delta t), X_k) \leq C\Delta t^p.$$

Applying Theorem 2.2 with  $l = 2p+2$  then gives the result.  $\square$

## 4 Strong Convergence

In this section, we apply the Strassen–Dudley theorem to show that, after being re-embedded in another probability space, weakly converging methods for stochastic differential equations converge strongly with a reduced order. This re-embedding immediately gives a rate of convergence in the Wasserstein distance. As a corollary, we establish a rate of convergence of  $\mathbb{E}f(X_{T/\Delta t})$  to  $\mathbb{E}f(X(T))$  that requires only that  $f$  is locally Lipschitz with a polynomial growth condition.

**THEOREM 4.1** Let conditions (a), (b), (c) of Theorem 3.1 be satisfied. There is a probability space on which random vectors  $Y$  and  $Z$  are defined such that  $Y$  has the same distribution as  $X(T)$  and  $Z$  has the same distribution as  $X_{T/\Delta t}$  and, for any  $\varepsilon > 0$

$$\mathbb{E}|Y - Z| \leq C\Delta t^{\frac{p}{(2p+3)} - \varepsilon},$$

for some constant  $C$ , for all sufficiently small  $\Delta t$ .

*Proof.* Let  $\alpha = K\Delta t^{p/(2p+3)}$  where  $K$  is as in Corollary 3.1. Theorem 2.1 together with Corollary 3.1 establish the existence of the random vectors  $Y$  and  $Z$  with the correct distributions such that

$$\mathbb{E}(\mathbf{1}_{|Y-Z|>\alpha}) = \mathbb{P}(|Y-Z| > \alpha) < \alpha.$$

Choose  $\varepsilon > 0$ . Now, the conditions of Theorem 3.1 ensure that both  $Y$  and  $Z$  and hence  $|Y-Z|$  have finite moments of all orders independent of  $\Delta t$ . Choose real numbers  $q_1, q_2 > 1$  such that  $1/q_1 + 1/q_2 = 1$  and  $\frac{p}{(2p+3)} \frac{1}{q_2} \geq \frac{p}{(2p+3)} - \varepsilon$ . Using Hölder's inequality, we obtain

$$\begin{aligned} \mathbb{E}|Y - Z| &= \mathbb{E}[|Y - Z|\mathbf{1}_{|Y-Z|>\alpha}] + \mathbb{E}[|Y - Z|\mathbf{1}_{|Y-Z|\leq\alpha}] \\ &\leq (\mathbb{E}|Y - Z|^{q_1})^{1/q_1} (\mathbb{E}\mathbf{1}_{|Y-Z|>\alpha})^{1/q_2} + \alpha \\ &\leq (\mathbb{E}|Y - Z|^{q_1})^{1/q_1} \alpha^{1/q_2} + \alpha \\ &\leq E\alpha^{1/q_2} = D\Delta t^{\frac{p}{(2p+3)} \frac{1}{q_2}} \leq C\Delta t^{\frac{p}{(2p+3)} - \varepsilon}, \end{aligned}$$

for all sufficiently small  $\Delta t$ , as required.  $\square$

Applying this theorem to the case of weak Euler–Maruyama (see Equation (1.5)) with  $p = 1$  implies a strong rate of convergence after re-embedding of  $1/5 - \varepsilon$  for any  $\varepsilon > 0$ .

Now we express our result in terms of the Wasserstein distance, also known as the Wasserstein-1 distance, the Monge–Wasserstein distance [3, p. 420], or the Kantorovich–Rubinstein distance [16, p. 206]. To define this metric, let  $X$  and  $Y$  be random elements of  $S$  and let  $M(X, Y)$  be the set of all probability measures  $\mu$  on  $S \times S$  such that the marginals of  $\mu$  are the probability measures induced on  $S$  by  $X$  and  $Y$  respectively. Then the Wasserstein distance is

$$W(X, Y) = \inf_{\mu \in M(X, Y)} \mathbb{E}_\mu d(x, y) = \inf_{\mu \in M(X, Y)} \int d(x, y) d\mu(x, y).$$

where  $\mathbb{E}_\mu$  denotes expectation with respect to the measure  $\mu$  for  $(x, y) \in S \times S$ . In words, the Wasserstein distance is the minimal  $L^1$  distance between  $X$  and  $Y$  after re-embedding. Therefore Theorem 4.1 shows the following:

