Weierstraß-Institut
für Angewandte Analysis und Stochastik
Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 2198-5855

Solving linear parabolic rough partial differential equations

Christian Bayer[1], Denis Belomestny[2], Martin Redmann[1],

Sebastian Riedel[3], John Schoenmakers[1]

submitted: May 8, 2018

[1] Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
E-Mail: christian.bayer@wias-berlin.de
martin.redmann@wias-berlin.de
john.schoenmakers@wias.berlin.de

[2] Universität Duisburg-Essen
Fachbereich Mathematik
Thea-Leymann-Str. 9
45127 Essen
Germany
E-Mail: denis.belomestny@uni-due.de

[3] TU Berlin
Institut für Mathematik
Straße des 17. Juni 136
10623 Berlin
Germany
E-Mail: riedel@math.tu-berlin.de

No. 2506

Berlin 2018

2010 *Mathematics Subject Classification.* Primary 65C30; Secondary 65C05, 60H15.

Key words and phrases. Rough paths, rough partial differential equations, Feynman-Kac formula, regression.

C.B., M.R., S.R., and J.S. gratefully acknowledge support from the DFG through the research unit FOR2402.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 20372-303
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Solving linear parabolic rough partial differential equations

Christian Bayer, Denis Belomestny, Martin Redmann,
Sebastian Riedel, John Schoenmakers

Abstract

We study linear rough partial differential equations in the setting of [Friz and Hairer, Springer, 2014, Chapter 12]. More precisely, we consider a linear parabolic partial differential equation driven by a deterministic rough path $\mathbf{W}$ of Hölder regularity α with $1/3 < \alpha \leq 1/2$. Based on a stochastic representation of the solution of the rough partial differential equation, we propose a regression Monte Carlo algorithm for spatio-temporal approximation of the solution. We provide a full convergence analysis of the proposed approximation method which essentially relies on the new bounds for the higher order derivatives of the solution in space. Finally, a comprehensive simulation study showing the applicability of the proposed algorithm is presented.

1 Introduction

We consider linear rough partial differential equations in the setting of Friz and Hairer [14, Chapter 12], see also Diehl, Oberhauser, Riedel [13] and Diehl, Friz and Stannat [12], i.e.,

$$\begin{aligned} -du &= L(u)\,dt + \sum_{k=1}^{d} \Gamma_k(u) d\mathbf{W}^k, \\ u(T,\cdot) &= g, \end{aligned}$$

where the differential operators L and $\Gamma = (\Gamma_1, \dots, \Gamma_d)$ are defined by

$$\begin{aligned} Lf(x) &= \frac{1}{2} \operatorname{trace}\left(\sigma(x)\sigma(x)^\top D^2 f(x)\right) + \langle b(x)\,, Df(x)\rangle + c(x) f(x), \\ \Gamma_k f(x) &= \langle \beta_k(x)\,, Df(x)\rangle + \gamma_k(x) f(x), \end{aligned}$$

see Section 2 for more details. We stress here that $\mathbf{W}$ is a *deterministic* rough path (of Hölder regularity α with $1/3 < \alpha \leq 1/2$), i.e., the PDE above is considered as a deterministic, not a stochastic equation. (This does not, of course, preclude choosing individual trajectories produced by a stochastic process, say a fractional Brownian motion.)

The goal of this paper is to provide a numerical algorithm for solving the above rough partial differential equation together with a proper numerical analysis of the algorithm and numerical examples. More precisely, we want to approximate the

DOI 10.20347/WIAS.PREPRINT.2506 Berlin 2018

function $(t, x) \mapsto u(t, x)$ as a linear combination of some easily computable basis functions depending on x with time dependent coefficients. Such approximations can then be, for example, used to solve optimal control problems for linear rough PDEs. In this respect, our approach can be viewed as an alternative to the space-time Galerkin proper orthogonal decomposition method used to solve optimal control problems for the standard linear parabolic PDEs (see, e.g. [4] and references therein). We analyze the corresponding approximation error which turns out to depend on smoothness properties of the solution u. As a by-product of this analysis, we also proved regularity of the solution u in x of degree larger than 1 under suitable conditions.

Stochastic partial differential equations driven by Brownian motion have, of course, many potential applications for instance in filtering theory ([27]) and mathematical finance ([6]), see also [21] for more examples. In many of these cases, it can be beneficial to go beyond the classical case and, for example, allow fractional Brownian motion, which allows for auto-correlation of the noise in time. Our framework naturally covers such extensions, concentrating on the (technically more challenging) case of Hurst index $H < \frac{1}{2}$.

1.1 Literature review

Terry Lyons' [24] theory of rough paths provides a deterministic, pathwise analysis of stochastic ordinary differential equations. This has interesting consequences both from a theoretical point of view—often based on the continuity of the solution w.r.t. the driving noise (which is not true in the classical stochastic analysis framework)—and from a practical point of view—see, for instance, [23]. We refer to [17, 14] for background information on rough path theory.

Nonetheless, the seemingly obvious step from rough ODEs the rough PDEs turns out to be quite difficult, mainly because of two essential limitations of standard rough path theory: regularity of the vector fields driving the differential equation (lacking in the case of (unbounded) partial differential operators), and the restriction to paths, i.e., functions parametrized by a one-dimensional variable. While not relevant for this paper, we should mention that the second restriction was overcome by seminal work of Hairer [20], thereby allowing space-time noise.

Despite those difficulties, rough partial differential equations (driven by a true path, i.e., a "noise" component only depending on time, but not space) have become a thriving field in the last few years, and several approaches have been developed to extend rough path analysis to rough PDEs (RPDEs). Most approaches are based on transformations of the problem separating the roughness of the drivers from the non-regularity of the differential operators. A series of papers by Friz and co-authors derives existence and uniqueness results for some classes of RPDEs by applying a flow-transformation to a classical PDE (with random coefficients), for instance see [7, 16]. Other works in this flavour are based on mild formulations of the RPDE, e.g., Deya, Gubinelli and Tindel [10].

