Approximations of the Wong–Zakai type for stochastic differential equations in M-type 2 Banach spaces with applications to loop spaces

Zdzisław Brzeźniak[1] and Andrew Carroll[2]

[1] Department of Mathematics, The University of Hull, Hull HU6 7RX, UK
e-mail: Z.Brzezniak@maths.hull.ac.uk

[2] St Bede's School, Hailsham, East Sussex, BN27 3QH, UK
e-mail: andrew.carroll@stbedesschool.org

Introduction

In the celebrated paper [51] Wong and Zakai investigated the convergence of certain ordinary differential equations (ODEs) which, in particular, involved piece-wise linear approximations of a one-dimensional Wiener process. They showed that the solutions to these ODEs converge almost surely to a solution of a certain Stratonovitch stochastic differential equation (SDE) and *not* an Itô SDE. This result initiated considerable (and varied) research into the approximation of solutions to SDEs. The multi-dimensional Wiener process[3] was first studied by Clark in his PhD thesis [16] (see also [17]), McShane in [42] and by Stroock–Varadhan in [48]. Let us mention also papers by Malliavin [35] and Ikeda–Watanabe [28], where the authors consider and compare approximations involving different regularisations of the Wiener process. Elworthy in [24] announced (and in [25] proved with full details ascribed to Dowell [23]) a Hilbert space version of the piece-wise linear approximation and applied his result to the approximation of stochastic flows on manifolds. Moulinier, [40], studies continuity properties and rates of convergence, whereas Mackevicius, [34], considers a more general case of approximations of SDEs driven by semimartingales. One should not forget to mention important works [5] by Bismut and [38] by Malliavin. Doss [22] and Sussman [49] independently studied the question of continuity of the solutions to Stratonovitch equations with respect to an individual path of Wiener process. Their very interesting results were, however, restricted (as in [51]) to a one dimensional BM or, in the case of [22], to a multi-dimensional BM but under commutativity assumptions on

[3] Let us point out here that the convergence of stochatic integrals with respect to multidimensional Wiener process has also been investigated by Wong and Zakai in [52].

J. Azéma et al. (Eds.): LNM 1832, pp. 251–289, 2003.
© Springer-Verlag Berlin Heidelberg 2003

the vector fields. The continuity question was later investigated by T. Lyons and coworkers using an ingenious method first described in [33].

Furthermore, there has also been much work done on the approximation of solutions to stochastic partial differential equations (SPDEs). We refer the reader to the paper by Brzeźniak and Flandoli, [11], which contains more recent results on parabolic and hyperbolic evolution equations. Moreover, in the introduction to this paper they provide a brief account (with references) of the work done in this area by various authors such as Gyöngy, Kunita and Pardoux.

In this paper we prove results which are in the same spirit of the results of Wong–Zakai. However, not only is our Wiener process infinite dimensional but we are also concerned with SDEs taking values in Banach spaces, which is one of the novelties of our work. It has long been known that there are problems even in defining a stochastic Itô integral for general Banach spaces. However, a theory of stochastic integration has been developed for M-type 2 or 2-uniformly smooth Banach spaces, see[4] Neidhard [44], Belopolskaya and Daletskii [3], Dettweiler [20] and references therein. SDEs and SPDEs in M-type 2 Banach spaces have been studied by both the authors and Elworthy, see [6], [9], [7], [14], with more recent work done by Brzeźniak and Elworthy concerning SDEs on Banach manifolds which are modeled on M-type 2 Banach spaces, see [10].

The results in the first part of the paper are an extension of results given in the thesis by Dowell, [23] (on which the earlier mentioned proof in [25] is based), who considers approximations of SDEs in infinite dimensional separable Hilbert spaces. The extension to the Banach space case is non trivial. Indeed, Dowell was familiar with the theory of stochastic integration in 2-uniformly smooth Banach spaces, but was unable to extend his results to this case. When considering Stratonovitch equations in Banach spaces the main difficulty lies in dealing with the 'trace' map, see the discussion in Section 2. Although our problem is technically more difficult, we actually prove stronger results than Dowell. Under the assumption that the coefficients are globally Lipschitz and bounded, Dowell proves convergence in L^2 and convergence in the space of continuous functions in probability. However, we prove convergence in the space of continuous functions in L^p, $p \geqslant 2$, and for $p \geqslant 2$ we prove estimates which give a rate of convergence. This in turn proves almost sure convergence of the approximated ODEs to the Stratonovitch SDE, analogous to the original result of Wong–Zakai. These results first appeared in the PhD thesis by Carroll, [14].

The second part of the paper is concerned with certain applications to SDEs on loop spaces. However, the assumptions on the coefficients described

[4] In fact, the authors have recently become aware of an earlier paper [27] by Hoffmann-Jorgensen and Pisier, where such an integral was constructed (although only for 1-dimensional square integrable martingales and for deterministic integrands).

above are too strong for the applications we have in mind. Making use of our earlier results we prove convergence in the space of continuous functions in probability in the case of equations whose coefficients are locally Lipschitz and of linear growth. The importance of these new results is that they can be applied to the recent results of Brzeźniak and Elworthy concerning solutions to SDEs on loop spaces. We prove convergence in probability of equations taking values in a Banach manifold $\mathcal{M}$ which is modeled on an M-type 2 Banach space. In particular, $\mathcal{M}$ is a certain Sobolev–Slobodetskii space of loops on a compact, finite dimensional manifold M. There are certain implications of this result when considering SDEs on (both finite dimensional and Banach) manifolds. In particular, it emphases the need for using Stratonovich integrals as opposed to Itô integrals when dealing with SDEs on manifolds.

The layout of the paper is as follows. In Section 1 we introduce the relevant definitions and results concerning stochastic integration in M-type 2 Banach spaces. This is to make the paper self-contained. In Section 2 we consider SDEs whose coefficients are globally Lipschitz and bounded. We compare our results with those of Dowell and discuss the technical difficulties that need to be overcome when dealing with the Banach space case. In Section 3 we consider SDEs whose coefficients are locally Lipschitz and of linear growth. The result of Section 3 is then applied to a class of SDEs on loops, as studied by Brzeźniak and Elworthy in [10] and to diffeomorphism groups as studied in [9].

At the end of our Introduction let us mention one important consequence of our results: '*the transfer principle*'. By this we mean a general statement of the form: *Whatever is true for ordinary differential equations remains true for the stochastic differential equations in the Stratonovitch form.* As an example of such a principle, we prove invariance of a manifold M under solutions to Stratonovitch SDEs in the case when the vector fields are tangent to M, see Theorem 5. See e.g. [18] for a finite dimensional case.

1 Stochastic Integration in M-type 2 Banach Spaces

The following definition is fundamental for our work.

Definition 1. *A Banach space X is called M-type 2 if and only if there exists a constant $C(X) > 0$ such that for any X-valued martingale $\{M_k\}$ the following inequality holds*

$$\sup_k \mathbb{E}\big[|M_k|^2\big] \leqslant C(X) \sum_k \mathbb{E}\big[|M_k - M_{k-1}|^2\big]. \tag{1}$$

Any Hilbert space is an M-type 2 Banach space. In such a case we then have equality in (1) with $C(X) = 1$. The Lebesgue Function spaces L^p, $p > 2$, are examples of M-type 2 Banach spaces which are not Hilbert spaces.

The theory of stochastic integration in infinite dimensional Hilbert spaces has been developed and is well understood. However, for general separable Banach spaces there are difficulties in defining a meaningful Itô integral. In an unpublished thesis by Neidhardt, [44], a theory of stochastic integration was developed for a certain class of Banach spaces known as 2-uniformly smooth Banach spaces. A Banach space X is said to be 2-uniformly smooth if and only if for each x, $y \in X$

$$\frac{1}{2}\big(|x+y|_X^2 + |x-y|_X^2\big) \leqslant |x|_X^2 + A|y|_X^2, \tag{2}$$

for some constant $A > 0$. If X is a Hilbert space then equality holds in (2) with $A = 1$, i.e., the norm $|\,.\,|_X$ satisfies the parallelogram law. Independently of Neidhardt, similar work on stochastic integrals was carried out by Dettweiler, see [20] and references therein. It is known, see [45], that a Banach space is 2-uniformly smooth if and only if it is M-type 2. Either of the above two inequalities make it possible to define a meaningful Itô integral for this class of Banach spaces. However, the M-type 2 inequality (1) will prove to be the most useful for our needs. We briefly outline the construction of the Itô integral in M-type 2 Banach spaces and refer the reader to [10] and [14] for a more detailed summary and further references.

Definition 2. *For separable Hilbert and Banach spaces H and X we set*

$$R(H,X) := \{T : H \to X : T \in L(H,X) \text{ and } T \text{ is } \gamma\text{-radonifying}\}, \tag{3}$$

where $L(H,X)$ denotes the Banach space of bounded linear operators between H and X. By T being γ-radonifying we mean that the image $T(\gamma_H) := \gamma_H \circ T^{-1}$ of the canonical finitely additive Gaussian measure γ_H on H is σ-additive on the algebra of cylindrical sets in X.

Remark 1. The algebra of cylindrical sets in X generates the Borel σ-algebra, $\mathcal{B}(X)$ on X, see [31]. Thus $T(\gamma_H)$ extends to a Borel measure on $\mathcal{B}(X)$ which we denote by ν_T. In particular, ν_T is a Gaussian measure on $\mathcal{B}(X)$, i.e., for each $\lambda \in X^*$ (the dual of X), the image measure $\lambda(\nu_T)$ is a Gaussian measure on $\mathcal{B}(\mathbb{R})$. The covariance operator of ν_T equals $TT^* : E^* \to E$.

For $T \in R(H,X)$ we put

$$\|T\|^2_{R(H,X)} := \int_X |x|^2 \,\mathrm{d}\nu_T(x). \tag{4}$$

As ν_T is Gaussian, then by the Fernique–Landau–Shepp Theorem, see [31], $\|T\|_{R(H,X)}$ is finite. Furthermore, see [44], $R(H,X)$ is a separable Banach space endowed with the norm (4).

Definition 3. *Let E be a separable Banach space. We say that $i : H \hookrightarrow E$ is an Abstract Wiener Space, AWS, if and only if i is a linear, one-to-one map and $i \in R(H,E)$. If $i : H \hookrightarrow E$ is an AWS, then the Gaussian measure ν_i on E will be denoted by μ and called the canonical Gaussian measure on E.*

Remark 2. Many authors require $i(H)$ to be dense in E in the definition of an AWS. This is an unnecessary restriction for us. In fact, Sato, [46], proved that given a separable Banach space with Gaussian measure μ, then there always exists a Hilbert subspace $H \subset E$ such that $i : H \hookrightarrow E$ is an AWS, with $\mu = \nu_i$, where i is the inclusion mapping. The imbedding i is not dense in general.

Remark 3. The Hilbert space H appearing in the above definition is often referred to as the reproducing kernel Hilbert space, RKHS, of (E, μ).

Suppose that a triple $(\Omega, \mathcal{F}, \mathbb{P})$ is a complete probability space and let $i : H \hookrightarrow E$ be an AWS. Let $w(t)$, $t \geqslant 0$, denote the corresponding E-valued Wiener process, i.e., a continuous process on E such that:

(i) $w(0) = 0$ a.s.;
(ii) the law of the random function $t^{-1/2}w(t) : \Omega \to E$ equals μ, for any $t > 0$;
(iii) if $\mathcal{F}_s$ is the σ-algebra generated by $w(r)$, $r \in [0, s]$, then $w(t) - w(s)$ is independent of $\mathcal{F}_s$ for any $t \geqslant s \geqslant 0$.

Remark 4. In view of (ii) it is not difficult to show that for $p \geqslant 0$,

$$m_p := \mathbb{E}\left[\left|\frac{w(t) - w(s)}{(t-s)^{1/2}}\right|_E^p\right] = \int_E |z|_E^p \,\mathrm{d}\mu(z). \tag{5}$$

Furthermore by the Fernique–Landau–Shepp Theorem, see [31], $m_p < \infty$ for each $p \geqslant 0$.

Let X be an M-type 2 Banach space and $T \in (0, \infty)$. For $p \geqslant 1$, let $M^p(0, T; L(E, X))$ be the space of (equivalence classes of) progressively measurable functions $\xi : [0, T] \times \Omega \to L(E, X)$ which satisfy

$$\mathbb{E}\left[\int_0^T |\xi(t)|_{L(E,X)}^p \,\mathrm{d}t\right] < \infty$$

(with an analogous definition for the space $M^p(0, T; R(H, X))$).

Let $M^p_{\mathrm{step}}(0, T; L(E, X))$ be the subspace of those $\xi \in M^p(0, T; L(E, X))$ for which there exists a partition $0 = t_0 < t_1 < \cdots < t_n = T$ such that $\xi(t) = \xi(t_k)$ for $t \in [t_k, t_{k+1})$, $0 \leqslant k \leqslant n - 1$, $k \in \mathbb{N}$.

For $\xi \in M^2_{\mathrm{step}}(0, T; L(E, X))$ define a measurable map $I(\xi) : \Omega \to X$ by

$$I(\xi) := \sum_{j=1}^{n-1} \xi(t_k)\big(w(t_{k+1}) - w(t_k)\big). \tag{6}$$

The following lemma is crucial for the successful construction of the Itô integral.

Lemma 1. *Suppose $i : H \to E$ is an AWS with canonical E-valued Wiener process $w(t)$, $t \geqslant 0$, X is an M-type 2 Banach space and $T \in (0,\infty)$. Then for $\xi \in M^p_{\text{step}}(0,T;L(E,X))$, $I(\xi) \in L^2(\Omega;X)$, $\mathbb{E}[I(\xi)] = 0$ and*

$$\mathbb{E}\big[|I(\xi)|_X^2\big] \leqslant C \int_0^T \mathbb{E}\big[\|\xi(t)\circ i\|^2_{R(H,X)}\big]\,\mathrm{d}t. \tag{7}$$

Remark 5. Lemma 1 may be proved using either the inequality (1) or the inequality (2), along with the fact that $L(E,X)$ is contained in $R(H,X)$ via the continuous map

$$L(E,X) \ni \xi \longmapsto \xi \circ i \in R(H,X).$$

Remark 6. In the case when X is a Hilbert space (7) reads

$$\mathbb{E}\big[|I(\xi)|_X^2\big] = \mathbb{E}\bigg[\int_0^T \|\xi(t)\circ i\|^2_{R(H,X)}\,\mathrm{d}t\bigg],$$

which, of course, is the well-known Itô Isometry. The existence of the Itô Isometry is due to the 'nice' geometrical properties of the Hilbert space, i.e., the existence of an inner product. In general Banach spaces we lose the notion of 'geometry' and this is where the difficulty lies when one wishes to construct an Itô Integral. Although we do not have the Itô Isometry, the inequality (7) is enough to ensure that we can control the 'size' of the random variable $I(\xi)$ given by (6).