**COROLLARY 4.1** Let conditions (a), (b), (c) of Theorem 3.1 be satisfied. Then for any  $\varepsilon > 0$

$$W(X(T), X_{T/\Delta t}) \leq C\Delta t^{\frac{p}{2p+3} - \varepsilon},$$

for some constant  $C$ , for sufficiently small  $\Delta t$ .  $\square$

As another corollary to Theorem 4.1, we show that  $\mathbb{E}f(X_{T/\Delta t}) \rightarrow \mathbb{E}f(X(T))$  given some polynomial growth conditions on  $f$ , even when  $f$  is only locally Lipschitz. This result is like the usual weak convergence result [11, p. 100], but with a relaxed smoothness requirement on  $f$  and a reduced rate. Compare with Mikulevicius and Platen’s result [9, p. 460] or Bally and Talay’s result [1], both of which only apply to strong Euler–Maruyama (see Equation (1.2)).

**COROLLARY 4.2** Let conditions (a), (b), (c) of Theorem 3.1 be satisfied. Let  $f$  be locally Lipschitz with

$$|f(x) - f(y)| \leq L_R|x - y|,$$

whenever  $|x| \leq R$  and  $|y| \leq R$ , where

$$L_R \leq C(1 + R^\kappa),$$

for some constants  $C$  and  $\kappa$ . Then for any  $\varepsilon > 0$

$$|\mathbb{E}f(X_{T/\Delta t}) - \mathbb{E}f(X(T))| \leq K\Delta t^{\frac{p}{(2p+3)} - \varepsilon},$$

for some constant  $K$ , for sufficiently small  $\Delta t$ .

*Proof.* Let  $\alpha = K\Delta t^{p/(2p+3)}$  where  $K$  is as in Corollary 3.1. Let  $Y$  and  $Z$  be as in the proof of Theorem 4.1, so that  $Y$  has the same distribution as  $X(T)$ ,  $Z$  has the same distribution as  $X_{T/\Delta t}$ , and  $\mathbb{P}(|Y - Z| > \alpha) < \alpha$ . Let  $M = \max(|Y|, |Z|)$ . We immediately have

$$|\mathbb{E}f(X_{T/\Delta t}) - \mathbb{E}f(X(T))| = |\mathbb{E}f(Y) - \mathbb{E}f(Z)| \leq \mathbb{E}|f(Y) - f(Z)|.$$

Let  $R$  be an arbitrary radius which we shall fix later. We can split the quantity of interest into three terms:

$$\begin{aligned} \mathbb{E}|f(Y) - f(Z)| &\leq \mathbb{E}|f(Y) - f(Z)|\mathbf{1}_{|Y-Z| \leq \alpha, M \leq R} + \mathbb{E}|f(Y) - f(Z)|\mathbf{1}_{|Y-Z| > \alpha, M \leq R} \\ &\quad + \mathbb{E}|f(Y) - f(Z)|\mathbf{1}_{M > R} \\ &=: T_1 + T_2 + T_3. \end{aligned}$$

To bound the first term, note that

$$T_1 \leq L_R \alpha \leq \alpha C(1 + R^\kappa).$$

To bound the second and third terms, note that

$$\begin{aligned} |f(x)| &\leq |f(0)| + |f(x) - f(0)| \\ &\leq |f(0)| + L_{|x|}|x| \\ &\leq D(1 + |x|^{\kappa+1}), \end{aligned}$$

for all  $x$  for some constant  $D$ . Then

$$\begin{aligned} T_2 &\leq \mathbb{E}\{|f(Y)| + |f(Z)|\} \mathbf{1}_{|Y-Z|>\alpha, M \leq R} \\ &\leq \mathbb{E}\{D(1 + |Y|^{\kappa+1}) + D(1 + |Z|^{\kappa+1})\} \mathbf{1}_{|Y-Z|>\alpha, M \leq R} \\ &\leq \mathbb{E}2D(1 + R^{\kappa+1}) \mathbf{1}_{|Y-Z|>\alpha} \leq 2D(1 + R^{\kappa+1})\alpha. \end{aligned}$$