Some more recent works have focused on more intrinsic formulations of rough

PDEs, trying to extend classical PDE techniques to the rough PDE context. This paper is based on the Feynman-Kac approach of Diehl and co-authors [12, 13, 14], which is presented in more detail in Section 2. In a quite different vein, Deya, Gubinelli, Hofmanová and Tindel [9] have provided a rough Gronwall lemma, which makes classical approaches to weak solutions of PDEs accessible.

Despite the increasing interest in rough PDEs, so far no numerical schemes have been suggested to the best of our knowledge. In this context, let us again mention [9], which could open up the field to finite element methods, as it provides variational techniques for some classes of rough PDEs. Of course, an abundance of numerical methods exist for classical stochastic PDEs and PDEs with random coefficients, see, for instance, [22].

In this work, we will use the stochastic representation of [12] in order to build a regression based approximation of the solution $u(t,\cdot)$ of the rough PDE, a technique that has been successfully applied both to stochastic PDEs by Milstein and Tretyakov [26] and to PDEs with random coefficients by Anker et al. [2].

1.2 Outline of the paper and main results

Diehl, Friz and Stannat [12] provide a solution theory to the rough partial differential equation above by means of a *stochastic representation*, i.e., they construct a stochastic process X which is driven by a stochastic rough path $\mathbf{Z}$ constructed from a Brownian motion B and our rough path $\mathbf{W}$ driving the rough PDE. The solution u of the rough PDE then is given as a conditional expectation of a functional of X, see Section 2 and, in particular, (2.6) for more details.

For the numerical approximation of $u(t,\cdot)$, it is very important to understand the regularity of this map. Note that the regularity in t quite clearly corresponds to the regularity of the driving path $\mathbf{W}$, whereas regularity in space alone can be much better depending on the coefficients and the terminal data g. In the theoretical work [12], spacial regularity of u is obtained from regularity of the stochastic process X in its initial value $X_0 = x$, which is well understood for rough differential equations. In order to show regularity of u one, however, needs to interchange differentiation with expectation, and the required integrability conditions were only available for the first derivative (see Cass, Litterer, Lyons [8]), but not for higher derivatives. In Section 3 we extend these results to higher derivatives, which enables us to show the following theorem, see Corollary 3.3:

Theorem 1.1. *Let $u(t,x)$ be as above. Assume that g is k-times differentiable with g and its derivatives having at most exponential growth. Assume further that b and β are bounded, $(2+k)$-times continuously differentiable with bounded derivatives, σ is bounded with $(3+k)$-times continuously differentiable with bounded derivatives, and γ and c are bounded, k-times continuously differentiable with bounded derivatives. Then $u(t,\cdot)$ is k times continuously differentiable and we provide explicit bounds on the derivatives.*

For any fixed $t > 0$, the spacial resolution of the function $x \mapsto u(t,x)$ can be approximated using regression with respect to properly chosen basis functions $\psi_1, \dots, \psi_K$,

$K \in \mathbb{N}$. More precisely, with respect to a specific probability measure μ on the state space $\mathbb{R}^n$, we try to minimize the error in the sense of $L^2(\mu)$, i.e., we would like to find

$$\operatorname*{arg\,min}_{\tilde{v}\in\operatorname{span}\{\psi_1,\dots,\psi_K\}} \int_{\mathbb{R}^n} |u(t,x) - \tilde{v}(x)|^2 \, \mu(dx).$$

The above loss function can, however, only serve as a guiding principle, since $u(t,x)$ is not available to us. Instead, we replace the above loss function by a proper Monte Carlo approximation: denoting the actual stochastic representation of u by $\mathcal{V} = \mathcal{V}(t,x,\omega)$ in the sense that $u(t,x) = E[\mathcal{V}(t,x)]$, we consider samples $\mathcal{V}^{(m)}$ of $\mathcal{V}$ obtained by

1 sampling initial values $x^{(m)}$ according to the distribution μ;

2 sampling the solution of X started at $X_t = x^{(m)}$ driven by independent (of each other and of $x^{(m)}$) samples of the Brownian motion.

Finally, we construct an approximation $\tilde{v}$ of $u(t,\cdot)$ by (essentially) solving the least squares problem

$$\operatorname*{arg\,min}_{\tilde{v}\in\operatorname{span}\{\psi_1,\dots,\psi_K\}} \sum_{m=1}^{M} \frac{1}{M} \left|\mathcal{V}^{(m)} - \tilde{v}(x^{(m)})\right|^2$$

via a "pseudo-regression" procedure (also used in [2] for PDEs with random coefficients) that is presented in detail in Section 4.) We obtain (cf. Theorem 4.1):

Theorem 1.2. *Under some boundedness conditions on the solutions and its stochastic representation, there is a constant $C > 0$ (which can be made explicit) such that*

$$E\left[\int_{\mathbb{R}^n} |u(t,x) - \tilde{v}(x)|^2 \, \mu(dx)\right] \le C\frac{K}{M} + \inf_{w\in\operatorname{span}\{\psi_1,\dots,\psi_K\}} \int_{\mathbb{R}^n} |u(t,x) - w(x)|^2 \, \mu(dx).$$

In order to find an approximation $u(t,x)$ on a time grid $0 < t_1 < \ldots < t_L < T$ the entire procedure (2) has to be repeated for every time step, i.e., we generate samples of X starting in $x^{(m)}$ at the respective initial time t_l. In Section 4.2 we propose an alternative regression type algorithm which allows for approximating the solution $u(t,x)$ on $0 < t_1 < \ldots < t_L < T$ using only one set of trajectories of the process X with $X_0 = x^{(m)}$.