The fundamental property of the map I is that it extends uniquely to a bounded linear map from $M^2(0,T;R(H,X))$ into $L^2(\Omega;X)$. This is a consequence of (7) and the fact, proven in [44], that $M^2_{\text{step}}(0,T;L(E,X))$ is dense in $M^2(0,T;R(H,X))$. For $\xi \in M^2(0,T;R(H,X))$, the value of this extension will be denoted by $\int_0^T \xi(t)\,\mathrm{d}w(t)$. Furthermore, we have

Theorem 1. *Suppose $i : H \to E$ is an AWS with corresponding E-valued Wiener process $w(t)$, $t \geqslant 0$, and X is an M-type 2 Banach space. Assume that for $T > 0$, $\xi \in M^2(0,T;R(H,X))$ and let $I(r) := \int_0^r \xi(t)\,\mathrm{d}w(t)$ for $r > 0$. Then, $I(r)$ is a continuous X-valued martingale and for any $p \in (1,\infty)$ there exists a constant $C_p > 0$, independent of T and ξ, such that*

$$\mathbb{E}\bigg[\sup_{0\leqslant r\leqslant T} |I(r)|_X^p\bigg] \leqslant C_p\bigg(\int_0^T \mathbb{E}\big[\|\xi(t)\|^2_{R(H,X)}\big]\,\mathrm{d}t\bigg)^{p/2}. \tag{8}$$

The inequality (8) is the Burkholder inequality. The case $p = 2$ was proved in [44] and later, using the M-type 2 inequality, was proved in [20] for any $p \in (1,\infty)$.

Remark 7. In the above we may replace $R(H,X)$ by $L(E,X)$, see Remark 5. In particular, $\int_0^T \xi(t)\,\mathrm{d}w(t)$ exists for any $\xi \in M^2(0,T;L(E,X))$ and satisfies

$$\mathbb{E}\left[\sup_{0\leqslant r\leqslant T}\left|\int_0^r \xi(t)\,\mathrm{d}w(t)\right|_X^p\right] \leqslant C_p\left(\int_0^T \mathbb{E}|\xi(t)|^2_{L(E,X)}\,\mathrm{d}t\right)^{p/2}. \tag{9}$$

For suitable maps $f : X \to X$ and $g : X \to R(H, X)$ we consider the following problem

$$\begin{cases} \mathrm{d}\xi(t) = f\big(\xi(t)\big)\,\mathrm{d}t, + g\big(\xi(t)\big)\,\mathrm{d}w(t) \\ \xi(0) \;\; = \xi_0, \end{cases} \tag{10}$$

where $\xi_0 : \Omega \to X$ is $\mathcal{F}_0$-measurable. A continuous and adapted process $\xi : [0, T] \times \Omega \to X$ is said to be a solution to the Itô equation (10) if and only if for all $t \in [0, T]$

$$\xi(t) = \xi(0) + \int_0^t f\big(\xi(r)\big)\,\mathrm{d}r + \int_0^t g\big(\xi(r)\big)\,\mathrm{d}w(r) \qquad \text{a.s.} \tag{11}$$

We have the following existence and uniqueness theorem (see Theorem 2.26 in [10], where only the case $p = 2$ was studied; however, the proof carries over to any $p \in [1, \infty)$ without any substantial difference).

Theorem 2. *Assume that $i : H \hookrightarrow E$ is an AWS, $\{w(t)\}_{t\geqslant 0}$ the corresponding Wiener process on E and X is an M-type 2 Banach space. Let $T > 0$ be fixed. Suppose the maps $f : X \to X$ and $g : X \to R(H, X)$ satisfy the following linear growth and Lipschitz conditions:*

(i) *(Linear Growth Condition) there exists $K > 0$ such that for each $x \in X$*

$$\max\big\{|f(x)|_X, \|g(x)\|_{R(H,X)}\big\} \leqslant K(1 + |x|_X);$$

(ii) *(local Lipschitz continuity) for any $x_0 \in X$ there exists an $r_0 > 0$ and $L_0 > 0$ such that for any x, $y \in \bar{B}(x_0, r_0) := \{x \in X : |x - x_0| \leqslant r_0\}$*

$$\max\big\{|f(x) - f(y)|_X, \|g(x) - g(y)\|_{R(H,X)}\big\} \leqslant L_0|x - y|_X.$$

Let $p \geqslant 1$ and $\xi_0 : \Omega \to X$ be $\mathcal{F}_0$-measurable such that $\mathbb{E}[|\xi_0|_X^p] < \infty$.

Then there exists a unique $\xi \in M^p(0, T; X)$ which is the solution to the problem (10)*. Moreover, the following estimate holds:*

$$\mathbb{E}\left[\sup_{0\leqslant t\leqslant T}|\xi(t)|_X^p\right] \leqslant C_p\big(\mathbb{E}\big[|\xi_0|_X^p\big] + T^p\big). \tag{12}$$

Remark 8. One should point out that the local Lipschitz condition (ii) above is weaker than the usual one:

(ii′) (Lipschitz continuity on balls) for any $R > 0$ there exists $C_R > 0$ such that

$$\max\big\{|f(x) - f(y)|_X, \|g(x) - g(y)\|_{R(H,X)}\big\} \leqslant C_R|x - y|_X$$

for all x, $y \in X$ with $|x|_X$, $|y|_X \leqslant R$.

The condition (ii) is more suitable for studying equations on Banach manifolds. Both conditions are equivalent if $\dim X < \infty$.

So far we have introduced the Itô integral and defined what we mean by a solution to an Itô equation. We now turn to Stratonovitch integrals and Stratonovitch equations. We first need to introduce some notation. By $L_2(E;X)$ we denote the space of bounded bilinear maps, $\Lambda : E \times E \to X$. Let $i : H \to E$ be an AWS. We define the map $\mathrm{tr} : L_2(E;X) \to X$ by

$$\mathrm{tr}\,\Lambda := \int_E \Lambda(e,e)\,\mathrm{d}\mu(e), \tag{13}$$

where μ is the canonical Gaussian measure on E. In view of the Fernique–Landau–Shepp Theorem, tr is a bounded linear map. Note that the tr map depends on the choice of AWS.

The following two definitions are taken from [10].

Definition 4. *Suppose $i : H \to E$ is an AWS with canonical E-valued Wiener process $w(t)$, $t \geqslant 0$, and X is an M-type 2 Banach space. Let $T \in (0,\infty)$ and $\xi(t)$, $t \in [0,T]$ be a stochastic process such that for any $t \geqslant 0$*

$$\xi(t) = \xi(0) + \int_0^t a(r)\,\mathrm{d}r + \int_0^t b(r)\,\mathrm{d}w(r) \qquad a.s.,$$

where $a \in M^1(0,T;X)$ and $b \in M^2(0,T;L(E,X))$. For a C^1 map $g : X \to L(E,X)$ we define the Stratonovitch Integral of $g(\xi(t))$ as

$$\int_0^t g\big(\xi(r)\big) \circ \mathrm{d}w(r) := \int_0^t g\big(\xi(r)\big)\,\mathrm{d}w(r) + \frac{1}{2}\int_0^t \mathrm{tr}\big[g'\big(\xi(r)\big)\,b(r)\big]\,\mathrm{d}r. \tag{14}$$

Remark 9. By a C^1 map we mean that $g : X \to L(E,X)$ is Fréchet differentiable with continuous Fréchet derivative $g' : X \to L(X, L(E,X))$. Furthermore, note that

$$g'\big(\xi(r)\big)\,b(r) \in L\big(E, L(E,X)\big) \simeq L_2(E;X)$$

so that $\mathrm{tr}[g'(\xi(r))\,b(r)]$ appearing in (14) is well defined.

Remark 10. In the definition of the Stratonovitch Integral, it is not accidental that we have chosen

$$b \in M^2\big(0,T;L(E,X)\big) \qquad \text{rather than} \qquad b \in M^2\big(0,T;R(H,X)\big).$$

For a discussion why one needs to consider processes in $M^2(0,T;L(E,X))$ and not in the larger space $M^2(0,T;R(H,X))$, see [10], Appendix A.

Definition 5. *Suppose $i : H \to E$ is an AWS with canonical E-valued Wiener process $w(t)$, $t \geqslant 0$, and X is an M-type 2 Banach space. Let $T \in (0,\infty)$. Let*

g be as above and let $f : X \to X$ be a continuous function. We say that an adapted and continuous X-valued process $\xi(t)$, $t \in [0,T]$, is a solution to the Stratonovitch equation

$$\mathrm{d}\xi(t) = f\big(\xi(t)\big)\,\mathrm{d}t + g\big(\xi(t)\big) \circ \mathrm{d}w(t) \tag{15}$$

if and only if it is a solution to the Itô equation

$$\mathrm{d}\xi(t) = \Big(f\big(\xi(t)\big) + \frac{1}{2}\operatorname{tr}\big[g'\big(\xi(t)\big)\, g\big(\xi(t)\big)\big]\Big)\,\mathrm{d}t + g\big(\xi(t)\big)\,\mathrm{d}w(t). \tag{16}$$

Thus $\xi(t)$ is a solution to (15) if and only if it satisfies for each $t \geqslant 0$

$$\begin{aligned}\xi(t) = \xi(0) &+ \int_0^t f\big(\xi(r)\big)\,\mathrm{d}r\\ &+ \frac{1}{2}\int_0^t \operatorname{tr}\big[g'\big(\xi(r)\big)\, g\big(\xi(r)\big)\big]\,\mathrm{d}r + \int_0^t g\big(\xi(r)\big)\,\mathrm{d}w(r) \qquad \textit{a.s.}\end{aligned} \tag{17}$$

2 Approximations of SDEs with Lipschitz and bounded coefficients

Let X be an M-type 2 Banach space and $i : H \hookrightarrow E$ an AWS with corresponding E-valued Wiener process $w(t)$, $t \geqslant 0$. We impose the following conditions on the coefficients f and g.

(A1) $f : X \to X$ is a C^1-map which is Lipschitz and bounded.
(B1) $g : X \to L(E,X)$ is a C^1 map such that the maps g and g' are Lipschitz and bounded.

We should point out that as a consequence of (B1), the map $\operatorname{tr}(g'g) : X \to X$ is Lipschitz and bounded, where $\operatorname{tr}(g'g)(x) := \operatorname{tr}[g'(x)\, g(x)]$, $x \in X$, see (13). Let $x_0 \in L^p(\Omega, X)$, $p \geqslant 2$ and $T > 0$, be fixed but arbitrary. In view of Theorem 2 there exists a unique continuous progressively measurable process $x : [0,T] \times \Omega \to X$ such that for each $t \in [0,T]$,

$$\begin{aligned}x(t) = x(0) &+ \int_0^t f\big(x(r)\big)\,\mathrm{d}r\\ &+ \frac{1}{2}\int_0^t \operatorname{tr}\big[g'\big(x(r)\big)\, g\big(x(r)\big)\big]\,\mathrm{d}r + \int_0^t g\big(x(r)\big)\,\mathrm{d}w(r), \qquad \text{a.s.}\end{aligned} \tag{18}$$

Moreover, we have the estimate

$$\mathbb{E}\Big[\sup_{0\leqslant t\leqslant T} |x(t)|_X^p\Big] \leqslant C_p\big(\mathbb{E}\big[|x_0|^p\big] + T^p\big). \tag{19}$$

Note that x is a solution to the Stratonovitch equation

$$\mathrm{d}x(t) = f\big(x(t)\big) + g\big(x(t)\big) \circ \mathrm{d}w(t) \tag{20}$$

and x may be written as

$$x(t) = x(0) + \int_0^t f\big(x(r)\big)\,\mathrm{d}r + \int_0^t g\big(x(r)\big) \circ \mathrm{d}w(r), \tag{21}$$

where the last integral on the RHS of (21) is the Stratonovitch integral.

For each $n \in \mathbb{N}$, let π_n be a partition of $[0,T]$, i.e.,

$$0 = t_0 < t_1 < t_2 < \cdots < t_{N(n)} = T.$$

We assume that each partition satisfies

$$\operatorname{mesh} \pi_n := \max_{0 \leqslant k \leqslant N(n)-1} |t_{k+1} - t_k| \leqslant \frac{C_1}{n}, \tag{22}$$

$$N(n) \leqslant C_2 n, \tag{23}$$

where C_1 and C_2 are constants independent of n. For each partition π_n, $n \in \mathbb{N}$, we consider the following piece-wise linear approximation of the E-valued Wiener process $w(t)$:

$$w_{\pi_n}(t) = w(t_i) + \frac{t - t_i}{t_{i+1} - t_i}\big(w(t_{i+1}) - w(t_i)\big), \quad t \in [t_i, t_{i+1}],\ 0 \leqslant i < N(n).$$

Let $x_{\pi_n} : [0,T] \times \Omega \to X$ be the solutions to the family of ODEs (indexed by $\omega \in \Omega$)

$$\begin{cases} \dfrac{\mathrm{d}x_{\pi_n}(t)}{\mathrm{d}t} &= f\big(x_{\pi_n}(t)\big) + g\big(x_{\pi_n}(t)\big)\dfrac{\mathrm{d}w_{\pi_n}(t)}{\mathrm{d}t}, \\ x_{\pi_n}(0) &= x_0. \end{cases} \tag{24}$$

The family of equations (24) may sometimes be written

$$\begin{cases} \mathrm{d}x_{\pi_n}(t) = f\big(x_{\pi_n}(t)\big)\,\mathrm{d}t + g\big(x_{\pi_n}(t)\big)\,\mathrm{d}w_{\pi_n}(t) \\ x_{\pi_n}(0) \ \ = x_0 \end{cases}$$

In particular, for $t \in (t_i, t_{i+1})$, $i = 0, \ldots, N(n)-1$, x_{π_n} takes the form

$$x_{\pi_n}(t) = x_{\pi_n}(t_i) + \int_{t_i}^t f\big(x_{\pi_n}(s)\big)\,\mathrm{d}s + \int_{t_i}^t g\big(x_{\pi_n}(s)\big)\left(\frac{w(t_{i+1}) - w(t_i)}{t_{i+1} - t_i}\right)\mathrm{d}s.$$

Using the above notation, we now state our first result.

Theorem 3. *For $p > 2$ and $n \in \mathbb{N}$*

$$\mathbb{E}\Big[\sup_{0 \leqslant t \leqslant T} |x(t) - x_{\pi_n}(t)|_X^p\Big] \leqslant C n^{-p/2}, \tag{25}$$

where C is a constant independent of n and depending only on the space X, p, T, m_p (see (5)*), C_1, C_2 and the bounds and Lipschitz constants of f, g, g' and* $\operatorname{tr}(g'g)$.