To bound the third term, we use the fact that all moments  $Y$  and  $Z$  are finite and the bounds on the moments of  $Z$  are independent of  $\Delta t$ . For any exponent  $q > \kappa + 1$  (which we shall choose later), we let  $m_q$  denote this bound so that  $\mathbb{E}|Y|^q \leq m_q$  and  $\mathbb{E}|Z|^q \leq m_q$ . These inequalities in turn imply that  $\mathbb{E}M^q \leq 2m_q$ . Then

$$\begin{aligned} T_3 &\leq \mathbb{E}\{|f(Y)| + |f(Z)|\} \mathbf{1}_{M>R} \\ &\leq \mathbb{E}2D(1 + M^{\kappa+1}) \mathbf{1}_{M>R} \\ &= 2D \mathbb{E} \frac{(1 + M^{\kappa+1})}{M^q} M^q \mathbf{1}_{M>R} \\ &\leq 2D \frac{(1 + R^{\kappa+1})}{R^q} 2m_q \leq C_q R^{\kappa+1-q}, \end{aligned}$$

for all sufficiently large  $R$ . Putting these three bounds together gives

$$\mathbb{E}|f(Y) - f(Z)| \leq \alpha ER^{\kappa+1} + C_q R^{\kappa+1-q},$$

for sufficiently large  $R$ . We get to choose both  $R$  and  $q$ . For any  $q \geq \kappa$ , if we choose  $R = \alpha^{-1/q}$  we obtain

$$\mathbb{E}|f(Y) - f(Z)| \leq E\alpha^{1-(\kappa+1)/q} + C_q \alpha^{1-(\kappa+1)/q}.$$

Choosing  $q$  large enough gives the desired result.  $\square$

## 5 Pathwise Convergence in Distribution

The results in previous sections concern the pointwise weak convergence of numerical methods for SDEs, that is, convergence at each point in time  $t$ . A stronger result is that entire trajectories generated by the numerical method weakly converge to those of the system of SDEs. Stroock and Varadhan prove pathwise convergence in distribution of numerical methods in great generality in [13] but they do not provide a rate. Here we review their result and apply a re-embedding theorem to establish the corresponding strong result for embedded random paths. Since no rate appears to be established for Stroock and Varadhan's result, we do not phrase results in terms of the Prokhorov metric and instead just consider convergence in distribution. Moreover, we can use Skorohod's theorem for re-embedding rather than the Strassen–Dudley theorem. The latter gives precise rates of strong convergence but the former allows one to construct a whole sequence of random paths and their limit on one probability space.

First we review the definition of convergence in distribution in  $C^n[0, T]$ , the space of continuous,  $\mathbb{R}^n$ -valued functions on  $[0, T]$ , [2]. For any fixed  $T$  and initial condition  $x_0$  the solution to the system of SDEs (3.1) gives a random element of  $C^n[0, T]$  which we denote by  $X$ . For the same  $T$  and initial conditions the numerical method with step-size  $\Delta t$  gives a sequence  $X_k$ ,  $k = 0, 1, \dots$ . We define the linear interpolant  $\bar{X}_{\Delta t}$  of the values  $X_k$  by

$$\bar{X}_{\Delta t}(t) = X_{\lfloor t/\Delta t \rfloor} + (t/\Delta t - \lfloor t/\Delta t \rfloor)(X_{\lfloor t/\Delta t \rfloor + 1} - X_{\lfloor t/\Delta t \rfloor}),$$

for  $t \in [0, T]$ . Thus  $\bar{X}_{\Delta t}$  is a random element of  $C^n[0, T]$ . If we equip  $C^n[0, T]$  with the norm  $\|\cdot\|_{\infty}$  we obtain a metric space with metric

$$d(x, y) := \|x - y\|_{\infty} = \sup_{t \in [0, T]} |x(t) - y(t)|,$$

for  $x, y \in C^n[0, T]$ . We say that  $\bar{X}_{\Delta t}$  converges in distribution to  $X$  if for all bounded continuous functions  $f: C^n[0, T] \rightarrow \mathbb{R}$

$$\mathbb{E}f(\bar{X}_{\Delta t}) \rightarrow \mathbb{E}f(X),$$

as  $\Delta t \rightarrow 0$ . Stroock and Varadhan's result gives conditions on the original system of SDEs and the numerical method under which  $\bar{X}_{\Delta t}$  converges in distribution to  $X$ .