Under somewhat more restrictive assumptions, our approximation $\overline{u}$ satisfies

$$\begin{aligned} E\left[\int_{\mathbb{R}^n} |\overline{u}(t_l,x) - u(t_l,x)| \, \mathrm{P}_{X_{t_l}}(dx)\right] \;\le\; & C\left[\frac{K\,\log(M)}{M}\right. \\ & \left. + \inf_{w\in\operatorname{span}\{\psi_1,\dots,\psi_K\}} \int_{\mathbb{R}^n} |u(t_l,x) - w(x)|^2 \, \mathrm{P}_{X_{t_l}}(dx)\right], \end{aligned}$$

for $l = 1,\dots,L$. We note that the above estimate is much more involved compared to Theorem 1.2 owing to the fact that the exact distribution $P_{X_{t_l}}$ (which essentially plays the rôle of μ) cannot be assumed to be known and, therefore, has to be approximated itself.

In order to bound the corresponding approximation errors

$$\inf_{f\in\mathrm{span}\{\psi_1,\dots,\psi_K\}} \|u(t_l,\cdot)-f(\cdot)\|^2_{L^2(\varrho)}, \quad l=1,\dots,L,$$

where the measure ϱ is either μ or $\mathrm{P}_{X_{t_l}}$, one needs to specify the basis functions.

In Section 4.3 we show that in the case of piecewise polynomial basis functions, the approximation error can be bounded (up to a constant) by $K^{-\varkappa}$ with $\varkappa=\frac{2(q+1)}{n}$, provided that each function $u(t_l,\cdot)$ is $q+1$ times differentiable, irrespective of the chosen regression type, and the support of ϱ above is finite. Now the smoothness of $u(t_l,\cdot)$ follows from the smoothness of the coefficients of the underlying PDE and g via Theorem 1.1.

A last puzzle piece is still missing if we want to provide a fully implementable approximation scheme, since we still need to solve the rough differential equation describing X. Here we employ a (simplified) Euler-type scheme including approximations of the needed signature terms by polynomials of the path itself, see Bayer, Friz, Riedel and Schoenmakers [5] for details. The scheme is recalled in Section 5. In the current context, the rate of convergence of the scheme is (almost) $2\alpha-1/2$, see Theorem 5.1 for details. Finally, we give several numerical examples in Section 6.

Remark 1.3. The scope of this paper is solving deterministic rough PDEs, i.e., PDEs driven by a deterministic but rough path $\mathbf{W}$. What happens when $\mathbf{W}$ is instead assumed to be random – implying randomness of u? If we apply the same algorithm as above, but with the samples of X based on i.i.d. samples of $\mathbf{W}$, then the regression based approximation is an estimate for $E[u(t,x)]$ (see [2] for the case of regular random noise). Of course, we can also use a regression approach for solving the full random solution $u(t,x;\omega)$. In this case, we need to choose basis functions in both x and ω. This means, proper basis functions need to be found in ω – or rather, in $\mathbf{W}$. The *signature* of $\mathbf{W}$ provides a useful parametrization for purposes of regression, see, for instance Lyons [23]. We will revisit this question in future works.

2 Stochastic representation

We consider rough partial differential equations in the setting studied [12, 13, 14]. Given a d-dimensional α-Hölder continuous geometric rough path $\mathbf{W}=(W,\mathbb{W})$, $\frac{1}{3}<\alpha\le\frac{1}{2}$, we consider the backward problem on $\mathbb{R}^n$

$$-du = L(u)\,dt + \sum_{k=1}^{d}\Gamma_k(u)d\mathbf{W}^k, \tag{2.1a}$$

$$u(T,\cdot)=g, \tag{2.1b}$$

where the differential operators L and $\Gamma=(\Gamma_1,\dots,\Gamma_d)$ are defined by

$$Lf(x)=\frac{1}{2}\,\mathrm{trace}\left(\sigma(x)\sigma(x)^\top D^2 f(x)\right)+\langle b(x)\,,Df(x)\rangle+c(x)f(x), \tag{2.2}$$

$$\Gamma_k f(x)=\langle\beta_k(x)\,,Df(x)\rangle+\gamma_k(x)f(x), \tag{2.3}$$

for a suitable test function $f : \mathbb{R}^n \to \mathbb{R}$ and given functions $\sigma : \mathbb{R}^n \to \mathbb{R}^{n\times m}$, $b : \mathbb{R}^n \to \mathbb{R}^n$, $c : \mathbb{R}^n \to \mathbb{R}$, $\beta_k : \mathbb{R}^n \to \mathbb{R}^n$, $\gamma_k : \mathbb{R}^n \to \mathbb{R}$, $k = 1, \dots, d$. All functions are "smooth enough".

A function $u = u(t, x; \mathbf{W})$ is called a "regular" solution[1] to (2.1) if $u \in C^{0,2}$ and

$$u(t,x) = g(x) + \int_t^T Lu(r,x)dr + \sum_{k=1}^d \int_t^T \Gamma_k u(r,x) d\mathbf{W}_r^k,$$

where the integral is understood in the rough path sense requiring $\Gamma_k u, \Gamma_j \Gamma_k u$ to be controlled by $\mathbf{W}$ as functions in t.

Solutions to the above rough PDE in the above sense are constructed by Feynman-Kac representations. We introduce an m-dimensional Brownian motion B, which will essentially be used to construct a diffusion process with generator L. Specifically, let

$$dX_t = \sigma(X_t)dB_t + b(X_t)dt + \beta(X_t)d\mathbf{W}_t, \tag{2.4}$$

where the dB-integral is understood in the Itô sense. More precisely, (2.4) is understood as a random (via B) rough ordinary differential equation with respect to a $(m + d)$-dimensional rough path $\mathbf{Z} = (Z, \mathbb{Z})$ defined by

$$Z_t := \begin{pmatrix} B_t \\ W_t \end{pmatrix}, \quad \mathbb{Z}_{s,t} := \begin{pmatrix} \mathbb{B}^{\text{Itô}}_{s,t} & \int_s^t W_{s,r} \otimes dB_r \\ \int_s^t B_{s,r} \otimes dW_r & \mathbb{W}_{s,t} \end{pmatrix}. \tag{2.5}$$

The following existence and uniqueness theorem is (part of) [12, Theorem 2.8].