Corollary 1. *For each $T > 0$ $x_{\pi_n}(\,.\,) \to x(\,.\,)$ in $C(0,T;X)$ in probability, i.e., for each $\varepsilon > 0$*

$$\mathbb{P}\big\{\omega : |x(\,.\,,\omega) - x_{\pi_n}(\,.\,,\omega)|_{C(0,T;X)} > \varepsilon\big\} \longrightarrow 0 \tag{26}$$

as $\operatorname{mesh}\pi_n \to 0$. *Here $C(0,T;X)$ is the space of X valued continuous functions on the interval $[0,T]$.*

Corollary 2. *For each $T > 0$,*

$$x_{\pi_n}(\,.\,) \longrightarrow x(\,.\,) \qquad \text{in } C(0,T;X) \text{ almost surely as } n \to \infty. \tag{27}$$

Remark 11. Theorem 3 is an extension of a result proved in the PhD thesis by Dowell, [23]. There, the case $p = 2$ with X being a Hilbert space was treated. In particular, Dowell proved the following two results (more or less independently of one another), see Theorems 5.2 and 5.7 in [23]:

- For each $T > 0$

$$\sup_{0 \leqslant t \leqslant T} \mathbb{E}\big[|x(t) - x_{\pi_n}(t)|_X^2\big] \longrightarrow 0 \qquad \text{as } \operatorname{mesh}\pi_n \to 0. \tag{28}$$

- For each $T > 0$ and $\varepsilon > 0$

$$\mathbb{P}\Big\{\omega : \sup_{0 \leqslant t \leqslant T} |x(t,\omega) - x_{\pi_n}(t,\omega)|_X > \varepsilon\Big\} \longrightarrow 0 \qquad \text{as } \operatorname{mesh}\pi_n \to 0. \tag{29}$$

Our result is a much stronger and more general result than Dowell's for several reasons. Firstly, Theorem 3 holds in the case when X is an M-type 2 Banach space. Secondly, we have convergence in $L^p(\Omega; C(0,T;X))$, $p \geqslant 2$, whereas Dowell only proved a weaker form of convergence, i.e., uniform convergence in $L^2(\Omega;X)$, see (28). With this stronger form of convergence, convergence in $C(0,T;X)$ in probability is then a simple consequence of the Chebyshev inequality and this gives us Corollary 1. Finally, for $p > 2$ we prove estimates which give a rate of convergence, see (25). Using these estimates it is straightforward to prove almost sure convergence in $C(0,T;X)$ (see Corollary 2). Indeed, the estimates (25) imply that, since $p > 2$,

$$\mathbb{E}\Big[\sum_{n=1}^{\infty} |x - x_{\pi_n}|_{C(0,T;X)}\Big] \leqslant C \sum_{n=1}^{\infty} \frac{1}{n^{p/2}} < \infty.$$

Thus, almost surely

$$\sum_{n=1}^{\infty} |x - x_{\pi_n}|_{C(0,T;X)} < \infty,$$

which implies that almost surely

$$|x - x_{\pi_n}|_{C(0,T;X)} \longrightarrow 0 \qquad \text{as } n \to \infty.$$

The method Dowell uses to prove (28), which itself is a generalization of a similar result in [43], carries over to some extent to the case $p > 2$ and X is an M-type 2 Banach space. The Burkholder inequality (9) is the main tool we use here. However, although Dowell was familiar with stochastic integration in 2-uniformly smooth Banach spaces and the Burkholder inequality (via the thesis of Neidhardt), he was not able to deal with the Banach space case because of the term involving the tr map. There is a considerable level of difficulty in dealing with the tr map in Banach spaces as opposed to Hilbert spaces. We deal with this problem by making use of the M-type 2 property of our space X, in particular, the inequality (1).

Proof of Theorem 3. Fix a partition $\pi = \pi_n = \{0 \leqslant t_0 \leqslant t_1 \cdots \leqslant t_{N(n)} = T\}$ and denote x_π by y. Set $x_j = x(t_j)$, $y_j = y(t_j) = x_\pi(t_j)$, $\Delta_j t = t_{j+1} - t_j$ and $\Delta_j w = w(t_{j+1}) - w(t_j)$. To simplify the notation we put f identically zero. This will not affect the result owing to the conditions put on f. Moreover, C will denote a generic constant depending only on the space X, p, T, m_p, C_1, C_2 the bounds and Lipschitz constants of g, g' and tr.

For $t \in [0,T]$, let k be the largest integer such that $t_k \leqslant t$. Moreover, for $r \in [0,T]$, set $R(n) = \max\{m : t_m \leqslant r\}$. Then, using the triangle inequality, we have

$$\mathbb{E}\Big[\sup_{0\leqslant t\leqslant r} |x(t)-y(t)|_X^p\Big] \leqslant C\,\mathbb{E}\Big[\sup_{0\leqslant t\leqslant r}\Big(|x(t)-x(t_k)|_X^p + |y(t_k)-y(t)|_X^p\Big)\Big] + C\,\mathbb{E}\Big[\sup_{0\leqslant k\leqslant R(n)} |x(t_k)-y(t_k)|_X^p\Big]. \tag{30}$$

Suppose, for the time being, we have the following estimates

$$\mathbb{E}\Big[\sup_{0\leqslant t\leqslant r}\Big(|x(t)-x(t_k)|_X^p + |y(t_k)-y(t)|_X^p\Big)\Big] \leqslant C\eta(\pi), \tag{31}$$

$$\mathbb{E}\Big[\sup_{0\leqslant k\leqslant R(n)} |x(t_k)-y(t_k)|_X^p\Big] \leqslant \eta(\pi) + C\int_0^r \mathbb{E}[\gamma(s)]\,\mathrm{d}s, \tag{32}$$

where

$$\gamma(s) = \sup_{0\leqslant l\leqslant s} |x(l)-y(l)|_X^p \tag{33}$$

and $\eta(\pi)$ is independent of k and satisfies

$$\eta(\pi) \leqslant C\,n^{-p/2}.$$

(Note, for example, that $(\mathrm{mesh}\,\pi)^{p/2}$ is a term of the form $\eta(\pi)$.) From (30), (31), (32) and (33) we may deduce that for all $r \in [0,T]$:

$$\mathbb{E}[\gamma(r)] = \mathbb{E}\Big[\sup_{0\leqslant t\leqslant r} |x(t)-y(t)|_X^p\Big] \leqslant C\,\eta(\pi) + C\int_0^r \mathbb{E}[\gamma(s)]\,\mathrm{d}s.$$

An application of Gronwall's Lemma implies that

$$\mathbb{E}[\gamma(T)] \leqslant C\,\eta(\pi)\exp(CT),$$

i.e.,

$$\mathbb{E}\left[\sup_{0\leqslant t\leqslant T}|x(t)-y(t)|_X^p\right] \leqslant C\,n^{-p/2}.$$

To complete the proof of Theorem 3 we need to prove the estimates (31) and (32). We begin with (31).

Lemma 2. *With the above notation,*

$$\mathbb{E}\left[\sup_{0\leqslant t\leqslant r}\Big(|x(t)-x(t_k)|_X^p+|y(t_k)-y(t)|_X^p\Big)\right] \leqslant C\,(\operatorname{mesh}\pi)^{p/2}. \tag{34}$$

Proof. Note first that from (18) and the boundedness of the maps g and $\operatorname{tr}(g'g)$ we have

$$\mathbb{E}\left[\sup_{0\leqslant t\leqslant r}|x(t)-x(t_k)|_X^p\right] \leqslant C\,(\operatorname{mesh}\pi)^p + C\,\mathbb{E}\left[\sup_{0\leqslant t\leqslant r}\left|\int_{t_k}^t g(x(s))\,\mathrm{d}w(s)\right|_X^p\right].$$

It then follows using the Burkholder inequality and the boundedness of g that

$$\mathbb{E}\left[\sup_{0\leqslant t\leqslant r}|x(t)-x(t_k)|_X^p\right] \leqslant CT^{p/2}(\operatorname{mesh}\pi)^{p/2}.$$

Recall Taylor's formula in integral form, see [15]:

$$y(a)-y(b)=\int_0^1 y'\big(b+r(a-b)\big)(a-b)\,\mathrm{d}r. \tag{35}$$

For some $0\leqslant s\leqslant 1$, we have, using (35), (24) and the boundedness of g,

$$\begin{aligned}
|y(t)-y(t_k)|_X^p &= |y(t_k+s\,\Delta_k t)-y(t_k)|_X^p\\
&= \left|\int_0^1 y'(t_k+r(s\,\Delta_k t))(s\,\Delta_k t)\,\mathrm{d}r\right|_X^p\\
&= \left|\int_0^s y'(t_k+r\,\Delta_k t)(\Delta_k t)\,\mathrm{d}r\right|_X^p\\
&= \left|\int_0^s g(y(t_k+r\,\Delta_k t))(\Delta_k w)\,\mathrm{d}r\right|_X^p\\
&\leqslant C|\Delta_k w|_E^p.
\end{aligned} \tag{36}$$

Using (5) we infer that

$$\mathbb{E}\left[\sup_{0\leqslant t\leqslant r}|y(t)-y(t_k)|_X^p\right] \leqslant C\,(\operatorname{mesh}\pi)^{p/2}.$$

This completes the proof of Lemma 2. ∎

Fix an interval $[t_i, t_{i+1}]$ in the partition π. We quote another form of Taylor's formula, see [15]:

$$y(a) - y(b) = y'(b)(a-b) + \int_0^1 (1-s) y''\big(b + s(a-b)\big)(a-b, a-b)\,\mathrm{d}s. \quad (37)$$

Using (37), the chain rule and (24) we obtain

$$y(t_{i+1}) - y(t_i) = y'(t_j)\,\Delta_j t + \int_0^1 (1-s)\, y''(t_j + s\,\Delta_j t)(\Delta_j t, \Delta_j t)\,\mathrm{d}s$$
$$= g(y_j)\,\Delta_j w + \int_0^1 (1-s)\Big(g'\big(y(t_j + s\,\Delta_j t)\big)\, g\big(y(t_j + s\,\Delta_j t)\big)(\Delta_j w, \Delta_j w)\Big)\,\mathrm{d}s.$$

It then follows, denoting $s_j := t_j + s\,\Delta_j t$, that

$$\begin{aligned}
y(t_k) - y(0) &= \sum_{j=0}^{k-1} (y_{j+1} - y_j) \\
&= \sum_{j=0}^{k-1} \Big(g(y_j)\,\Delta_j w + \frac{1}{2}\, g'(y_j)\, g(y_j)(\Delta_j w, \Delta_j w) \Big) \\
&\quad + \sum_{j=0}^{k-1} \int_0^1 (1-s)\, g'\big(y(s_j)\big)\, g\big(y(s_j)\big)(\Delta_j w, \Delta_j w)\,\mathrm{d}s \\
&\quad - \sum_{j=0}^{k-1} \int_0^1 (1-s)\, g'(y_j)\, g(y_j)(\Delta_j w, \Delta_j w)\,\mathrm{d}s.
\end{aligned}$$

Recalling that

$$x(t_k) = x(0) + \int_0^{t_k} g\big(x(s)\big)\,\mathrm{d}w(s) + \frac{1}{2} \int_0^{t_k} \mathrm{tr}\big[g'\big(x(s)\big)\, g\big(x(s)\big)\big]\,\mathrm{d}s,$$

we may write

$$y(t_k) - x(t_k) = A_k + B_k + \frac{1}{2}\,\overline{C}_k + D_k + \frac{1}{2}\, E_k + \frac{1}{2}\, F_k,$$

where

$$\begin{aligned}
A_k &= \sum_{j=0}^{k-1} \int_0^1 (1-s)\Big(g'\big(y(s_j)\big)\, g\big(y(s_j)\big) - g'(y_j)\, g(y_j)\Big)(\Delta_j w, \Delta_j w)\,\mathrm{d}s \\
B_k &= \sum_{j=0}^{k-1} \big(g(y_j) - g(x_j)\big)\,\Delta_j w \\
\overline{C}_k &= \sum_{j=0}^{k-1} \Big(g'(y_j)\, g(y_j) - g'(x_j)\, g(x_j)\Big)(\Delta_j w, \Delta_j w)
\end{aligned}$$

$$D_k = \sum_{j=0}^{k-1} g(x_j)\,\Delta_j w - \int_0^{t_k} g\big(x(s)\big)\,\mathrm{d}w(s)$$

$$E_k = \sum_{j=0}^{k-1}\big(g'(x_j)\,g(x_j)(\Delta_j w, \Delta_j w) - \mathrm{tr}[g'(x_j)\,g(x_j)]\,\Delta_j t\big)$$

$$F_k = \sum_{j=0}^{k-1} \mathrm{tr}[g'(x_j)\,g(x_j)]\,\Delta_j t - \int_0^{t_k} \mathrm{tr}\big[g'\big(x(t)\big)\,g\big(x(t)\big)\big]\,\mathrm{d}t.$$

We begin with proving:

Lemma 3. *Using the above notation we have*

$$\mathbb{E}\Big[\sup_{1\leqslant k\leqslant R(n)} |A_k + D_k + E_k + F_k|_X^p\Big] \leqslant C\,(\mathrm{mesh}\,\pi)^{p/2}.$$

Proof. Consider first the term $A_k = \sum_{j=0}^{k-1} \Gamma_j$, where

$$\Gamma_j := \int_0^1 (1-s)\Big(g'\big(y(s_j)\big)\,g\big(y(s_j)\big)(\Delta_j w, \Delta_j w) - g'(y_j)\,g(y_j)(\Delta_j w, \Delta_j w)\Big)\,\mathrm{d}s.$$

The boundedness and Lipschitz properties of g' and g, along with (36), imply that

$$\begin{aligned}|\Gamma_j|_X &\leqslant \int_0^1 \Big|\Big(g'\big(y(s_j)\big) - g'(y_j)\Big)\,g\big(y(s_j)\big)(\Delta_j w, \Delta_j w)\Big|_X\,\mathrm{d}s \\ &\qquad + \int_0^1 \Big|g'(y_j)\Big(g\big(y(s_j)\big) - g(y_j)\Big)(\Delta_j w, \Delta_j w)\Big|_X\,\mathrm{d}s \\ &\leqslant C|\Delta_j w|_E^2\,|y(s_j) - y_j|_X \\ &\leqslant C|\Delta_j w|_E^3. \end{aligned}\tag{38}$$

Using (38) and Hölder's inequality for sums we have

$$\mathbb{E}\Big[\sup_{1\leqslant k\leqslant N(n)} |A_k|_X^p\Big] \leqslant CN(n)^{p-1}\,\mathbb{E}\Big[\sum_{j=0}^{N(n)-1} |\Delta_j w|_E^{3p}\Big].$$

Applying (5) (with p replaced by $3p$) gives us

$$\mathbb{E}\Big[\sup_{1\leqslant k\leqslant N(n)} |A_k|_X^p\Big] \leqslant CN(n)^{p-1} \sum_{j=0}^{N(n)-1} |\Delta_j t|^{3p/2}.$$

It then follows, using (22) and (23), that

$$\mathbb{E}\Big[\sup_{1\leqslant k\leqslant N(n)} |A_k|_X^p\Big] \leqslant C\,n^{p-1}(\mathrm{mesh}\,\pi)^{3p/2} n \leqslant C(\mathrm{mesh}\,\pi)^{p/2}.\tag{39}$$

Consider then the term $D_k = \sum_{j=0}^{k-1} g(x_j)\,\Delta_j w - \int_0^{t_k} g(x(s))\,\mathrm{d}w(s)$. Define

$$\tilde{g}(s) = \begin{cases} g(x_j) & \text{for } t_j \leqslant s < t_{j+1}, \\ 0 & \text{if } s > t_k. \end{cases}$$

$\tilde{g}(s)$ is well-defined, adapted to the filtration $\{\mathcal{F}_s\}_{s\geqslant 0}$ and moreover, the integral $\int_0^t \tilde{g}(s)\,\mathrm{d}w(s)$ makes sense for all $t \in [0,T]$. We may write

$$D_k = \int_0^{t_k} \Big(\tilde{g}(s) - g\big(x(s)\big)\Big)\,\mathrm{d}w(s).$$

Using the Burkholder inequality, the Lipschitz property of g and the properties (22) and (23), it follows that

$$\begin{aligned}
\mathbb{E}\Big[\sup_{1\leqslant k\leqslant R(n)} |D_k|_X^p\Big] &\leqslant \mathbb{E}\Big[\sup_{0\leqslant t\leqslant r}\Big|\int_0^t \Big(\tilde{g}(s) - g\big(x(s)\big)\Big)\,\mathrm{d}w(s)\Big|_X^p\Big] \\
&\leqslant C\,\mathbb{E}\Big[\Big(\int_0^r \big|\tilde{g}(s) - g\big(x(s)\big)\big|_{L(E,X)}^2\,\mathrm{d}s\Big)^{p/2}\Big] \\
&= C\,\mathbb{E}\Big[\Big(\sum_{j=0}^{R(n)-1}\int_{t_j}^{t_{j+1}} \big|g(x_j) - g\big(x(s)\big)\big|_{L(E,X)}^2\,\mathrm{d}s\Big)^{p/2}\Big] \\
&\leqslant C\,\mathbb{E}\Big[\Big(\sum_{j=0}^{R(n)-1}\int_{t_j}^{t_{j+1}} |x_j - x(s)|_X^2\,\mathrm{d}s\Big)^{p/2}\Big] \\
&\leqslant C\,\mathbb{E}\Big[\sup_{0\leqslant t\leqslant r} |x(t_l) - x(t)|_X^p\Big]
\end{aligned}$$

where l is such that $t \in [t_l, t_{l+1})$. Using Lemma 2 we deduce that

$$\mathbb{E}\Big[\sup_{1\leqslant k\leqslant R(n)} |D_k|_X^p\Big] \leqslant C(\operatorname{mesh}\pi)^{p/2}. \tag{40}$$