The system of SDEs (3.1) we consider is determined by its coefficients  $a$  and  $\sigma_r$ . We define the matrix  $b$  from  $\sigma_r$  by

$$b_{ij}(x) = \sum_r \sigma_{r,i}(x) \sigma_{r,j}(x), \quad (5.1)$$

for  $i, j = 1, \dots, n$ . Recall that the increment of the numerical method (3.2) starting from  $x$  is denoted  $\bar{\delta}(x, \Delta t)$ . For our numerical method we define corresponding coefficients  $a_{\Delta t}$  and  $b_{\Delta t}$  by

$$a_{\Delta t, i}(x) = \frac{1}{\Delta t} \mathbb{E} \bar{\delta}_i(x, \Delta t) \mathbf{1}_{|\bar{\delta}(x, \Delta t)| \leq 1},$$

and

$$b_{\Delta t, ij}(x) = \frac{1}{\Delta t} \mathbb{E} \bar{\delta}_i(x, \Delta t) \bar{\delta}_j(x, \Delta t) \mathbf{1}_{|\bar{\delta}(x, \Delta t)| \leq 1}.$$

Finally, we define  $\Gamma_{\Delta t}^\varepsilon$  by

$$\Gamma_{\Delta t}^\varepsilon(x) = \frac{1}{\Delta t} \mathbb{P}(|\bar{\delta}(x, \Delta t)| \geq \varepsilon).$$

In the following let  $\|\cdot\|$  denote any norm on the space of  $n \times n$  matrices.

**THEOREM 5.1** Suppose that

- (a) the coefficients  $a$  and  $\sigma_r$  of the SDEs (3.1) are locally Lipschitz continuous;
- (b) there is a constant  $C$  such that for all  $x \in \mathbb{R}^n$

$$x^T a(x) \leq C(1 + |x|^2) \quad \text{and} \quad \|b(x)\| \leq C(1 + |x|^2);$$

- (c) for all  $R > 0$

$$\limsup_{\Delta t \rightarrow 0} \sup_{|x| \leq R} |a_{\Delta t}(x) - a(x)| = 0, \quad \limsup_{\Delta t \rightarrow 0} \sup_{|x| \leq R} \|b_{\Delta t}(x) - b(x)\| = 0,$$

and

$$\limsup_{\Delta t \rightarrow 0} \sup_{|x| \leq R} \Gamma_{\Delta t}^\varepsilon(x) = 0,$$

for all  $\varepsilon > 0$ .

Then  $\bar{X}_{\Delta t}$  converges in distribution to  $X$  in  $C^n[0, T]$ .

*Proof.* This result is Theorem 11.2.3 of [13], using Theorems 5.3.1 and 5.3.2 of [5] to obtain the well-posedness of the martingale problem.  $\square$

We remark that condition (b) of Theorem 5.1 is stronger than necessary. The result holds with (b) replaced by the weaker condition that the martingale problem with coefficients  $a$  and  $b$  has a unique solution.

To give a feeling for the power of this result, here are some examples of functions  $f$  for which it applies. Firstly, we can recover the simpler pointwise results (without rates) if we let  $f: C^n[0, T] \rightarrow \mathbb{R}$  be defined by  $f(X) := g(X(t))$ , for some time  $t \in [0, T]$ . If  $g: \mathbb{R} \rightarrow \mathbb{R}$  is continuous and bounded then  $f$  is continuous and bounded and the previous theorem tells us that  $\mathbb{E}g(\bar{X}_{\Delta t}(t)) \rightarrow \mathbb{E}g(X(t))$ . More generally, we can choose  $f$  to depend on  $X$  through a number of points  $t_1, \dots, t_k \in [0, T]$ . Suppose  $g: \mathbb{R}^k \rightarrow \mathbb{R}$  is bounded and continuous. Letting  $f(X) := g(X(t_1), \dots, X(t_k))$  gives us

$$\mathbb{E}g(\bar{X}_{\Delta t}(t_1), \dots, \bar{X}_{\Delta t}(t_k)) \rightarrow \mathbb{E}g(X(t_1), \dots, X(t_k)),$$

as  $\Delta t \rightarrow 0$ . Even more generally, we can look at functions that do not depend on any finite number of times  $t_i$ . For example, let  $g: \mathbb{R} \rightarrow \mathbb{R}$  be bounded and continuous and define  $f(X) := g(\max_{t \in [0, T]} X(t))$ . The previous theorem tells us that  $\mathbb{E}f(\bar{X}_{\Delta t}) \rightarrow \mathbb{E}f(X)$  as  $\Delta t \rightarrow 0$ .