Theorem 2.1. *Assume that the coefficients satisfy $\sigma_i, \beta_j, \gamma_k \in C_b^6(\mathbb{R}^n)$, $c, g \in C_b^4(\mathbb{R}^n)$. Define*

(2.6)

$$u(t,x;\mathbf{W}) := E^{t,x}\left[g(X_T)\exp\left(\int_t^T c(X_r)dr + \int_t^T \gamma(X_r)d\mathbf{W}_r\right)\right], \quad (t,x) \in [0,T]\times\mathbb{R}^n.$$

Then $u \in C_b^{0,4}([0,T]\times\mathbb{R}^n)$ solves the problem (2.1) *in the regular sense. The solution is unique among all $C_b^{0,4}([0,T]\times\mathbb{R}^n)$ "which are controlled by $\mathbf{W}$". Moreover, if g additionally has exponential decay, then the same is true for u.*

Remark 2.2. The authors of this paper are in doubt to what extent Theorem 2.1 was indeed proved in [12], in particular with respect to regularity. Differentiability of u in space is obtained by the corresponding differentiability of the solution map $x = X_0 \mapsto X_t$ of the mixed stochastic/rough differential equation (2.4). Cass, Litterer and Lyons [8] (see also [15]) have proved the existence *and integrability* of the first variation of RDEs like (2.4), i.e., the first derivative, which extends to the statement that $u \in C^{0,1}$ in the above theorem. However, to the best of our knowledge, this result has not been extended to higher order derivatives in the literature before. We fill this gap in Section 3, see Theorem 3.1 for the result on regularity of the flow of an RDE and Corollary 3.3 for the extended version of Theorem 2.1 above.

[1] [12] also provide a weak notion of solution. In what follows, the construction for both notions of solutions is the same, but weak solutions can be established under weaker regularity conditions on the coefficients.

Remark 2.3. It is possible to consider the problem (2.1) for slightly more general operators L and Γ, by adding to L and Γ an autonomous term, say $h(x)$ and $\eta(x) \in \mathbb{R}^d$, respectively. This will result in an extended stochastic representation

$$E\left[g(X_T^{t,x})Y_T^{t,x,1} + Z_T^{t,x,1,0}\right], \quad t \le T, \quad x \in \mathbb{R}^n, \tag{2.7}$$

for the solution of (2.1), where $X_t^{t,x} = x$, and

$$\begin{aligned} Y_s^{t,x,1} &:= \exp\left(\int_t^s c(X_r^{t,x})dr + \int_t^s \gamma(X_r^{t,x})dW_r\right), \\ Z_T^{t,x,1,0} &:= \int_t^T Y_r^{t,x,1}\left(h(X_r)dt + \eta^\top(X_r)\,dW_r\right). \end{aligned} \tag{2.8}$$

Remark 2.4. By defining a mean-zero process $\widetilde{Z}_\cdot^{t,x,1,0}$ as the solution to

$$d\widetilde{Z}_s = Y_s F^\top(s, X_s)dB_s, \quad \widetilde{Z}_t = 0$$

for an arbitrary column vector function $F(s,y) \in \mathbb{R}^m$, $y \in \mathbb{R}^n$, and Y_s given in (2.8), we obtain another modification of the standard stochastic representation, (2.6), which provides a stochastic representation with a free parameter that has smaller (pointwise) variance if this parameter is chosen accordingly. Indeed, from Theorem 2.1 it is a trivial observation that

$$E\left[g(X_T^{t,x})Y_T^{t,x,1} + \widetilde{Z}_T^{t,x,1,0}\right], \quad t \le T, \quad x \in \mathbb{R}^n, \tag{2.9}$$

is a stochastic representation to the solution of (2.1). In fact, via the chain rule for geometric rough paths it is possible to show that the variance of the random variable

$$g(X_T^{t,x})Y_T^{t,x,1} + \widetilde{Z}_T^{t,x,1,0}$$

vanishes if F satisfies $\sigma^\top Du + F = 0$. (Cf. Milstein and Tretyakov [25] for this result in the standard SDE setting.) Of course such an "optimal" F involves the solution of the problem itself, and as such is not directly available. A comprehensive study of constructing "good" variance reducing parameters F in the present context is deferred to subsequent work.

Remark 2.5. From a regression point of view, it might be simpler to consider the Dirichlet problem on a domain $D \subset \mathbb{R}^n$, i.e.,

$$\begin{aligned} -du &= L(u)\,dt + \sum_{k=1}^{d} \Gamma_k(u)d\mathbf{W}^k, \\ u(T,x) = g(x),\ x \in D, &\quad u(\cdot,x) = f(x),\ x \in \partial D. \end{aligned}$$

There are a few challenges here:

- A new existence and uniqueness theorem following the lines of [12] is required. In particular, the Feynman-Kac representation in terms of stopped processes has to be derived.
- Numerical schemes for stopped rough differential equations have, to the best of our knowledge, not yet been considered.

3 Regularity of the solution

In order to understand the convergence of the regression based approximation to u as a function of the input data (including the rough path $\mathbf{W}$), we need to control the derivative $\partial_x u(t,x)$ and higher order derivatives explicitly in terms of the data.