Consider next the term F_k. We have

$$\begin{aligned}
|F_k|_X &= \Big|\sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}} \Big(\mathrm{tr}[g'(x_j)\,g(x_j)] - \mathrm{tr}\big[g'\big(x(t)\big)\,g\big(x(t)\big)\big]\Big)\,\mathrm{d}t\Big|_X \\
&\leqslant \sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}} \Big|\mathrm{tr}\big[g'(x_j)\,g(x_j) - g'\big(x(t)\big)\,g\big(x(t)\big)\big]\Big|_X\,\mathrm{d}t \\
&\leqslant C\sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}} \big|g'(x_j)\,g(x_j) - g'\big(x(t)\big)\,g\big(x(t)\big)\big|_{L_2(E,X)}\,\mathrm{d}t \\
&\leqslant C\sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}} \Big(\big|g'(x_j)\,g(x_j) - g'(x_j)\,g\big(x(t)\big)\big|_{L_2(E,X)} \\
&\qquad + \big|g'(x_j)\,g\big(x(t)\big) - g'\big(x(t)\big)\,g\big(x(t)\big)\big|_{L_2(E,X)}\Big)\,\mathrm{d}t.
\end{aligned}$$

Using the boundedness and the Lipschitz properties of functions g and g', we deduce that

$$\begin{aligned}|F_k|_X &\leqslant C\sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}} |x_j - x(t)|_X \,\mathrm{d}t \\ &\leqslant CT \sup_{0\leqslant t\leqslant r} |x(t_l) - x(t)|_X,\end{aligned}$$

where l is such that $t \in [t_l, t_{l+1})$. Again, using Lemma 2, we conclude that

$$\mathbb{E}\Big[\sup_{1\leqslant k\leqslant R(n)} |F_k|_X^p\Big] \leqslant C(\operatorname{mesh}\pi)^{p/2}. \tag{41}$$

Finally, we deal with the term E_k and we will prove

$$\mathbb{E}\Big[\sup_{1\leqslant k\leqslant R(n)} |E_k|_X^p\Big] \leqslant C(\operatorname{mesh}\pi)^{p/2}. \tag{42}$$

This part of the proof differs considerably from [23]. Dowell proves (42) using the properties of the inner product on a Hilbert space and the proof is quite straightforward. We do not have an inner product to work with and instead we make use of the M-type 2 property of our space X. Let $E_k = \sum_{j=0}^{k-1} \Lambda_j$, where

$$\Lambda_j = g'(x_j)\, g(x_j)(\Delta_j w, \Delta_j w) - \operatorname{tr}\big(g'(x_j)\, g(x_j)\big)\, \Delta_j t \tag{43}$$

We first show that E_k is an X-valued martingale with respect to the discrete filtration $\{\mathcal{F}_{t_k}\}_{1\leqslant k\leqslant R(n)}$. For $0 \leqslant j \leqslant k-1$, $x_j : \Omega \to X$ is $\mathcal{F}_{t_j}$-measurable and $w(t_{j+1}) - w(t_j) : \Omega \to E$ is $\mathcal{F}_{t_{j+1}}$-measurable. Using the continuity of the maps g, g' and $\operatorname{tr}(g'g)$ it follows that each Λ_j is $\mathcal{F}_{t_{j+1}}$-measurable. We deduce that E_k is $\mathcal{F}_{t_k}$-measurable. To prove E_k is a martingale we are left with showing that $\mathbb{E}[E_k \,|\, \mathcal{F}_{t_{k-1}}] = E_{k-1}$. For this it suffices to prove that $\mathbb{E}[\Lambda_{k-1} \,|\, \mathcal{F}_{t_{k-1}}] = 0$.

Denote

$$\Psi_{k-1} := g'(x_{k-1})\, g(x_{k-1})(\Delta_{k-1} w, \Delta_{k-1} w).$$

Then

$$\begin{aligned}\mathbb{E}[\Psi_{k-1} \,|\, \mathcal{F}_{k-1}] &= \mathbb{E}[g'(x_{k-1})\, g(x_{k-1})(\Delta_{k-1} w, \Delta_{k-1} w) \,|\, \mathcal{F}_{t_{k-1}}] \\ &= g'(x_{k-1})\, g(x_{k-1})\mathbb{E}[(\Delta_{k-1} w, \Delta_{k-1} w)] \\ &= (t_k - t_{k-1}) \int_E g'(x_{k-1})\, g(x_{k-1})(e, e)\, \mathrm{d}\mu(e) \\ &= (\Delta_{k-1} t)\operatorname{tr}\big(g'(x_{k-1})\, g(x_{k-1})\big) \\ &= \mathbb{E}\big[(\Delta_{k-1} t)\operatorname{tr}\big(g'(x_{k-1})\, g(x_{k-1})\big) \;\big|\; \mathcal{F}_{t_{k-1}}\big].\end{aligned} \tag{44}$$

As x_{k-1} is $\mathcal{F}_{t_{k-1}}$-measurable, then so is $\operatorname{tr}(g'(x_{k-1})\, g(x_{k-1}))$, which explains the final step. Thus (43) and (44) imply that $\mathbb{E}[\Lambda_{k-1} \,|\, \mathcal{F}_{t_{k-1}}] = 0$. We conclude that $\{E_k\}_{k=1}^{R(n)}$ is an X-valued martingale with respect to the discrete filtration $\{\mathcal{F}_{t_k}\}_{1\leqslant k\leqslant R(n)}$. Since X is an M-type 2 Banach space it follows that, see (1),

$$\mathbb{E}\Big[\sup_{1\leqslant k\leqslant R(n)} |E_k|_X^p\Big] \leqslant C\,\mathbb{E}\Bigg[\bigg(\sum_{j=1}^{R(n)-1} |E_j - E_{j-1}|_X^2\bigg)^{p/2}\Bigg].$$

Thus

$$\mathbb{E}\Bigg[\sup_{1\leqslant k\leqslant R(n)}\bigg|\sum_{j=0}^{k-1} \Lambda_j\bigg|_X^p\Bigg] \leqslant C\,\mathbb{E}\Bigg[\bigg(\sum_{j=1}^{R(n)} |\Lambda_{j-1}|_X^2\bigg)^{p/2}\Bigg].$$

Applying the Hölder inequality for sums gives

$$\begin{aligned}\mathbb{E}\Bigg[\sup_{1\leqslant k\leqslant R(n)}\bigg|\sum_{j=0}^{k-1} \Lambda_j\bigg|_X^p\Bigg] &\leqslant CR(n)^{p/2-1}\,\mathbb{E}\Bigg[\sum_{j=1}^{R(n)} |\Lambda_{j-1}|_X^p\Bigg]\\ &\leqslant CN(n)^{p/2-1}\sum_{j=1}^{N(n)} \mathbb{E}\big[|\Lambda_{j-1}|_X^p\big]. \end{aligned}\tag{45}$$

Note that

$$\begin{aligned}\mathbb{E}\big[|\Lambda_j|_X^p\big] &\leqslant \mathbb{E}\Big[\Big(\big|g'(x_j)\,g(x_j)(\Delta_j w, \Delta_j w)\big|_X + \big|\mathrm{tr}\big(g'(x_j)\,g(x_j)\big)(\Delta_j t)\big|_X\Big)^p\Big]\\ &\leqslant C\,\mathbb{E}\big[|\Delta_j w|_E^{2p} + |\Delta_j t|^p\big]\\ &\leqslant C(\Delta_j t)^p.\end{aligned}\tag{46}$$

It follows from (45) and (46) that

$$\begin{aligned}\mathbb{E}\Big[\sup_{1\leqslant k\leqslant R(n)} |E_k|_X^p\Big] &\leqslant CN(n)^{p/2-1}\sum_{j=1}^{N(n)} (\Delta_j t)^p\\ &\leqslant CN(n)^{p/2-1}\sum_{j=1}^{N(n)} (\mathrm{mesh}\,\pi)^p\\ &\leqslant CN(n)^{p/2}(\mathrm{mesh}\,\pi)^p \leqslant C(\mathrm{mesh}\,\pi)^{p/2}.\end{aligned}\tag{47}$$

Lemma 3 now follows from (39), (40), (41) and (47). □

Lemma 4. *For a constant C independent of k and r,*

$$\mathbb{E}\Big[\sup_{1\leqslant k\leqslant R(n)}\big|B_k + \overline{C}_k\big|_X^p\Big] \leqslant C\int_0^r \mathbb{E}[\gamma(s)]\,\mathrm{d}s \tag{48}$$

Proof. As in the proof of Lemma 3, define

$$Y(s) = \begin{cases} g(y_j) - g(x_j) & \text{if } t_j \leqslant s < t_{j+1}, \text{ where } 0\leqslant j\leqslant k-1,\\ 0 & \text{if } s > t_k.\end{cases}$$

$Y(s)$ is well-defined, adapted to the filtration $\{\mathcal{F}_s\}_{s\geqslant 0}$ and $\int_0^t Y(s)\,\mathrm{d}w(s)$ makes sense for all $t \in [0,T]$. Moreover,

$$|B_k|_X^p = \left|\int_0^{t_k} Y(s)\,\mathrm{d}w(s)\right|_X^p.$$

Using the Burkholder inequality and the Lipschitz properties of g, it follows that

$$\begin{aligned}
\mathbb{E}\Bigg[\sup_{1\leqslant k\leqslant R(n)} |B_k|_X^p\Bigg] &= \mathbb{E}\Bigg[\sup_{0\leqslant t\leqslant r}\left|\int_0^t Y(s)\,\mathrm{d}w(s)\right|_X^p\Bigg] \\
&\leqslant C\,\mathbb{E}\Bigg[\bigg(\int_0^r |Y(s)|_{L(E,X)}^2\,\mathrm{d}s\bigg)^{p/2}\Bigg] \\
&= C\,\mathbb{E}\Bigg[\Bigg(\sum_{j=0}^{R(n)-1}\int_{t_j}^{t_{j+1}} |g(y_j)-g(x_j)|_{L(E,X)}^2\,\mathrm{d}s\Bigg)^{p/2}\Bigg] \\
&\leqslant C\,\mathbb{E}\Bigg[\Bigg(\sum_{j=0}^{R(n)-1} |y_j-x_j|_X^2\,\Delta_j t\Bigg)^{p/2}\Bigg] \\
&\leqslant C\,\mathbb{E}\Bigg[\Bigg(\sum_{j=0}^{R(n)-1} \gamma(t_j)^{2/p}\,\Delta_j t\Bigg)^{p/2}\Bigg].
\end{aligned}$$

Applying the Hölder inequality for sums gives

$$\begin{aligned}
\mathbb{E}\Bigg[\sup_{1\leqslant k\leqslant R(n)} |B_k|_X^p\Bigg] &\leqslant CR(n)^{p/2-1}\,\mathbb{E}\Bigg[\sum_{j=0}^{R(n)-1}\gamma(t_j)(\Delta_j t)^{p/2}\Bigg] \\
&\leqslant CN(n)^{p/2-1}(\operatorname{mesh}\pi)^{p/2-1}\sum_{j=0}^{R(n)-1}\mathbb{E}[\gamma(t_j)\,\Delta_j t] \\
&\leqslant C\int_0^r \mathbb{E}[\gamma(s)]\,\mathrm{d}s,
\end{aligned}$$

which constitutes the first in proving Lemma 4. Consider the final term $\overline{C}_k$. Then

$$\begin{aligned}
\big|\overline{C}_k\big|_X &= \left|\sum_{j=0}^{k-1}\big(g'(y_j)g(y_j)-g'(x_j)g(x_j)\big)(\Delta_j w,\Delta_j w)\right|_X \\
&\leqslant \sum_{j=0}^{k-1}\Big(\big|\big(g'(y_j)-g'(x_j)\big)\,g(x_j)(\Delta_j w,\Delta_j w)\big|_X \\
&\qquad + \big|g'(y_j)\big(g(y_j)-g(x_j)\big)(\Delta_j w,\Delta_j w)\big|_X\Big) \\
&\leqslant C\sum_{j=0}^{k-1}|x_j-y_j|_X\,|\Delta_j w|_E^2.
\end{aligned}$$

Applying the Hölder inequality gives

$$\left|\overline{C}_k\right|^p \leqslant CN(n)^{p-1} \sum_{j=0}^{k-1} |x_j - y_j|_X^p \, |\Delta_j w|_E^{2p}.$$

On taking supremum over k and then expectations we get

$$\mathbb{E}\Big[\sup_{1\leqslant k\leqslant R(n)} \left|\overline{C}_k\right|_X^p\Big] \leqslant CN(n)^{p-1} \sum_{j=0}^{R(n)-1} \mathbb{E}\big[|x_j - y_j|_X^p \, |\Delta_j w|_E^{2p}\big].$$

Since both x_j and y_j are $\mathcal{F}_{t_j}$-measurable and $\Delta_j w$ is independent of $\mathcal{F}_{t_j}$ then using the properties of conditional expectation and (5) we have

$$\begin{aligned}\mathbb{E}\big[|x_j - y_j|_X^p |\Delta_j w|_{E_0}^{2p}\big] &= \mathbb{E}\Big[\mathbb{E}\big[|x_j - y_j|_X^p \, |\Delta_j w|_E^{2p} \mid \mathcal{F}_{t_j}\big]\Big] \\ &= \mathbb{E}\Big[|x_j - y_j|_X^p \, \mathbb{E}\big[|\Delta_j w|_E^{2p} \mid \mathcal{F}_{t_j}\big]\Big] \\ &= \mathbb{E}\Big[|x_j - y_j|_X^p \, \mathbb{E}\big[|\Delta_j w|_E^{2p}\big]\Big] \\ &\leqslant C|\Delta_j t|^p \, \mathbb{E}\big[|x_j - y_j|_X^p\big]. \end{aligned} \tag{49}$$

It then follows using (49), (22) and (23) that

$$\begin{aligned}\mathbb{E}\Big[\sup_{1\leqslant k\leqslant R(n)} \left|\overline{C}_k\right|_X^p\Big] &\leqslant CN(n)^{p-1} \sum_{j=0}^{R(n)-1} \mathbb{E}\big[|x_j - y_j|_X^p\big] |\Delta_j t|^p \\ &\leqslant CN(n)^{p-1} (\mathrm{mesh}\,\pi)^{p-1} \sum_{j=0}^{R(n)-1} (\Delta_j t) \, \mathbb{E}\Big[\sup_{0\leqslant r\leqslant t_j} |x(r) - y(r)|_X^p\Big] \\ &\leqslant C \sum_{j=0}^{R(n)-1} \mathbb{E}[(\gamma(t_j) \, \Delta_j t]. \end{aligned}$$

Since $\gamma(s)$ is non-decreasing we can conclude that

$$\mathbb{E}\Big[\sup_{1\leqslant k\leqslant R(n)} \left|\overline{C}_k\right|_X^p\Big] \leqslant C \int_0^r \mathbb{E}[\gamma(s)] \, \mathrm{d}s,$$

which concludes the proof of Lemma 4. The proof of Theorem 3 is now complete. □

Remark 12. In a very recent preprint [32] by M. Ledoux, T. Lyons and Z.Qian, the authors extend the main results of [33] to a wide class of Banach spaces. The finite dimensional case of the rough path theory, see [33], gives deep understanding of what approximation procedure leads to Stratonovitch stochastic differential equations. The infinite dimensional case discussed in the above cited preprint should give greater understanding of Corollary 2. On the other hand, our results could be used to show that the rough path theory agrees with classical theory of stochastic differential equations in M-type 2 Banach spaces. One can point out a difference concerning regularity assumptions between our paper and [33], [32]. While we assume that the coefficient g is of C^2-class (i.e., g' is Lipschitz), the assumption in the above two papers is that g is of $C^{2+\varepsilon}$-class for some $\varepsilon > 0$ depending on the roughness of the driving rough path.