These example all rely on  $f$  being continuous and bounded. However, these assumptions can be weakened considerably in some cases. One result of this type is that if  $f$  is bounded and measurable and the probability of  $X$  falling in the set of discontinuities of  $f$  is zero, then it still holds that  $\mathbb{E}f(\bar{X}_{\Delta t}) \rightarrow \mathbb{E}f(X)$  [2, p. 21]. As an example, let  $f(X) = \mathbf{1}_{X(t_1) \in A_1} \mathbf{1}_{X(t_2) \in A_2}$ . Then if we can show  $\mathbb{P}(X(t_1) \in \partial A_1, X(t_2) \in \partial A_2) = 0$ , it follows that

$$\mathbb{P}(\bar{X}_{\Delta t}(t_1) \in A_1, \bar{X}_{\Delta t}(t_2) \in A_2) \rightarrow \mathbb{P}(X(t_1) \in A_1, X(t_2) \in A_2)$$

as  $\Delta t \rightarrow 0$ . Similarly, the condition on the boundedness of  $f$  can be relaxed given some a priori knowledge on the distributions of  $X_n$  and  $X$  [2, p. 31].

We now put the conditions of Theorem 5.1 in terms more familiar in the numerical analysis of SDEs.

**THEOREM 5.2** Suppose the coefficients  $a$  and  $\sigma$  of (3.1) are locally Lipschitz continuous and condition (b) of Theorem 5.1 is satisfied. Suppose that the following limits hold as  $\Delta t \rightarrow 0$  uniformly on bounded subsets of  $\mathbb{R}^n$ :

$$\frac{1}{\Delta t} |\mathbb{E} \delta_i - \mathbb{E} \bar{\delta}_i| \rightarrow 0, \quad (5.2)$$

$$\frac{1}{\Delta t} |\mathbb{E} \delta_i \delta_j - \mathbb{E} \bar{\delta}_i \bar{\delta}_j| \rightarrow 0, \quad (5.3)$$

and

$$\frac{1}{\Delta t} \mathbb{E} |\bar{\delta}_i \bar{\delta}_j \bar{\delta}_k| \rightarrow 0, \quad (5.4)$$

for all  $i, j, k = 1, \dots, n$ . Then  $\bar{X}_{\Delta t}$  converges in distribution to  $X$  in  $C^n[0, T]$ .

*Proof.* We prove this result by showing that (5.2), (5.3), and (5.4) imply condition (c) of Theorem 5.1. In the following we suppress the arguments of  $\delta(x, \Delta t)$  and  $\bar{\delta}(x, \Delta t)$ . Fix a bounded set in  $\mathbb{R}^n$ . We start by proving that (5.4) implies  $\Gamma_{\Delta t}^\varepsilon(x) \rightarrow 0$  uniformly. Using Chebyshev's inequality:

$$\Gamma_{\Delta t}^\varepsilon(x) = \frac{1}{\Delta t} \mathbb{P}(|\bar{\delta}(x, \Delta t)| \geq \varepsilon) \leq \sum_{i=1}^n \frac{1}{\Delta t} \mathbb{P}(|\bar{\delta}_i| \geq \varepsilon/n) = \sum_{i=1}^n \frac{1}{\Delta t} \mathbb{E} |\bar{\delta}_i|^3 (n/\varepsilon)^3,$$

which goes to zero as  $\Delta t \rightarrow 0$ .

Next we prove that  $\|b_{\Delta t}(x) - b(x)\|$  goes to zero uniformly as  $\Delta t \rightarrow 0$ . The argument for  $|a_{\Delta t}(x) - a(x)|$  is analogous so we omit it. Note that it suffices to show that  $|b_{\Delta t, i, j} - b_{i, j}| \rightarrow 0$  uniformly on the chosen set for all  $i, j$ . For each  $i$  and  $j$

$$\begin{aligned} |b_{\Delta t, i, j} - b_{i, j}| &= \left| \frac{1}{\Delta t} \mathbb{E} \bar{\delta}_i \bar{\delta}_j \mathbf{1}_{|\bar{\delta}| \leq 1} - b_{i, j} \right| \\ &\leq \frac{1}{\Delta t} |\mathbb{E} \bar{\delta}_i \bar{\delta}_j \mathbf{1}_{|\bar{\delta}| \leq 1} - \mathbb{E} \bar{\delta}_i \bar{\delta}_j| + \frac{1}{\Delta t} |\mathbb{E} \bar{\delta}_i \bar{\delta}_j - \mathbb{E} \delta_i \delta_j| + \left| \frac{1}{\Delta t} \mathbb{E} \delta_i \delta_j - b_{i, j} \right| \end{aligned}$$

The second term goes to zero from (5.3). The third term goes to zero by general properties of SDEs. The first term is equal to

$$\frac{1}{\Delta t} |\mathbb{E} \bar{\delta}_i \bar{\delta}_j \mathbf{1}_{|\bar{\delta}| > 1}| \leq \sum_{k=1}^n \frac{1}{\Delta t} \mathbb{E} |\bar{\delta}_i \bar{\delta}_j| \mathbf{1}_{|\bar{\delta}_k| > 1/n} \leq n \sum_{k=1}^n \frac{1}{\Delta t} \mathbb{E} |\bar{\delta}_i \bar{\delta}_j \bar{\delta}_k|,$$

which goes to zero by (5.4). □

In order to show a strong re-embedding type of result, we use the following theorem, sometimes called Skorohod's Theorem.