We start with an e-dimensional weakly geometric rough path $\mathbf{Z}$ with finite p-variation norm ($2 \le p < 3$).[2] Recall that standard stability estimates for solutions of rough differential equations driven by $\mathbf{Z}$ lead to estimates of the form

$$\exp\left(\|\mathbf{Z}\|_{p-\mathrm{var}} \vee \|\mathbf{Z}\|_{p-\mathrm{var}}^p\right),$$

see, for instance, [17, Theorem 10.38]. If we replace $\mathbf{Z}$ by a Brownian rough path, we see that terms of the above form are not integrable, due to the pth power. Hence, these estimates, which are sufficient (and sharp) in the deterministic setting, are impractical in the stochastic setting. Cass, Litterer and Lyons [8] were able to derive alternative estimates for the first derivative of the solution flow induced by a rough differential equation, which retain integrability in (most) Gaussian contexts, cf. also [15]. In the following section, we extend their results to higher order derivatives.

3.1 Higher order derivatives of RDE flows

Consider the rough differential equation

$$X_t^x = x + \int_0^t V(X_s^x)\, d\mathbf{Z}_s \in \mathbb{R}^n, \tag{3.1}$$

where $V \in C(\mathbb{R}^n, L(\mathbb{R}^e, \mathbb{R}^n))$. Formally, the derivative $X^{(1)} := D_x X^x$ should solve the equation

$$X_t^{(1)} = \mathrm{Id} + \int_0^t DV(X_s^x)(d\mathbf{Z}_s) X_s^{(1)} \in \mathbb{R}^{n\times n}$$

with

$$DV\colon \mathbb{R}^n \to L(\mathbb{R}^n, L(\mathbb{R}^e, \mathbb{R}^n)) \cong L(\mathbb{R}^e, L(\mathbb{R}^n, \mathbb{R}^n)).$$

The higher order derivatives of the vector field V are functions

$$D^k V\colon \mathbb{R}^n \to L((\mathbb{R}^n)^{\otimes k}, L(\mathbb{R}^e, \mathbb{R}^n)) \cong L(\mathbb{R}^e, L((\mathbb{R}^n)^{\otimes k}, \mathbb{R}^n)),$$

and the k-th derivative of the flow $X^{(k)} := D_x^k X^x$ should be a function

$$D^k X_t^{\cdot}\colon \mathbb{R}^n \to L((\mathbb{R}^n)^{\otimes k}, \mathbb{R}^n).$$

[2] Note that $\mathbf{Z}$ as defined in (2.5) is *not* weakly geometric. As outlined below, we have to transform the equation (2.4) for X into Stratonovich form first.

Taking formally the second derivative in (3.1), we obtain the equation

$$X_t^{(2)} = \int_0^t D^2V(X_s^x)(d\mathbf{Z}_s)(X_s^{(1)} \otimes X_s^{(1)}) + \int_0^t DV(X_s^x)(d\mathbf{Z}_s)X_s^{(2)},$$

and for the third derivative,

$$\begin{aligned} X_t^{(3)} = &\int_0^t D^3V(X_s^x)(d\mathbf{Z}_s)(X_s^{(1)} \otimes X_s^{(1)} \otimes X_s^{(1)}) + 2\int_0^t D^2V(X_s^x)(X_s^{(1)} \otimes X_s^{(2)}) \\ &+ \int_0^t D^2V(X_s^x)(d\mathbf{Z}_s)(X_s^{(2)} \otimes X_s^{(1)}) + \int_0^t DV(X_s^x)(d\mathbf{Z}_s)X_s^{(3)}. \end{aligned}$$

These formal calculations can be performed for any order k. The forthcoming theorem justifies these calculations. Moreover, it provides estimates for the solution which are especially useful for tail estimates when the equation is driven by a Gaussian process. For given $0 \le s < t \le T$, these estimates are based on the following sequence of times τ_i, iteratively defined by $\tau_0 = s$ and

$$\tau_{i+1} := \inf\left\{ \tau_i < u < t \;\middle|\; \|\mathbf{Z}\|^p_{p-\mathrm{var};[\tau_i,u]} \ge \alpha \right\} \wedge t,$$

where α is a positive parameter. Define

$$N_\alpha(\mathbf{Z};[s,t]) := \max\{\, n \mid \tau_n < t \,\}. \tag{3.2}$$

For $\alpha = 1$, we will omit the parameter and simply write N instead of N_1. The important insight of [8] was that $\|\mathbf{Z}\|^p_{p-\mathrm{var}}$ can often be replaced by N in rough path estimates, and that N does have Gaussian tails when $\mathbf{Z}$ is replaced by Gaussian processes respecting certain regularity assumptions. But for now we remain in a purely deterministic setting.

Next, we state the main theorem of this section. Since the proof is a bit lengthy, we decided to give it in the appendix, cf. page 41.

Theorem 3.1. *Fix $s \in [0,T]$ and let $\mathbf{Z}$ be a weakly geometric p-rough path for $p \in [2,3)$. Let $V \in C_b^{2+k}(\mathbb{R}^n, L(\mathbb{R}^e,\mathbb{R}^n))$ for some $k \ge 1$. Consider the unique solution $X^{s,x}$ to*

$$X_t^{s,x} = x + \int_s^t V(X_u^{s,x})\, d\mathbf{Z}_u \in \mathbb{R}^n, \quad t \in [s,T]. \tag{3.3}$$

Then for every fixed $t \in [s,T]$, the map $x \mapsto X_t^{s,x}$ is k-times differentiable. Moreover, the kth derivative $X_t^{(k)} = D_x^k X_t^{s,x}$ solves a rough differential equation which is obtained by formally differentiating (3.3) *k-times with respect to x. Setting $\omega(u,v) := \|\mathbf{Z}\|^p_{p-var;[u,v]}$, we have the bounds*

(3.4)

$$\|X^{(k)}\|_{p-\omega} := \|X^{(k)}\|_{p-\omega;[s,T]} \le C\|V\|_{C_b^{2+k}} \exp\left(C\|V\|^p_{C_b^{2+k}}(N(\mathbf{Z};[s,T])+1)\right) \quad \textit{and}$$

(3.5)

$$\|X^{(k)}\|_\infty := \|X^{(k)}\|_{\infty;[s,T]} \le |X_s^{(k)}| + C\|\mathbf{Z}\|_{p-var;[s,T]}\|V\|_{C_b^{2+k}} \exp\left(C\|V\|^p_{C_b^{2+k}}(N(\mathbf{Z};[s,T])+1)\right)$$

where C depends on p and k.