In another recent work [19] the author employs the Euler method to prove local existence of solutions to differential equations in finite dimensional spaces driven by a finite dimensional rough path. It would be interesting to extend his result to an infinite dimensional case and to also consider the global existence of solutions when the input is a p-rough path with $p > 2$. Such results would help to give a better understanding of the relationship between our paper and the T. Lyons theory, in particular with the above mentioned preprint [32]. The authors would like to thank the anonymous referee for informing them about the interesting paper by A.M. Davie [19].

3 Approximation of SDEs whose coefficients are locally Lipschitz

In this section we improve the result given as Corollary 1. We no longer assume that the maps f and g satisfy a global Lipschitz condition nor that they are bounded. We assume the following conditions hold true.

(A2) $f : X \to X$ is a C^1-map which is Lipschitz on balls.
(B2) $g : X \to L(E, X)$ is a C^1 map such that the maps g and g' are Lipschitz on balls.
(C2) The functions f, g and $\mathrm{tr}(g'g)$ are of linear growth.

We should point out here that if the condition (B2) is satisfied, the map $\mathrm{tr}(g'g) : X \to X$ is also Lipschitz on balls. We would like also to stress that we have imposed the condition (C2) in order to ensure that there exists a global solution to the problems (50) and (51) below.

In addition to the assumption that X is an M-type 2 Banach space, we assume also the following.

(D2) There exists a C^1-class bump function $\phi : X \to \mathbb{R}$ such that

$$\phi(x) = \begin{cases} 1 & \text{if } |x|_X \leqslant 1, \\ 0 & \text{if } |x|_X \geqslant 2, \end{cases}$$

$0 \leqslant \phi(x) \leqslant 1$ for $x \in X$, and the first derivative of ϕ, ϕ', is Lipschitz and bounded.

Remark 13. Concerning the assumptions on X, it would suffice, for example, to assume that for some $p \geqslant 2$, X satisfies the following condition:

(H_p) The function $\phi_p : X \to \mathbb{R}$ given by $\phi_p(x) = |x|_X^p$ is of C^2 class and there exists constants $k_1,\ k_2 > 0$ such that $|\phi_p'(x)| \leqslant k_1 |x|_X^{p-1}$ and $|\phi_p''(x)| \leqslant k_2 |x|_X^{p-2}$, for $x \in X$.

It is straightforward to show the existence of the bump function ϕ if (H_p) holds. Secondly, any Banach space satisfying (H_p) is of M-type 2, see [13].

It is worthwhile noticing that for any $q \geqslant p$ the Lebesgue spaces L^q and the Sobolev–Slobodetskii spaces $W^{\theta,q}$ (see Section 4), satisfy (H_p), see [21].

Fix $T > 0$, $p \geqslant 2$ and $x_0 \in L^p(\Omega, X)$. For a partition π of $[0,T]$ let $x : [0,T] \times \Omega \to X$ and $x_\pi : [0,T] \times \Omega \to X$ be the respective solutions to the problems

$$\begin{cases} \mathrm{d}x(t) = f\big(x(t)\big)\,\mathrm{d}t + g\big(x(t)\big) \circ \mathrm{d}w(t), & t \geqslant 0, \\ x(0) \ \ = x_0. \end{cases} \tag{50}$$

and

$$\begin{cases} \mathrm{d}x_\pi(t) = f\big(x_\pi(t)\big)\,\mathrm{d}t + g\big(x_\pi(t)\big)\,\mathrm{d}w_\pi(t), & t \geqslant 0, \\ x_\pi(0) \ \ = x_0. \end{cases} \tag{51}$$

The assumptions on f and g are sufficient to guarantee the existence of the solutions x and x_π, see Theorem 2. Note also that x is continuous, i.e.,

$$\mathbb{P}\{\omega \in \Omega : x(\omega) \in C(0,T;X)\} = 1. \tag{52}$$

For each $n \in \mathbb{N}$ let π_n be a partition of $[0,T]$ as described in the previous section. In particular, we assume that the conditions (22–23) are satisfied.

Theorem 4. *With the above assumptions and notation, for each $\delta > 0$*

$$\mathbb{P}\Big\{\omega : \sup_{0 \leqslant t \leqslant T} |x(t,\omega) - x_{\pi_n}(t,\omega)|_X > \delta\Big\} \longrightarrow 0 \qquad \text{as } \operatorname{mesh} \pi_n \to 0, \tag{53}$$

i.e., $x_{\pi_n}(\,.\,) \to x(\,.\,)$ in $C(0,T;X)$ in probability.

Remark 14. The extension of Corollary 1 to Theorem 4 is important as it allows us to apply this approximation result to a class of diffusion processes on loops, see Section 4.

Proof. The proof we give is analogous to a proof given in [25] (see Theorem 10, page 153). Throughout the proof we denote the norm on X by $|\,.\,|$ and the norm on $C(0,T;X)$ by $|\,.\,|_\infty$. For $R \in \mathbb{N}$ set

$$\mathcal{B}_R := \{\gamma \in C(0,T;X) : |\gamma|_\infty \leqslant R\}.$$

(52) implies that

$$1 = \mathbb{P}\Big\{\omega \in \Omega : x(\omega) \in \bigcup_{R\in\mathbb{N}} \mathcal{B}_R\Big\} = \lim_{R\to\infty} \mathbb{P}\{\omega \in \Omega : x(\omega) \in \mathcal{B}_R\}. \tag{54}$$

Set $\Omega_R := \{\omega \in \Omega : x(\omega) \in \mathcal{B}_R\}$. Let $\varepsilon > 0$ be given. (54) implies that we may choose $R \in \mathbb{N}$ so large that

$$\mathbb{P}(\Omega_R^c) = 1 - \mathbb{P}(\Omega_R) < \frac{\varepsilon}{2}. \tag{55}$$

Henceforth we keep R fixed such that (55) holds. Let

$$x_0^R(\omega) = \begin{cases} x_0(\omega) & \text{if } |x_0(\omega)|_X \leqslant R+1, \\ 0 & \text{if } |x_0(\omega)|_X > R+1. \end{cases} \tag{56}$$

Lemma 5. *For any $R > 0$ there exists C^1-class maps $f_R : X \to X$ and $g_R : X \to L(E,X)$ such that:*

(i) *f_R, g_R and g_R' are globally Lipschitz and bounded;*
(ii) *f_R and g_R coincide with f and g on the closed ball $\bar{B}(0,R+1) \subset X$.*

Proof. Fix $R > 0$ and define $\phi_R : X \to \mathbb{R}$ by

$$\phi_R(x) := \phi\Big(\frac{x}{R+1}\Big), \qquad x \in X,$$

where $\phi : X \to \mathbb{R}$ is the bump function described earlier. It is clear that ϕ_R is C^2-class, Lipschitz and bounded. ϕ_R satisfies

$$\phi_R(x) = \begin{cases} 1 & \text{if } |x|_X \leqslant R+1, \\ 0 & \text{if } |x|_X \geqslant 2R+2, \end{cases}$$

and $0 \leqslant \phi_R \leqslant 1$. Furthermore ϕ_R' is Lipschitz and bounded.

Define $f_R : X \to X$ and $g_R : X \to L(E,X)$ by

$$f_R(x) = \phi_R(x) f(x), \qquad g_R(x) = \phi_R(x)\, g(x), \qquad x \in X.$$

It is not difficult to see that f_R and g_R are C^1-maps which coincide with f and g respectively on the closed ball $\bar{B}(0,R+1)$. We are thus left with proving (i). Note the following two facts:

a) If $\gamma : X \to \mathbb{R}$ and $g : X \to X$ are Lipschitz and bounded, then so is γg.

b) If $g : X \to X$ is Lipschitz and bounded on the closed ball $\bar{B}(0,2R+2) \subset X$, then $\phi_R g$ is Lipschitz and bounded. Indeed, by Appendix, there exists a Lipschitz and bounded function $\tilde{g} : X \to X$ such that $\tilde{g} = g$ on $\bar{B}(0,2R+2)$. By *a*), $\phi_R \tilde{g}$ is Lipschitz and bounded. The equality $\phi_R \tilde{g} = \phi_R g$ concludes the proof.

The point b) implies that $f_R = \phi_R f$ and $g_R = \phi_R g$ are Lipschitz and bounded. For the same reasons the maps $\phi_R' g$ and $\phi_R g'$ are also Lipschitz and bounded. Thus, since

$$g_R' = \phi_R' g + \phi_R g',$$

it follows that g_R' is Lipschitz and bounded. This completes the proof of Lemma 5. $\square$

Denote by $x^R(t)$, $t \in [0,T]$, the unique solution to the problem

$$\begin{cases} \mathrm{d}x^R(t) = f_R\big(x^R(t)\big)\,\mathrm{d}t + g_R\big(x^R(t)\big) \circ \mathrm{d}w(t), & t \geqslant 0, \\ x^R(0) \ = x_0^R, \end{cases} \tag{57}$$

where f_R and g_R are the maps from Lemma 5. (Of course, the solution x^R exists by Theorem 2.)

Remark 15. Note that if $\omega \in \Omega_R$ then $x(\,.\,,\omega) \in \mathcal{B}_R$, i.e.,

$$|x(\,.\,,\omega)|_\infty = \sup_{0 \leqslant t \leqslant T} |x(t,\omega)| \leqslant R.$$

As $f = f_R$ and $g = g_R$ on $\bar{B}(0, R+1)$, then by uniqueness of the solutions to the problems (50) and (57), $x(t,\omega) = x^R(t,\omega)$ for all $t \in [0,T]$. It follows that

$$\big|x^R(\,.\,,\omega)\big|_\infty = \sup_{0 \leqslant t \leqslant T} \big|x^R(t,\omega)\big| \leqslant R.$$

For a partition π_n of the interval $[0,T]$, let $x_{\pi_n}^R : [0,T] \times \Omega \to X$ be the solution to

$$\begin{cases} \mathrm{d}x_{\pi_n}^R(t) = f_R\big(x_{\pi_n}^R(t)\big)\,\mathrm{d}t + g_R\big(x_{\pi_n}^R(t)\big)\,\mathrm{d}w_{\pi_n}(t), & t \geqslant 0, \\ x_{\pi_n}^R(0) \ = x_0^R. \end{cases} \tag{58}$$

Take $\delta > 0$ such that $0 < \delta < 1$. The maps f_R and g_R satisfy the assumptions of Theorem 3 and so we may apply Corollary 1. In particular, there exists $N_\varepsilon \in \mathbb{N}$ such that $\forall n \geqslant N_\varepsilon$

$$\mathbb{P}\big\{\omega \in \Omega : \big|x^R(\,.\,,\omega) - x_{\pi_n}^R(\,.\,,\omega)\big|_\infty > \delta\big\} < \frac{\varepsilon}{2}. \tag{59}$$

For $n \geqslant N_\varepsilon$ set $\Omega_{n,\delta} := \{\omega \in \Omega : |x^R(\,.\,,\omega) - x_{\pi_n}^R(\,.\,,\omega)|_\infty > \delta\}$.

Lemma 6. *If $\omega \in (\Omega_R^c \cup \Omega_{n,\delta})^c$, $n \geqslant N_\varepsilon$, then*

$$x_{\pi_n}^R(\,.\,,\omega) = x_{\pi_n}(\,.\,,\omega) \qquad \text{on } [0,T]. \tag{60}$$

Proof. Let $\omega \in (\Omega_R^c \cup \Omega_{n,\delta})^c = \Omega_R \cap \Omega_{n,\delta}^c$. As $\omega \in \Omega_R$ then, see Remark 15,

$$\big|x^R(\,.\,,\omega)\big|_\infty \leqslant R.$$

Furthermore, if $\omega \in \Omega_{n,\delta}^c$ then $|x^R(\,.\,,\omega) - x^R_{\pi_n}(\,.\,,\omega)|_\infty \leqslant \delta < 1$ and hence

$$\left|x^R_{\pi_n}(\,.\,,\omega)\right|_\infty \leqslant 1 + \left|x^R(\,.\,,\omega)\right|_\infty < 1 + R,$$

i.e., $x^R_{\pi_n}(\,.\,,\omega) \in \mathcal{B}_{R+1}$. Following the arguments in Remark 15, the uniqueness of the solutions to the problems (51) and (58) implies

$$x^R_{\pi_n}(\,.\,,\omega) = x_{\pi_n}(\,.\,,\omega) \qquad \text{on } [0,T]. \qquad \square$$

If $\omega \in (\Omega_R^c \cup \Omega_{n,\delta})^c$, $n \geqslant N_\varepsilon$, then by Lemma 6

$$|x(\,.\,,\omega) - x_{\pi_n}(\,.\,,\omega)|_\infty = \left|x^R(\,.\,,\omega) - x^R_{\pi_n}(\,.\,,\omega)\right|_\infty \leqslant \delta.$$

It follows that

$$\{\omega \in \Omega : |x(\,.\,,\omega) - x_{\pi_n}(\,.\,,\omega)|_\infty > \delta\} \subset \Omega_R^c \cup \Omega_{n,\delta},$$

which implies that

$$\mathbb{P}\{\omega \in \Omega : |x(\,.\,,\omega) - x_{\pi_n}(\,.\,,\omega)|_\infty > \delta\} \leqslant \mathbb{P}(\Omega_R^c \cup \Omega_{n,\delta}) \leqslant \mathbb{P}(\Omega_R^c) + \mathbb{P}(\Omega_{n,\delta}).$$

(55) and (59) now imply that for $n \geqslant N_\varepsilon$

$$\mathbb{P}\{\omega \in \Omega : |x(\,.\,,\omega) - x_{\pi_n}(\,.\,,\omega)|_\infty > \delta\} < \varepsilon. \tag{61}$$

We have proved that (61) holds for $0 < \delta < 1$. Clearly (61) then also holds for any $\delta > 0$, i.e., $x_{\pi_n}(\,.\,) \to x(\,.\,)$ in $C(0,T;X)$ in probability. This completes the proof of Theorem 4. $\square$

Remark 16. As it should be clear from the presented proof the condition (C2) can be replaced by a weaker one:

(C3) The problem (50) has a unique X-valued solution and for each partition π of the interval $[0,T]$, for each $\omega \in \Omega$, the problem (51) has a unique X-valued solution.

This condition, used as well in [25], will prove useful in Section 5.

Next, we will show the following result on the ‘transfer principle’.