**THEOREM 5.3** (See [2, p. 70].) Let  $S$  be a separable metric space with metric  $d$ . Suppose that  $X_n, n \geq 1$  and  $X$  are random variables taking values in  $S$ , and that  $X_n$  converges in distribution to  $X$ . Then there are random variables  $Y_n, n \geq 1$  and  $Y$  all defined on the same probability space such that the distribution of  $Y_n$  is the same as  $X_n$  for all  $n$ , the distribution of  $Y$  is the same as  $X$ , and  $Y_n$  converges to  $Y$  almost surely.

Applying this theorem to  $\bar{X}_{\Delta t}$ , and  $X$  gives the following result.

**THEOREM 5.4** Let either the conditions of Theorem 5.1 or the conditions of Theorem 5.2 hold. Let  $\Delta t_n$  be a sequence of positive step-sizes converging to 0. Then there are random elements  $Y_n$  and  $Y$  of  $C^n[0, T]$  such that  $Y_n$  has the same distribution as  $\bar{X}_{\Delta t_n}$ ,  $Y$  has the same distribution as  $X$ , and  $Y_n \rightarrow Y$  in  $C^n[0, T]$  almost surely. Thus, almost surely,

$$\lim_{n \rightarrow \infty} \sup_{t \in [0, T]} |Y_n(t) - Y(t)| = 0.$$



**Acknowledgments.** BC was funded by a Canadian National Science and Engineering Research Council (NSERC) postdoctoral fellowship. YS was supported by an Institut des Sciences Mathématiques undergraduate research fellowship. PFT was funded by an NSERC Discovery Grant. The authors would like to thank the reviewers for comments that greatly improved the quality of the manuscript.

## References

- [1] V. Bally and D. Talay. The law of the Euler scheme for stochastic differential equations. I. Convergence rate of the distribution function. *Probab. Theory Related Fields*, 104(1):43–60, 1996.
- [2] P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. A Wiley-Interscience Publication.
- [3] R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.
- [4] R. Durrett. *Probability: theory and examples*. Duxbury Press, Belmont, CA, second edition, 1996.
- [5] R. Durrett. *Stochastic calculus*. Probability and Stochastics Series. CRC Press, Boca Raton, FL, 1996.
- [6] L. C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1998.
- [7] R. Fortet and E. Mourier. Convergence de la répartition empirique vers la répartition théorique. *Ann. Sci. Ecole Norm. Sup. (3)*, 70:267–285, 1953.
- [8] S. Kanagawa. On the rate of convergence for Maruyama’s approximate solutions of stochastic differential equations. *Yokohama Math. J.*, 36(1):79–86, 1988.
- [9] P. E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1992.
- [10] H. Kunita. Stochastic differential equations and stochastic flows of diffeomorphisms. In *École d’été de probabilités de Saint-Flour, XII—1982*, volume 1097 of *Lecture Notes in Math.*, pages 143–303. Springer, Berlin, 1984.
- [11] G. N. Milstein and M. V. Tretyakov. *Stochastic numerics for mathematical physics*. Scientific Computation. Springer-Verlag, Berlin, 2004.
- [12] V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Statist.*, 36:423–439, 1965.
- [13] D. W. Stroock and S. R. S. Varadhan. *Multidimensional diffusion processes*, volume 233 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1979.
- [14] D. Talay. Discrétisation d’une équation différentielle stochastique et calcul approché d’espérances de fonctionnelles de la solution. *RAIRO Modél. Math. Anal. Numér.*, 20(1):141–179, 1986.
- [15] A. D. Vent-Tsel’, S. A. Gladyshev, and G. N. Mil’shteyn. Piecewise constant approximation for the monte-carlo calculation of wiener integrals. *Theory of Probability and its Applications*, 29(4):744–752, 1985.
- [16] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.
- [17] D. Williams. *Probability with martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991.