3.2 Bounding higher variations of (2.4)

In order to apply Theorem 3.1 to the rough stochastic differential equation (2.4), we first rewrite it in Stratonovich form which leads to the following hybrid Stratonovich-rough differential equation

$$\begin{aligned} dX_t &= [b(X_t) - a(X_t)]\,dt + \sigma(X_t) \circ dB_r + \beta(X_t) d\mathbf{W}_t \\ &= \widehat{b}(X_t)dt + \sigma(X_t) \circ dB_r + \beta(X_t) d\mathbf{W}_t \end{aligned} \tag{3.6}$$

where $a(\cdot) = \frac{1}{2}\sum_{i=1}^{m} D\sigma_i(\cdot)\cdot\sigma_i(\cdot)$ is the Itô-Stratonovich correction, σ_i is the ith column of σ and $\widehat{b}(\cdot) := b(\cdot) - a(\cdot)$. Equation (3.6) is now indeed of the form (3.1) if we set $V = (\widehat{b}, \sigma, \beta)\colon \mathbb{R}^n \to L(\mathbb{R}^{1+m+d}, \mathbb{R}^n)$ and consider the joint (geometric) rough path lift $\mathbf{Z}$ of $t \mapsto (t, B_t, W_t) =: (\tilde{B}_t, W_t)$ obtained from the Stratonovich rough path lift of the Brownian motion $\tilde{B}$ and $\mathbf{W}$ (recall that d is the dimension of $\mathbf{W}$, m is the dimension of the Brownian motion B). The lift $\mathbf{Z}$ is given as in (2.5), but $\mathbb{B}^{\text{Itô}}$ is substituted by the Stratonovich integral $\mathbb{B}^{\text{Str}}$, cf. [13] for further details.

Applying the bounds of Theorem 3.1 to the solution of equation (3.6) with initial condition $X_t = x$, we see that the expected value of the norm of the kth variation $D^k X$ is bounded in terms of the moment generating function of $N(\mathbf{Z}; [t,T])$. In [13], such bounds are provided considering the moment generating function of $\|\mathbf{Z}\|^2_{p-\text{var}}$. The following is a version of [13, Corollary 23], which differs in two respect: first, we consider the moment generating function of $N(\mathbf{Z}; [t,T])$ instead of $N(\mathbf{Z}; [t,T])^2$, and secondly we try to make the constants explicit (instead of only providing the existence of the exponential moment).

Lemma 3.2. *Given $\frac{1}{\alpha} < p < 3$ and $\delta > 0$, we let*

$$\kappa_p(\delta, \mathbf{W}) := E\left[\exp\left(\delta \|\mathbf{Z}\|^2_{p-\text{var};[t,T]}\right)\right],$$

assuming that δ is small enough such that $\kappa < \infty$. Then for all $\lambda > 0$ we have the bound

$$E\left[\exp\left(\lambda N(\mathbf{Z}; [t,T])\right)\right] \le \exp\left(2^{1/p}\lambda\left[\frac{\log(2\kappa_p(\delta, \mathbf{W})}{\delta}\right]^{p/2}\right) + \sqrt{2\pi}\lambda\sigma e^{2\lambda^2\sigma^2}, \tag{3.7}$$

where $\sigma := \sqrt{T-t}$.

Proof. Choose $K > \sqrt{\frac{\log \kappa_p(\delta, \mathbf{W})}{\delta}}$ and define

$$r_0 := 2^{1/p}K^p, \quad a := 1 - \frac{\kappa_p(\delta, \mathbf{W})}{\exp(\delta K^2)}, \quad \alpha := \Phi^{-1}(a),$$

where Φ denotes the c.d.f. of the standard normal distribution and we note that $0 < a < 1$ by our conditions. The result follow from the Fernique type estimate in [13, Theorem 17], which shows that

$$P\left(N(\mathbf{Z}; [t,T]) > r\right) \le 1 - \Phi\left(\alpha + \frac{r}{2\sigma}\right), \quad r \ge r_0.$$

(The choice of constants follows from [13, Lemma 19, Theorem 20, Lemma 22, proof of Corollary 23], for $q = 1$.) Using this estimate, the integration by parts formula

$$E\left[\exp(\lambda N(\mathbf{Z};[t,T])\right] = \int_0^\infty P\left(N(\mathbf{Z};[t,T]) > \frac{1}{\lambda}\log x\right)dx$$

and the estimate $1-\Phi(x) \le \frac{1}{2}e^{-x^2/2}$ (where the Gaussian tail estimate applies, i.e., for $r = \frac{1}{\lambda}\log x \ge r_0$) together with the trivial estimate of any probability by 1 (where the estimate does not apply, i.e., for $r = \frac{1}{\lambda}\log x < r_0$), directly gives

$$E\left[\exp\left(\lambda N(\mathbf{Z};[t,T])\right)\right] \le e^{\lambda r_0} + \sqrt{\frac{\pi}{2}}\sigma\lambda\exp\left(2\sigma\lambda(\sigma\lambda-\alpha)\right)\times\cdots \\ \cdots\times \operatorname{erfc}\left(\frac{1}{\sqrt{2}}\left(\alpha - 2\sigma\lambda + \frac{r_0}{2\sigma}\right)\right). \tag{3.8}$$