Theorem 5. *Suppose X is an M-type 2 Banach space satisfying (D2), E is a Banach space and $w(t)$, $t \geqslant 0$, is an E-valued Wiener process on some filtered probability space $\mathfrak{A} = (\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$. Suppose that $\mathcal{M}$ is a closed submanifold of the Banach space X. Suppose that in addition to the conditions (A2), (B2) and (C3) the following condition is satisfied*

(M3) For each $x \in \mathcal{M}$, $f(x) \in T_x\mathcal{M}$ and $\operatorname{range} g(x) \subset T_x\mathcal{M}$.

If $x_0 \in \mathcal{M}$, then the solution x to (50) takes values in $\mathcal{M}$, a.s.

Proof. Let us fix $T > 0$. It follows from (53) that for some subsequence of the sequence π_n, still denoted by π_n to avoid too complicated notation, one has $\sup_{0\leqslant t\leqslant T} |x(t) - x_{\pi_n}(t)|_X \to 0$ a.s. On the other hand, from classical analysis it is known (due to (M3)), that the solution $x_{\pi_n}(t)$ takes values in $\mathcal{M}$ for all $t \in [0, T]$. Therefore, as $\mathcal{M}$ is closed, the result follows. □

We conclude this section with the following result on approximation when the solutions may blow up. However, we do not study as in [10] the case when all the coefficients are locally Lipschitz in a weak sense but assume that the conditions (A2) and (B2) hold true. What we do not assume is (C2) and (C3). The result we present can be seen as a generalization of Theorem VII.10 from Elworthy's book [25] from a Hilbert setting to a Banach one. Yet, it is a weaker result because our coefficients are Lipschitz on balls. A technical reason for this drawback of our result lies in the fact that we do not know if a Lipschitz map from a subset of a Banach space has a Lipschitz extension to the whole space (possibly with a bigger Lipschitz constant), see also Remark 24.

Theorem 6. *Suppose X is a Banach space satisfying (D2), E is a Banach space and $w(t)$, $t \geqslant 0$, is an E-valued Wiener process on some filtered probability space $\mathfrak{A} = (\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$. Suppose finally that the conditions (A2) and (B2) hold true. Let $x(t)$, $0 \leqslant t < \tau$ be the maximal solution to* (50) *and let $x_{\pi_n}(t)$, $0 \leqslant t < \tau_n$ be a family of the maximal solutions to the family of ordinary differential equations (51). Then x_{π_n} converges to x in measure in the sense that for each $t \geqslant 0$ and each $\delta > 0$,*

$$\mathbb{P}\Big\{\omega \in \Omega_t(\tau) : \sup_{0\leqslant s\leqslant t} |x(s,\omega) - x_{\pi_n}(s,\omega)|_X > \delta\Big\} \longrightarrow 0 \qquad \text{as } n \to \infty, \tag{62}$$

where $\Omega_t(\tau) = \{\omega \in \Omega : t < \tau(\omega)\}$. In particular, $t \wedge \tau_{\pi_n} \to t$ in measure on Ω_t as $n \to \infty$.

Proof. Our argument is principally a modification of the proof of Theorem VII.10 from [25]. Let us fix $\varepsilon > 0$ and $t > 0$. As in the proof of Theorem 4 we can find $R > 0$ and a measurable subset Ω_R of $\Omega_t(\tau)$ such that

$$x(s,\omega) \in \mathcal{B}_R, \quad \text{for } (s,\omega) \in [0,t] \times \Omega_R, \qquad \text{and} \qquad \mathbb{P}(\Omega_R^c) < \frac{\varepsilon}{2}. \tag{63}$$

Then we define the initial condition x_0^R by (56) and using Lemma 5 find C^1-class functions $f_R : X \to X$ and $g_R : X \to L(E, X)$ satisfying the conditions (i) and (ii) of that Lemma. Next we denote by x_R the global solution to the problem (57). As in Remark 15 we infer that by the uniqueness of solutions, $\sup_{0\leqslant s\leqslant t} |x^R(t,\omega)| \leqslant R$, for $\omega \in \Omega_t(\tau)$. Next, for $n \in \mathbb{N}$, let $x^R_{\pi_n} : [0, T] \times \Omega \to X$ be the solution to (58). Take next $\delta \in (0, 1)$. From Corollary 1 we infer that there exists $N_\varepsilon \in \mathbb{N}$ such that $\forall n \geqslant N_\varepsilon$ the inequality (59) holds true. For $n \geqslant N_\varepsilon$ set $\Omega_{n,\delta,t}(\tau) := \{\omega \in \Omega(\tau) : \sup_{0\leqslant s\leqslant t} |x^R(s,\omega) - x^R_{\pi_n}(s,\omega)| > \delta\}$. Then, arguing as in the proof of Lemma 6 we have that if $\omega \in \Omega_R \setminus \Omega_{n,\delta,t}(\tau)$ then

$$x^R_{\pi_n}(\,.\,,\omega) = x_{\pi_n}(\,.\,,\omega) \qquad \text{on } [0,t]. \tag{64}$$

Hence, if $\omega \in \Omega_R \setminus \Omega_{n,\delta,t}(\tau)$,

$$\sup_{0\leqslant s\leqslant t} \left|x(s,\omega) - x_{\pi_n}(s,\omega)\right| = \sup_{0\leqslant s\leqslant t} \left|x^R(s,\omega) - x^R_{\pi_n}(s,\omega)\right| \leqslant \delta$$

and therefore

$$\Big\{\omega \in \Omega_t(\tau) : \sup_{0\leqslant s\leqslant t} |x(s,\omega) - x_{\pi_n}(s,\omega)| > \delta\Big\} \subset \Omega_R^c \cup \Omega_{n,\delta,t}(\tau).$$

This, together with (63) and (64) implies that for $n \geqslant N_\varepsilon$

$$\mathbb{P}\Big\{\omega \in \Omega_t(\tau) : \sup_{0\leqslant s\leqslant t} |x(s,\omega) - x_{\pi_n}(s,\omega)| > \delta\Big\} < \varepsilon \tag{65}$$

what proves the theorem. □

4 Applications to diffusion processes on loop spaces

4.1 Diffusion processes on loop manifolds

In this first subsection we briefly outline recent results of Brzeźniak and Elworthy concerning the existence of diffusion processes as solutions to Stratonovitch stochastic differential equations on certain loop manifolds. All of what we present (unless otherwise stated) can be found in [10].

Let M be a smooth compact riemannian manifold. We imbed M into some Euclidean space $\mathbb{R}^d$ and identify M with its image. Let S^1 denote the unit circle.

For $\theta \in (0,1)$, $p \geqslant 1$, the Sobolev–Slobodetskii space of loops on $\mathbb{R}$, $W^{\theta,p}(S^1,\mathbb{R})$, is defined by

$$W^{\theta,p}(S^1,\mathbb{R}) := \left\{u \in L^p(S^1,\mathbb{R}) : \int_{S^1}\int_{S^1} \frac{|u(s_1)-u(s_2)|^p}{|s_1-s_2|^{1+\theta p}}\,\mathrm{d}s_1\,\mathrm{d}s_2 < \infty\right\}.$$

The vector space $W^{\theta,p}(S^1,\mathbb{R})$ is a Banach space with the norm

$$|u|_{\theta,p} := \int_{S^1} |u(s)|^p\,\mathrm{d}s + \int_{S^1}\int_{S^1} \frac{|u(s_1)-u(s_2)|^p}{|s_1-s_2|^{1+\theta p}}\,\mathrm{d}s_1\,\mathrm{d}s_2.$$

Furthermore, $W^{\theta,p}(S^1,\mathbb{R})$ is an M-type 2 Banach space and $W^{\theta,p}(S^1,\mathbb{R})$ satisfies the condition (H_p), see Section 3.

Remark 17. The spaces $W^{\theta,p}(S^1,\mathbb{R})$ may be considered as intermediate spaces lying between $L^p(S^1,\mathbb{R})$ and $W^{1,p}(S^1,\mathbb{R})$, where $W^{1,p}(S^1,\mathbb{R})$ is the space of loops on $\mathbb{R}$ whose first weak derivative lies in $L^p(S^1,\mathbb{R})$. Indeed, $W^{\theta,p}(S^1,\mathbb{R})$ may be identified with the real interpolation space

$$\big(L^p(S^1,\mathbb{R}), W^{1,p}(S^1,\mathbb{R})\big)_{\theta,p}.$$

See [14] for more details.

We say that a loop on $\mathbb{R}^d$, $u : S^1 \to \mathbb{R}^d$, belongs to the Sobolev–Slobodetskii space $W^{\theta,p}(S^1,\mathbb{R}^d)$ if and only if the coordinate functions $u^j : S^1 \to \mathbb{R}$ belong to $W^{\theta,p}(S^1,\mathbb{R})$ for $j = 1,\dots,d$.

Henceforth we choose θ to lie in the interval $(1/p, 1/2), p > 2$. In particular, this implies:

(i) the imbedding map $W^{\theta,p}(S^1,\mathbb{R}^d) \hookrightarrow C(S^1,\mathbb{R}^d)$ is continuous, where $C(S^1,\mathbb{R}^d)$ is the space of continuous loops on $\mathbb{R}^d$;
(ii) $i : H^{1,2}(S^1,\mathbb{R}^d) \hookrightarrow W^{\theta,p}(S^1,\mathbb{R}^d)$ is an AWS, where $H^{1,2}(S^1,\mathbb{R}^d)$ is the space of loops on $\mathbb{R}^d$ whose first weak derivative belongs to $L^2(S^1,\mathbb{R}^d)$.

Using the notation from the previous sections we set

$$X = E = W^{\theta,p}(S^1,\mathbb{R}^d) \qquad \text{and} \qquad H = H^{1,2}(S^1,\mathbb{R}^d).$$

We denote by $\{w(t)\}_{t\geqslant 0}$ the corresponding E-valued Wiener process. In view of (i), the paths of $w(t)$, $t \geqslant 0$, take values in $C(S^1,\mathbb{R}^d)$. In particular, for each $s \in S^1$, $w_s(t) := w(t)(s)$, $t \geqslant 0$, is an $\mathbb{R}^d$-valued Wiener process.

For $x : [0,T] \times S^1 \times \Omega \to M$ consider the following family (indexed by $s \in S^1$) of Stratonovitch stochastic differential equations on M:

$$\mathrm{d}x_s(t) = f\big(x_s(t)\big)\,\mathrm{d}t + g\big(x_s(t)\big) \circ \mathrm{d}w_s(t), \tag{66}$$

$t > 0$, $s \in S^1$, where we write $x_s(t) := x(t,s)$ and we have suppressed the dependence on $\omega \in \Omega$. We explain the notation used in (66):

a) $f \in C^\infty(M, TM)$, i.e., f is a smooth vector field on M;

b) $g \in C^\infty(M, L(\mathbb{R}^d, TM))$, i.e., g is a smooth section of a bundle $\mathbb{F}$ over M, whose fibres are $\mathbb{F}_x = L(\mathbb{R}^d, T_xM)$, $x \in M$. (Here TM is the tangent bundle of M and for $x \in M$, T_xM is the tangent space to M at x.)

Instead of considering the above family of SDEs on M, we reformulate (66) as a single SDE on a certain loop manifold. We define $\mathcal{M} = W^{\theta,p}(S^1, M)$ by

$$W^{\theta,p}(S^1,M) := \big\{u \in W^{\theta,p}(S^1,\mathbb{R}^d) : u(s) \in M,\ \forall s \in S^1\big\}.$$

In view of (i) $\mathcal{M}$ is well defined. Moreover, $\mathcal{M}$ is a closed submanifold of the infinite dimensional Banach space $W^{\theta,p}(S^1,\mathbb{R}^d)$, see [8]. The tangent space to $\mathcal{M}$ at a point $\gamma \in \mathcal{M}$ is given by

$$T_\gamma\mathcal{M} = \big\{\eta \in W^{\theta,p}\big(S^1,\mathbb{R}^d\big) : \eta(s) \in T_{\eta(s)}M,\ \forall s \in S^1\big\}.$$

Let $\gamma \in \mathcal{M}$, $\eta \in E = W^{\theta,p}(S^1,\mathbb{R}^d)$, $s \in S^1$. Given f and g as above, we define their corresponding Nemytski maps F and G through the following formulas

$$F(\gamma)(s) := f\big(\gamma(s)\big), \tag{67}$$

$$G(\gamma)(\eta)(s) := g\big(\gamma(s)\big)\,\eta(s). \tag{68}$$

In particular, F and G are C^∞ maps which satisfy

$$\mathcal{M} \ni \gamma \longmapsto F(\gamma) \in T_\gamma \mathcal{M},$$
$$\mathcal{M} \ni \gamma \longmapsto G(\gamma) \in L(E, T_\gamma \mathcal{M}).$$

Using the above notation we may rewrite the family of SDEs (66) as

$$\mathrm{d}x(t) = F\big(x(t)\big)\,\mathrm{d}t + G\big(x(t)\big) \circ \mathrm{d}w(t). \tag{69}$$

The equation (69) is a SDE on the loop manifold $\mathcal{M} = W^{\theta,p}(S^1, M)$. If the initial value lies on $\mathcal{M}$ then there exists a unique global $\mathcal{M}$-valued solution to (69), i.e., if $x_0 \in \mathcal{M}$ then for any $T > 0$, there exists a unique continuous, progressively measurable $\mathcal{M}$-valued process x such that for each $t \in [0, T]$

$$x(t) = x(0) + \int_0^t F\big(x(r)\big)\,\mathrm{d}r + \int_0^t G\big(x(r)\big) \circ \mathrm{d}w(r), \qquad \text{a.s.,}$$

with $x(0) = x_0$. In particular, x is a diffusion process on the loop manifold $\mathcal{M}$.

Remark 18. The family of M-valued processes $\{x_s(t)\}_{s \in S^1}$, $t \in [0, T]$, is a solution to (66), with initial value $\{x_{0,s}\}_{s \in S^1}$.