Note that $K^2 \ge \frac{1}{\delta}\log(2\kappa_p)$ implies $\alpha \ge 0$. Using the trivial bound $\operatorname{erfc} \le 2$, we further obtain

$$E\left[\exp\left(\lambda N(\mathbf{Z};[t,T])\right)\right] \le e^{2^{1/p}\lambda K^p} + \sqrt{2\pi}\sigma\lambda e^{2\lambda^2\sigma^2}.$$

The right hand side is now minimized by $K = \sqrt{\frac{1}{\delta}\log(2\kappa_p)}$, which gives (3.7). □

Finally, we consider bounds for the derivatives of the solution $u(t,x)$ of (2.1). For ease of notation, we will *formally* only consider the case $c \equiv 0$, $\gamma \equiv 0$, i.e.,

$$u(t,x;\mathbf{W}) = E^{t,x}\left[g(X_T)\right].$$

However, note that we do allow g and its derivatives to have exponential growth in what follows, and it is thus easy to incorporate the general setting by extending the state space. For this, we just need to add an additional component Y_t solving

$$dY_s = c(X_s)ds + \gamma(X_s)d\mathbf{W}_s, \quad Y_t = 0,$$

and consider

$$u(t,x;\mathbf{W}) = E^{t,x}\left[g(X_T)\exp(Y_T)\right].$$

Corollary 3.3. *Let $u(t,x,\mathbf{W})$ be as above. Assume that g is k-times differentiable and that there are constants ζ_1, $\zeta_2 \ge 0$ such that*

$$|D^l g(x)| \le \zeta_1 e^{\zeta_2|x|}$$

for all $x \in \mathbb{R}^n$ and $l = 1,\dots,k$. Assume that $\widehat{b}$, σ and β are bounded, $(2+k)$-times differentiable with bounded derivatives, and let $K > 0$ be a bound for their norms, i.e.

$$\|\widehat{b}\|_{C_b^{2+k}} \vee \|\sigma\|_{C_b^{2+k}} \vee \|\beta\|_{C_b^{2+k}} \le K.$$

Then there are constants $C = C(p,k)$ and $C_1 = C_1(p)$ such that

$$|\partial_x^k u(t,x,\mathbf{W})| \le C\zeta_1\lambda e^{\zeta_2(|x|+1)}\left(1+\sqrt{(T-t)}(\zeta_2 C_1 + C)K^p\right) \\ \times\exp\left((\zeta_2 C_1 + C)K^p\left(1+\left(\frac{\log\left(2\kappa_p(\delta,\mathbf{W})\right)}{\delta}\right)^{p/2} + (T-t)(\zeta_2 C_1 + C)K^p\right)\right)$$

where we use the same notation as in Lemma 3.2.

Proof. Recall that the solution $X_v^{t,x}$ to

$$dX_v^{t,x} = \widehat{b}(X_v^{t,x})\,dv + \sigma(X_v^{t,x}) \circ dB_v + \beta(X_v^{t,x})\,d\mathbf{W}_v; \quad X_t^{t,x} = x$$

equals the solution to

$$dX_v^{t,x} = V(X_v^{t,x})\,d\mathbf{Z}_v; \quad X_t^{t,x} = x$$

where $V = (\widehat{b}, \sigma, \beta)\colon \mathbb{R}^n \to L(\mathbb{R}^{1+m+d}, \mathbb{R}^n)$ and $\mathbf{Z}$ denotes the joint geometric rough path lift of $v \mapsto (v, B_v, W_v)$. Iterating the chain rule, we see that

$$\partial_x^k g(X_v^{t,x}) = \sum_{l=1}^{k} \sum_{i_1+\ldots+i_l=k} \lambda_{i_1,\ldots,i_k} (D^l g)(X_v^{t,x})(D_x^{i_1} X_v^{t,x} \otimes \ldots \otimes D_x^{i_l} X_v^{t,x})$$

where $1 \le i_1, \ldots i_l \le k$ and $\lambda_{l_1,\ldots,i_k}$ are nonnegative integers which can be calculated explicitly (using e.g. Faà di Bruno's formula). Thus we obtain

$$|\partial_x^k u(t,x)| = |E(\partial_x^k g(X_T^{t,x}))| \le \sum \lambda_{i_1,\ldots,i_l} \zeta_1 E\left(\exp(\zeta_2 |X_T^{t,x}|) |D_x^{i_1} X_T^{t,x}| \cdots |D_x^{i_l} X_T^{t,x}|\right).$$

As in the proof of Theorem 3.1, one can see that there is a constant C_1 depending only on p such that

$$|X_T^{t,x}| \le |x| + N(X^{t,x}; [t,T]) + 1 \le |x| + C_1 \|V\|^p_{C_b^{2+k}} (N(\mathbf{Z}; [t,T]) + 1) + 1.$$

The bounds of Theorem 3.1 imply that

$$\begin{aligned} |D_x^{i_1} X_T^{t,x}| \cdots |D_x^{i_l} X_T^{t,x}| &\le \left(1 + C\|\mathbf{Z}\|_{p-\mathrm{var};[t,T]} \|V\|_{C_b^{2+k}} \exp\left(C\|V\|^p_{C_b^{2+k}} (N(\mathbf{Z}; [t,T]) + 1)\right)\right)^l \\ &\le 2^{k-1} + C\|\mathbf{Z}\|^l_{p-\mathrm{var};[t,T]} \|V\|^l_{C_b^{2+k}} \exp\left(C\|V\|^p_{C_b^{2+k}} (N(\mathbf{Z}; [t,T]) + 1)\right) \end{aligned}$$

for a constant C depending on p and k. Therefore, we see that there is a constant λ depending on k only such that

$$\begin{aligned} |\partial_x^k u(t,x)| \le{}& \zeta_1 \lambda e^{\zeta_2(|x|+1)} E\left(\exp\left(\zeta_2 C_1 \|V\|^p_{C_b^{2+k}} (N(\mathbf{Z}; [t,T]) + 1)\right)\right) \\ &+ C\zeta_1 \lambda e^{\zeta_2(|x|+1)} \max_{l=1,\ldots,k} \|V\|^l_{C_b^{2+k}} E\left(\|\mathbf{Z}\|^l_{p-\mathrm{var};[t,T]} \exp\left((\zeta_2 C_1 + C)\|V\|^p_{C_b^{2+k}} (N(\mathbf{Z}; [t,T]) + 1)\right)\right). \end{aligned}$$