4.2 An approximation result for solutions to SDEs on $\mathcal{M}$

Fix $T > 0$ and $x_0 \in \mathcal{M}$. Let $x : [0, T] \times \Omega \to \mathcal{M}$ be the unique solution to the problem

$$\begin{cases} \mathrm{d}x(t) = F\big(x(t)\big)\,\mathrm{d}t + G\big(x(t)\big) \circ \mathrm{d}w(t), \\ x(0) \ = x_0, \end{cases} \tag{70}$$

where F and G are given by (67) and (68) respectively. It is known, see [10] that the problem (70) has a unique global solution. Moreover, in [12], the Feller property of solutions is investigated. For each partition π of $[0, T]$, let $x_\pi : [0, T] \times \Omega \to \mathcal{M}$ be the solutions to the family of ODEs (indexed by $\omega \in \Omega$)

$$\begin{cases} \mathrm{d}x_\pi(t) = F\big(x_\pi(t)\big)\,\mathrm{d}t + G\big(x_\pi(t)\big)\,\mathrm{d}w_\pi(t), \\ x_\pi(0) \ = x_0. \end{cases} \tag{71}$$

Theorem 7. *Suppose π_n is a sequence of partitions of the interval $[0, T]$ satisfying the conditions* (22–23). *Then $x_{\pi_n}(\,.\,) \to x(\,.\,)$ in $C(0, T; \mathcal{M})$ in probability.*

Proof. Let f and g be defined as in the previous subsection. We can extend f and g smoothly to be defined on the whole of $\mathbb{R}^d$ so that they are both of compact support. We denote these extensions $\hat{f}$ and $\hat{g}$ respectively. For each $m \in M$, we identify the tangent space $T_m M$ with the corresponding subspace of $\mathbb{R}^d$. In particular

$$\hat{f} : \mathbb{R}^d \to \mathbb{R}^d \qquad \text{and} \qquad \hat{g} : \mathbb{R}^d \to L(\mathbb{R}^d, \mathbb{R}^d),$$

and, for $m \in M$,

$$\hat{f}(m) = f(m), \tag{72}$$
$$\hat{g}(m) = g(m). \tag{73}$$

Given $\hat{f}$ and $\hat{g}$ we define their corresponding Nemytski maps $\widehat{F}$ and $\widehat{G}$ through the following formulas

$$\widehat{F}(\gamma)(s) := \hat{f}\big(\gamma(s)\big), \tag{74}$$
$$\widehat{G}(\gamma)(\eta)(s) := \hat{g}\big(\gamma(s)\big)\eta(s), \tag{75}$$

where $\gamma \in X = W^{\theta,p}(S^1, \mathbb{R}^d)$, $\eta \in E = W^{\theta,p}(S^1, \mathbb{R}^d)$ and $s \in S^1$. The maps $\widehat{F}$ and $\widehat{G}$ are smooth maps which satisfy

$$\widehat{F} : X \to X,$$
$$\widehat{G} : X \to L(E, X).$$

The maps $\widehat{F}$, $\widehat{G}$ and $\mathrm{tr}(\widehat{G}'\widehat{G})$ are Lipschitz continuous on balls and are of linear growth. (In fact, all the derivatives of $\widehat{F}$ and $\widehat{G}$ are Lipschitz on balls.) Thus, see Theorem 2, given $x_0 \in \mathcal{M} \subset X$ there exists a unique global X-valued solution to the problem

$$\begin{cases} \mathrm{d}\hat{x}(t) = \widehat{F}\big(\hat{x}(t)\big)\,\mathrm{d}t + \widehat{G}\big(\hat{x}(t)\big) \circ \mathrm{d}w(t), \\ \hat{x}(0) \;\; = x_0. \end{cases} \tag{76}$$

For a partition π let $\hat{x}_\pi : [0,T] \times \Omega \to X$ be the solution to the family of ODEs, indexed by $\omega \in \Omega$,

$$\begin{cases} \mathrm{d}\hat{x}_\pi(t) = \widehat{F}\big(\hat{x}_\pi(t)\big)\,\mathrm{d}t + \widehat{G}\big(\hat{x}_\pi(t)\big)\,\mathrm{d}w_\pi(t) \\ \hat{x}_\pi(0) \;\; = x_0. \end{cases} \tag{77}$$

The conditions of Theorem 4 are satisfied and so we deduce that

$$\hat{x}_{\pi_n}(\,.\,) \to \hat{x}(\,.\,) \qquad \text{in } C(0,T;X) \text{ in probability.} \tag{78}$$

However, note that if $\gamma \in \mathcal{M}$ then for each $s \in S^1$, $\gamma(s) \in M$ and so

$$\widehat{F}(\gamma)(s) = \hat{f}\big(\gamma(s)\big) = f\big(\gamma(s)\big) = F(\gamma)(s),$$
$$\widehat{G}(\gamma)(s) = \hat{g}\big(\gamma(s)\big) = g\big(\gamma(s)\big) = G(\gamma)(s)$$

(see (72), (73) and the definitions of the Nemytski maps, (67), (68), (74) and (75)). So, if $\gamma \in \mathcal{M}$, we have

$$\widehat{F}(\gamma) = F(\gamma) \qquad \text{and} \qquad \widehat{G}(\gamma) = G(\gamma).$$

Thus, the $\mathcal{M}$-valued solution to (70) is also the solution to the SDE (76), provided that we take the same initial value $x_0 \in \mathcal{M}$. Hence, by uniqueness we have that $\hat{x} = x$. A similar argument yields $\hat{x}_{\pi_n} = x_{\pi_n}$ for each $n \in \mathbb{N}$. It follows from this observation and (78) that $x_{\pi_n}(\,.\,) \to x(\,.\,)$ in $C(0,T;\mathcal{M})$ in probability. This completes the proof of Theorem 7. □

Remark 19. The argument above can be also used to give an alternative proof of existence of solutions to (69). We first prove, as in [10], the existence of global solutions to (76). Then, noting that due to our construction the coefficients $\widehat{F}$ and $\widehat{H}$ satisfy the assumption (M3) of Theorem 5. Applying then the last result implies that the solution $x(t)$ to (76) takes values in $\mathcal{M}$. Therefore, as the restrictions of $\hat{F}$ and $\hat{H}$ to $\mathcal{M}$ are simply F and G respectively, we infer that $x(t)$ is a solution to (69).

Remark 20. If in addition to (66) one considers a family of random ODEs:

$$\begin{cases} \mathrm{d}x_s^{\pi_n}(t) = f\big(x_s^{\pi_n}(t)\big)\,\mathrm{d}t + g\big(x_s^{\pi_n}(t)\big)\,\mathrm{d}w_s^{\pi_n}(t), \; t > 0, \\ x_s^{\pi_n}(0) \;\; = x_0(s), \end{cases}$$

indexed by $s \in S^1$, then $x^{\pi_n} \to x$ in probability in the following sense. For all $\varepsilon > 0$

$$\mathbb{P}\Big\{ \sup_{s \in S^1, t \in [0,T]} |x_s(t) - x_s^{\pi_n}(t)|_{\mathbb{R}^d} > \varepsilon \Big\} \longrightarrow 0 \qquad \text{as mesh } \pi_n \to 0. \tag{79}$$

5 Applications to stochastic flows

Suppose that M is a compact smooth riemannian manifold of dimension m. P. Baxendale in [2] defined a $\mathrm{Diff}^r(M)$-valued, $r = 1, 2, \ldots, \infty$, Brownian Motion (BM) and showed that an $\mathrm{Diff}^r(M)$-valued BM, $r = 3, 4, \ldots, \infty$, generates a Hilbert space $H \subset C^{r-2}(TM)$ and a vector field $f \in C^{r-3}(TM)$, where TM denotes the tangent vector bundle on M and $C^k(TM)$ the space of all sections of TM (i.e., vector fields on M) of class C^k. In what follows, $C^{k,1}(TM)$ will denote the space of all sections of TM of class C^k such that the k-th derivative $f^{(k)}$ is Lipschitz. Converse results were known through works of Elworthy, see [25] and Kunita [30]. In a recent paper [9] the authors proved the following. Suppose $H \subset C^{2,1}(TM)$ is such that the natural imbedding $H \hookrightarrow C^{2,1}(TM)$ is γ-radonifying and that $f \in C^{1,1}(TM)$. Let $\mathbb{H}$ be a vector bundle over M with a fiber at x to be equal $\mathcal{L}(H, T_xM)$. Define a section g of the bundle $\mathbb{H}$ by $g(x)(h) = h(x)$. The natural extension of g to $C^{2,1}(TM)$ we will denote also by g. Then, for $\theta \in (m/p, 1)$ there exists a global stochastic flow of $W^{\theta+1,p}$ diffeomorphisms of M to the problem (with $x \in M$):

$$\begin{cases} \mathrm{d}x(t) = f\big(x(t)\big)\,\mathrm{d}t + g\big(x(t)\big)\circ \mathrm{d}w(t), \\ x_0 = x. \end{cases} \tag{80}$$

Here we assume that $w(t)$, $t \geqslant 0$ is the canonical E-valued Wiener process defined on some complete filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$, where E is some separable Banach space such that $H \hookrightarrow E \subset C^{2,1}(TM)$ with the first imbedding being an AWS.

The construction of a stochastic flow of diffeomorphisms for (80) can be achieved by lifting the problem (80) to a stochastic differential equation on the group of diffeomorphisms of M (of appropriate regularity). This has been first done by Elworthy in a 'Hilbertian' framework, see [25], and later developed by the first author and Elworthy in a 'Banachian' framework, see [9]. Here we will follow the second reference. Let us choose θ and $p \geqslant 2$ such that $m/p < \theta < 1$. Let X be the Sobolev–Slobodetskii space $W^{\theta+1,p}(M, \mathbb{R}^d)$, where we assume that M is imbedded into $\mathbb{R}^d$ (and so, in particular, $f(x) \in \mathbb{R}^d$ for $x \in M$). Next, let $\mathcal{M}$ be the Banach manifold $W^{\theta+1,p}(M, M)$, see [8] and references therein for a definition and all basic properties. Finally, let $\mathrm{Diff}^{1+\theta,p}(M)$ be the open set in $\mathcal{M}$ consisting of all $\phi \in \mathcal{M}$ which are $\mathcal{C}^1$ diffeomorphisms of M. With the maps F and G defined by $F : X \ni u \mapsto f \circ u \in X$ and $G : X \to \mathcal{L}(H, X)$, $G(u)(h) = \{M \ni x \mapsto g(x)(h) \in \mathbb{R}^d\}$, the equation (80) lifted to the Banach space X takes the following form

$$\begin{cases} \mathrm{d}u(t) = F\big(x(t)\big)\,\mathrm{d}t + G\big(u(t)\big)\circ \mathrm{d}w(t) \\ u_0 = \mathrm{Id}, \end{cases} \tag{81}$$

where Id is the identity map of M. One proves by essentially the same methods as in [9] the following result

Theorem 8. *The problem* (81) *has a unique X-valued solution $u(t)$, $t \geqslant 0$. This process takes values in $\mathcal{M}$, and in fact is* $\mathrm{Diff}^{1+\theta,p}(M)$*-valued.*

The first part of the Theorem is proven first for the Banach space $\widehat{X} = W^{\theta,p}(M, \mathbb{R}^d)$. Then by studying an equation for the derivative flow one can show that the solution $u(t)$ actually takes values in X. The reason for proving this part of the Theorem in two steps lies in the fact that while the maps F and G are of linear growth in $\widehat{X}$ they are not of linear growth in E. They are locally Lipschitz in both cases.

The second part of the Theorem follows from invariance of the manifold $\mathcal{M}$ with respect to the problem (81) (in a similar way to [10]). The third part follows from the second by employing an ergodic type argument due to Itô (and used in a similar context by Elworthy in [26]).

As in the previous section we fix a positive $T > 0$ and for each partition π of the interval $[0, T]$ we consider a family of ODEs (indexed by $\omega \in \Omega$) on M:

$$\begin{cases} \mathrm{d}x_\pi(t) = f\big(x_\pi(t)\big)\,\mathrm{d}t + g\big(x_\pi(t)\big)\,\mathrm{d}w_\pi(t), \\ x_\pi(0) \;\; = x_0 \end{cases} \tag{82}$$

and its lift to X:

$$\begin{cases} \mathrm{d}u_\pi(t) = F\big(u_\pi(t)\big)\,\mathrm{d}t + G\big(u_\pi(t)\big)\,\mathrm{d}w_\pi(t) \\ u_\pi(0) \;\; = \mathrm{Id}. \end{cases} \tag{83}$$

It is a classical result that for each $\omega \in \Omega$ the (ordinary) differential equation (83) has a unique E-valued solution. Moreover this solution takes values in $\mathrm{Diff}^{1+\theta,p}(M)$. Similarly to Theorem (7) (but see also Remark 16) we have

Theorem 9. *Suppose π_n is a sequence of partitions of the interval $[0,T]$ satisfying the conditions* (22–23)*. Then $x_{\pi_n}(\,.\,) \to x(\,.\,)$ in $C(0,T;\mathrm{Diff}^{1+\theta,p}(M))$ in probability.*

Remark 21. The distance in $\mathrm{Diff}^{1+\theta,p}(M)$ is the distance inherited from the distance in $X = W^{1+\theta,p}(M,\mathbb{R}^d)$.

Remark 22. In a Hilbert manifolds framework Theorem 9 was stated and proved much earlier by K.D. Elworthy in [24] and [25], Corollary 1C.1 of chapter VIII. Comparing with [25] our results cover more general driving processes and, even in the more classical description of Elworthy, we allow vector fields to be of lower regularity. It is possible to apply Theorem 9 to get approximation of C^∞ flows, compare with Corollary 1C.3 therein.

One should also bring to the attention of the reader that convergence of stochastic flows was also stated by Malliavin in [36], see Theorem 3.3.2.1 on p. 91. However, a detailed proof of this result has not been provided by the author until his monograph [38], where the author works with C^∞-vector fields, see Theorem 6.2 therein. Malliavin proposed there a different approach to the question of approximation of stochastic flows based on mixture of arguments and techniques from [37] and [25] and proved tightness of the sequence of diffeomorphism flows corresponding to the approximated equation. He also identified the limit as a flow corresponding to the Stratonovitch equation. His approach seem to be more deterministic when compared to ours.

One should not forget to mention that by partially employing classical approximation results of Bismut [5], Malliavin and Nualart in [39] have given a quasi-sure version of this result.

We are grateful to the anonymous referee for pointing these facts out to us.

Example 1. Suppose that M is a compact manifold and $f : M \to TM$ is vector fields on M of $\mathcal{C}^{1,1}$ and $\sigma_j : M \to TM$, $j = 1,\ldots,k$ are a finite number vector fields on M of $\mathcal{C}^{2,1}$ class. Define H to be the finite dimensional Hilbert space spanned by σ_j, $j = 1,\ldots,k$ with a image norm, i.e.,

$$\|h\|^2 = \sum_j |y_j|^2, \qquad h = \sum_{j=1}^k y_j\sigma_j, \qquad y \in \mathbb{R}^k.$$

Note, that $\sigma_1, \dots, \sigma_k$ is an ONB of H. We use the notation (introduced earlier) of $\mathbb{H}$ and g, i.e., $\mathbb{H}$ denotes the vector bundle over M with a fiber at x equal $\mathcal{L}(H, T_xM)$ and g is a section of $\mathbb{H}$ by $g(x)(h) = h(x) = \sum_j y_j \sigma_j(x)$, $h = \sum_{j=1}^k y_j \sigma_j$. We consider an $\mathbb{R}^k$ valued Wiener process $w(t) = (w^1(t), \dots, w^k(t))$, $t \geqslant 0$, defined on some complete filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$. Then, with $E = H$, $w(t) := \sum w^j(t)\sigma_j$ is the E-valued canonical Wiener process.

It can be shown that, for a fixed $x \in M$, a solution the following stochastic differential equation on M,

$$\begin{cases} \mathrm{d}x(t) = f\big(x(t)\big)\,\mathrm{d}t + \sum_{j=1}^k \sigma\big(x(t)\big) \circ \mathrm{d}w^j(t), \\ x_0 = x \end{cases} \tag{84}$$

is also a solution to the problem (80) (and vice versa). Thus our results are, in particular, applicable to standard finite dimensional stochastic differential equations of the form (84). One should point out that Elworthy's results from [25] are also applicable in this situation. However, we allow coefficients of lower regularity.

Appendix

The aim of this section is to prove the following well known results.

Lemma 7. *Let X be a normed vector space with norm by $|\,.\,|$. Define $\psi : X \to X$ through the formula*

$$\psi(x) = \begin{cases} x & \text{if } |x| \leqslant 1, \\ x/|x| & \text{if } |x| > 1. \end{cases} \tag{85}$$

Then for all x, $y \in X$

$$|\psi(x) - \psi(y)| \leqslant 3|x - y|. \tag{86}$$

Corollary 3. *Let X and Y be normed vector spaces with norms denoted by $|\,.\,|$. Suppose that a map $g : X \to Y$ is Lipschitz on the closed ball $\bar{B}(0,R)$, $R > 0$, with Lipschitz constant C. Then, there exists a bounded map $\tilde{g} : X \to Y$ such that $\tilde{g} = g$ on $\bar{B}(0,R)$ and $\tilde{g}$ is Lipschitz on X, with Lipschitz constant $3C$.*

Proof of Lemma 7. Let $B := \bar{B}(0,1) = \{x \in X : |x| \leqslant 1\}$. Clearly, from the definition of ψ, (86) holds for x, $y \in B$. There are two other cases which need to be considered.

Case 1. Let $x \notin B$ and $y \notin B$. Then

$$
\begin{aligned}
|\psi(x)-\psi(y)| &= \left|\frac{x}{|x|}-\frac{y}{|y|}\right| \\
&= \frac{\big|x|y|-y|x|\big|}{|x|\,|y|} \\
&= \frac{\big|(x-y)|y|+(|y|-|x|)y\big|}{|x|\,|y|} \\
&\leqslant \frac{|x-y|}{|x|}+\frac{\big||y|-|x|\big|}{|x|} \\
&\leqslant \frac{2}{|x|}\,|x-y| \qquad (87)
\end{aligned}
$$

As $|x|>1$, then $2/|x|<2$. Thus (86) follows from (87).

Remark 23. Note, in particular, that (87) implies

$$
\frac{\big|x\,|y|-y\,|x|\big|}{|y|} \leqslant 2|x-y|, \qquad x,y\in X. \tag{88}
$$

Case 2. Let $x\in B$ and $y\notin B$. Then

$$
|\psi(x)-\psi(y)| = \left|x-\frac{y}{|y|}\right| \leqslant \left|x-|x|\,\frac{y}{|y|}\right| + \left||x|\,\frac{y}{|y|}-\frac{y}{|y|}\right|. \tag{89}
$$

Considering the first term on the RHS of (89), then

$$
\left|x-|x|\,\frac{y}{|y|}\right| = \frac{\big||y|\,x-|x|\,y\big|}{|y|} \leqslant 2|x-y|, \tag{90}
$$

where we have applied (88). Considering now the second term on the RHS of (89), then

$$
\left||x|\,\frac{y}{|y|}-\frac{y}{|y|}\right| = \left|(|x|-1)\frac{y}{|y|}\right| = \big|1-|x|\big| \leqslant \big||y|-|x|\big| \leqslant |x-y|. \tag{91}
$$

It follows from (89), (90) and (91) that for $x\in B$ and $y\notin B$ we have

$$
|\psi(x)-\psi(y)| \leqslant 3|x-y|.
$$

This completes the proof of Lemma 7. □

Proof of Corollary 3. For $R>0$ define $\psi_R : X\to X$ through the formula

$$
\psi_R(x) = R\,\psi\Big(\frac{x}{R}\Big), \qquad x\in X,
$$

where $\psi : X\to X$ is given by (85) (see Lemma 7). Then

$$\psi_R(x) = \begin{cases} x & \text{if } |x| \leqslant R, \\ x/|x| & \text{if } |x| > R. \end{cases}$$

Clearly, from (86) we have that

$$|\psi_R(x) - \psi_R(y)| = R\Big|\psi\Big(\frac{x}{R}\Big) - \psi\Big(\frac{y}{R}\Big)\Big| \leqslant 3|x-y|. \tag{92}$$

Thus ψ_R is Lipschitz. Note also that $\psi_R\,(X) \subseteq \bar{B}(0,R)$.

Set $\tilde{g} := g \circ \psi_R$. Then, $\tilde{g} : X \to X$ is well defined and coincides with g on the closed ball $\bar{B}(0,R)$. Since g is Lipschitz, and hence bounded on $\bar{B}(0,R)$, and $\psi_R : X \to \bar{B}(0,R)$, we infer that $\tilde{g}$ is bounded. Moreover, $\tilde{g}$ is Lipschitz. Indeed, for x, $y \in X$, the Lipschitz property of g and (92) imply that

$$\begin{aligned} |\tilde{g}(x) - \tilde{g}(y)| &= |g \circ \psi_R(x) - g \circ \psi_R(y)| \\ &\leqslant C|\psi_R(x) - \psi_R(y)| \\ &\leqslant 3C|x-y|. \end{aligned}$$

This completes the proof of Corollary 3. □

Remark 24. A pair (X,Y) of Banach spaces is said to have the Contraction Extension Property (CEP) iff for any subset A of X and any Lipschitz map $f : A \to Y$, there exists a Lipschitz map $\tilde{f} : X \to Y$ such that the restriction of $\tilde{f}$ to the set A equals f and the Lipschitz constant of $\tilde{f}$ equals to the Lipschitz constant of f. A space X has the CEP iff the pair (X,X) has it. It is well known, see Kirszbraun [29], that any pair (X,Y) of Hilbert spaces has the CEP. It is also known, see Schönbeck [47] (and Theorem 2.11 in [4]) that if a strictly convex Banach space X has CEP, then it is a Hilbert space. One should emphasize here, that although the M-type 2 (i.e, 2-uniformly smooth) Banach spaces are strictly convex (possibly with an equivalent norm), there is no contradiction between Corollary 3 and the above result of Schönbeck and Benyamini–Lindenstrauss. Indeed, we prove existence of a Lipschitz extension with the Lipschitz constant being 3 times the Lipschitz constant of the original map. Furthermore, our set A is only a ball and we do not know if Corollary 3 holds true for general sets A.

Acknowledgement. The research of the second named author was supported by an EPSRC Earmarked studentship while he was at the Department of Mathematics, The University of Hull. The authors would like to thank David Elworthy for the interest in this work and, in particular, for informing them about the Dowell thesis [23].

References

1. Bass, R., Hambly, B. and Lyons T., *Extending the Wong–Zakai theorem to reversible Markov processes*, J. Eur. Math. Soc. (JEMS) **4**, no. 3, pp. 237–269 (2002).

2. Baxendale, P., *Brownian Motions in the Diffeomorphism Group I*, Composito Mathematica, **53**, pp. 19–50 (1984).
3. Belopolskaya, Ya.L. and Daletskii, Yu.L., *Diffusion processes in smooth Banach manifolds. I*, Trans. Moscow Math. Soc. **1**, pp. 113–150 (1980). Russian original: Trudy Moskov. Mat. Obsh. **37**, pp. 107–141 (1978).
4. Benyamini, Y. and Lindenstrauss, J., Geometric nonlinear functional analysis. Volume 1. Colloquium Publications. American Mathematical Society. 48. Providence, RI: American Mathematical Society (AMS), (2000).
5. Bismut, J.M., Mécanique aléatoire, LNM **866**, Springer-Verlag, Berlin, 1981.
6. Brzeźniak, Z., *Stochastic Partial Differential Equations in M-type* 2 *Banach spaces,* Potential Analysis **4**, no.1, pp. 1–47, 1995.
7. Brzeźniak, Z., *Stochastic Convolution in Banach spaces,* Stochastics and Stochastics Reports **61**, pp. 245–295, 1997.
8. Brzeźniak, Z., *Manifold Structure of some Functional spaces: on the axiomatic approach of Palais,* manuscript (in preparation).
9. Brzeźniak, Z. and Elworthy, K.D., *Stochastic Flows of Diffeomorphisms for Stochastic Differential Equations,* pp. 107-138 in Proceedings of the Fifth Gregynog Stochastic Analysis Symposium Stochastic Analysis and Applications, Gregynog, Wales (1995), edts. I.M. Davies, A. Truman and K.D Elworthy, World Scientific, Singapore, 1996.
10. Brzeźniak, Z. and Elworthy, K.D., *Stochastic Differential Equations on Banach Manifolds; applications to diffusions on loop spaces*, MFAT (a special volume dedicated to the memory of Professor Yuri Daletski), **6**, no.1, pp. 43–84 (2000).
11. Brzeźniak, Z. and Flandoli, F., *Almost sure approximation of Wong–Zakai type for stochastic partial differential equations,* Stochastic Processes and their Applications **55**, pp. 329–358, 1995.
12. Brzeźniak, Z. and Léandre, R., Stochastic Pants on Riemannian Manifolds, Nancy preprint (1999).
13. Brzeźniak, Z. and Peszat, S., *Maximal Inequalities and Exponential estimates for Stochastic Convolutions in Banach Spaces,* pp. 55–64 in Stochastic Processes, Physics and Geometry: New Interplays. I A Volume in Honour of Sergio Albeverio, CMS Conference Proceedings, v. **28**, (2000).
14. Carroll, A., *The Stochastic Nonlinear Heat Equation,* PhD Thesis, The University of Hull, 1999.
15. Cartan, H., Differential Calculus, Herman, Paris, 1971.
16. Clark, J.M.C., PhD thesis.
17. Clark, J.M.C., *An introduction to stochastic differentiable equations on manifolds*, in Geometric Methods in Systems Theory, ed. D.Q. Meyne and R.W. Brochett, Reidel 1973.
18. Cohen, S. and Estrade, A., *Non-symmetric approximations for manifold-valued semimartingales.* Ann. Inst. Henri Poincaré, Probab. Stat. **36**, No.1, pp. 45–70 (2000).
19. Davie, A.M., *Differential equations driven by rough signals: an approach via discrete approximations*, a manuscript (in preparation).
20. Dettweiler, E., *Stochastic Integration Relative To Brownian Motion on a general Banach Space,* Doga-Tr. J. of Mathematics **15**, pp. 6–44, 1991.
21. R. Deville, G. Godefroy and V. Zizler, Smoothness and Renorming in Banach Spaces, Pitman Monographs and Surveys in Pure and Applied Mathematics 64, Longman Scientific and Technical, 1993.

22. H. Doss, *Liens entre équations différentielles stochastiques et ordinaires*, Ann. Inst. H. Poincaré **13**(2), pp. 99–125 (1977).
23. Dowell, R.M., *Differentiable Approximations to Brownian Motion on Manifolds,* PhD Thesis, University of Warwick, 1980.
24. Elworthy, K.D., *Stochastic dynamical systems and their flows*, in STOCHASTIC ANALYSIS, Proc. Int. Conf., Evanston/Ill. 1978, ed. A. Friedman and M. Pinsky, pp. 79–95, Academic Press, London New York, 1978.
25. Elworthy, K.D., *Stochastic Differential Equations On Manifolds*, London Mathematical Society LNS vol. 70, Cambridge University Press, 1982.
26. Elworthy, K.D., *Geometric Aspects of Diffusions on Manifolds*, École d'Été de Probabilités de Saint-Flour XVII, 1985, 1987, LNM 1362, Springer Verlag 1988.
27. Hoffman-Jorgensen J. and Pisier, G., *The Law of Large Numbers and the Central Limit Theorem in Banach Spaces*, Annals of Probability **4**, pp. 587–599 (1976).
28. Ikeda, N. and Watanabe, S., STOCHASTIC DIFFERENTIAL EQUATIONS AND DIFFUSION PROCESSES, North-Holland, 1981.
29. Kirszbraun, M.D. *Über die zusammenziehenden und Lipschitzschen Transformationen*, Fundam. Math. **22**, pp. 77–108 (1934).
30. Kunita, H., *Stochastic Differential Equations and Stochastic Flow of Diffeomorphisms*, pp. 144–303, École d'Été de Probabilités de Saint-Flour XII – 1982, LNM 1097, Springer-Verlag (1984).
31. Kuo, H.H., GAUSSIAN MEASURES ON BANACH SPACES, LNM 463, Springer Verlag, Berlin, 1975.
32. Ledoux, M., Lyons, T. and Qian, Z., *Lévy Area of Wiener processes in Banach spaces*, . Ann. Probab. **30**, no. 2, pp. 546–578 (2002).
33. Lyons, T.J., *Differential equations driven by rough signals*, Rev. Mat. Iberoamericana **14**, n. 2, pp. 215–310 (1998).
34. Mackevicius, V., S^p *stability of solutions of symmetric stochastic differential equations with discontinuous driving semimartingales,* Annales de l'Institut Henri Poincaré, Probabilités et Statistiques, Vol **23**, no 4, pp. 575–592, (1987).
35. Malliavin, P. *Stochastic Calculus of Variations and Hypoelliptic Operators,* in Proceedings of the International Conference on Stochastic Differential Equations, Kyoto, pp. 195–263, Wiley, 1978.
36. Malliavin, P. GÉOMÉTRIE DIFFÉRENTIELLE STOCHASTIQUE, Notes rédigées par D. Dehen et D. Michel, Les Presses de l'Université de Montréal, Montréal, 1978.
37. Malliavin, P., *A bootstrap proof of the limit theorem for linear SDE*, Cranston, Michael C. (ed.) et al., STOCHASTIC ANALYSIS, Proceedings of the Summer Research Institute on stochastic analysis, held at Cornell University, Ithaca, NY, USA, July 11–30, 1993. Providence, RI: American Mathematical Society. Proc. Symp. Pure Math. 57, pp. 361–372 (1995).
38. Malliavin, P., STOCHASTIC ANALYSIS, Springer, Berlin, 1997.
39. Malliavin, P. and Nualart, D., *Quasi-sure analysis of stochastic flow and Banach space valued smooth functionals on the Wiener space*, J. Funct. Anal. **112**, pp. 287–317 (1993).
40. Moulinier, J.M., *Théorème limite pour équations différentielles stochastiques,* Bull. Sci. Math. (2) **112**, pp. 185–209. (1988)
41. Nakao, S. and Yamato, Y., *Approximation theorem on stochastic differential equations*, Proc. Intern. Symp. SDE Kyoto 1976 (ed. by K. Itô), pp. 283–296, Kinokuniya, Tokyo (1978).

42. McShane, E.J., *Stochastic differential equations and models of random processes*, in: Proc. 6th Berkeley Sympos. Math. Statist. Probab. **III**, Univ. Calif. 1970, pp. 263–294, University of California Press, (1972).
43. McShane, E.J., STOCHASTIC CALCULUS AND STOCHASTIC MODELS, Academic Press, New York–London, 1974.
44. Neidhardt, A.L., *Stochastic Integrals in* 2*-uniformly smooth Banach spaces,* University of Wisconsin, 1978.
45. Pisier, G., *Martingales with values in uniformly convex spaces*, Israel Journal of Mathematics **20**, pp. 326–350 (1976).
46. Sato, H., *Gaussian Measures on a Banach space and abstract Wiener measure*, Nagoya Math. J. 36, pp. 65–81, 1969.
47. Schönbeck, S.O., *On the extension of Lipschitz maps*, Ark. Mat. **7**, pp. 201–209 (1967).
48. Stroock, D.W. and Varadhan, S.R.S., *On the support of diffusion processes with applications to the strong maximum principle,* in: Proc. 6th Berkeley Sympos. Math. Statist. Probab., Univ. Calif. 1970, 3, pp. 333–359, University of California Press, (1972).
49. Sussman, H.J., *On the gap between deterministic and stochastic ordinary differential equations*, Ann. Prob. **6**, pp. 19–41 (1978).
50. Triebel, H., INTERPOLATION THEORY, FUNCTION SPACES, DIFFERENTIAL OPERATORS, North-Holland, Amsterdam New York Oxford, 1978.
51. Wong, E., and Zakai, M., *On the convergence of ordinary integrals to stochastic integrals,* Ann. Math. Statist., **36**, pp. 1560–1564 (1965).
52. Wong, E., and Zakai, M., *Riemann–Stieltjes Approximations of Stochastic Integrals*, Z. Wahrscheinlishkeitstheorie verw. Geb. **12**, pp. 87–97 (1969).