We use [15, Lemma 4, Lemma 1 and Lemma 3] to see that for every $l = 1, \ldots, k$,

$$(\|V\|_{C_b^{2+k}} \|\mathbf{Z}\|_{p-\mathrm{var};[t,T]})^l \le \exp(k\|V\|^p_{C_b^{2+k}} (2N(\mathbf{Z}; [t,T]) + 1)).$$

This implies that

$$|\partial_x^k u(t,x)| \le C\zeta_1 \lambda e^{\zeta_2(|x|+1)} E\left(\exp\left((\zeta_2 C_1 + C)\|V\|^p_{C_b^{2+k}} (N(\mathbf{Z}; [t,T]) + 1)\right)\right).$$

Now we use Lemma 3.2 to obtain the bound

$$
\begin{aligned}
&C\zeta_1\lambda e^{\zeta_2(|x|+1)}E\left(\exp\left((\zeta_2C_1+C)\|V\|^p_{C_b^{2+k}}(N(\mathbf{Z};[t,T])+1)\right)\right)\\
&\leq C\zeta_1\lambda e^{\zeta_2(|x|+1)+(\zeta_2C_1+C)\|V\|^p_{C_b^{2+k}}}E\left(\exp\left((\zeta_2C_1+C)\|V\|^p_{C_b^{2+k}}(N(\mathbf{Z};[t,T]))\right)\right)\\
&\leq C\zeta_1\lambda e^{\zeta_2(|x|+1)}\left(1+\sqrt{2\pi(T-t)}(\zeta_2C_1+C)\|V\|^p_{C_b^{2+k}}\right)\\
&\quad\times\exp\left((\zeta_2C_1+C)\|V\|^p_{C_b^{2+k}}\left(1+2^{1/p}\left(\frac{\log\left(2\kappa_p(\delta,\mathbf{W})\right)}{\delta}\right)^{p/2}+2(T-t)(\zeta_2C_1+C)\|V\|^p_{C_b^{2+k}}\right)\right)
\end{aligned}
$$

and our claim follows. □

4 Regression

From the numerical point of view it is desirable to have a functional approximation for the solution u, i.e., to have an approximation of the form

$$u(t,x)\approx\sum_{k=0}^{K}a_k(t)\psi_k(x), \tag{4.1}$$

for some natural $K>0$, where $(\psi_k(x))$ are some simple basis functions and the coefficients $(a_k(t))$ depend only on t. Such an approximation can be then used to perform integration, differentiation and optimization of $u(t,x)$ in a fast way. In this section we are going to use non-parametric regression to construct approximations of the type (4.1). First we turn to the problem of approximating $u(t,x)$ for a fixed $t>0$ and then consider approximation of the solution u in space and time. While the first problem can be solved using a simplified version of linear regression called pseudo-regression, for the second task we need to use general non-parametric regression algorithms.

4.1 Spacial resolution obtained by regression

The representation (2.6) implies,

$$u(t,X_t^{0,x})=E_{\mathcal{F}_t}\left[g(X_T^{t,X_t^{0,x}})Y_T^{t,X_t^{0,x},1}\right], \tag{4.2}$$

where $(\mathcal{F}_s)_{0\leq s\leq T}$ denotes the filtration generated by B. (Recall the notation introduced in Remark 2.3.) From (2.9) we observe that for $s\geq t$,

$$
\begin{aligned}
Y_s^{t,X_t^{0,x},1}&=\exp\left[\int_t^s c\left(X_r^{t,X_t^{0,x}}\right)dr+\gamma^\top\left(X_r^{t,X_t^{0,x}}\right)d\mathbf{W}_r\right]\\
&=\exp\left[\int_t^s c\left(X_r^{0,x}\right)dr+\gamma^\top\left(X_r^{0,x}\right)d\mathbf{W}_r\right]\\
&=\frac{\exp\left[\int_0^s c\left(X_r^{0,x}\right)dr+\gamma^\top\left(X_r^{0,x}\right)d\mathbf{W}_r\right]}{\exp\left[\int_0^t c\left(X_r^{0,x}\right)dr+\gamma^\top\left(X_r^{0,x}\right)d\mathbf{W}_r\right]}=\frac{Y_s^{0,x,1}}{Y_t^{0,x,1}}.
\end{aligned} \tag{4.3}
$$

Due to (4.3), (4.2) yields

$$u(t, X_t^{0,x}) = E_{\mathcal{F}_t}\left[g(X_T^{0,x})\frac{Y_T^{0,x,1}}{Y_t^{0,x,1}}\right]. \tag{4.4}$$

We now aim at estimating $u(t,x)$ for a fixed t, $0 \le t \le T$, globally in $x \in \mathbb{R}^n$, based on the stochastic representation (4.4). Let us consider a random variable $\mathcal{U}$ ranging over some domain $\mathcal{D} \subset \mathbb{R}^n$, with distribution μ. Given $\mathcal{U}$, we then consider the random trajectory

$$\left(X_s^{0,\mathcal{U}}, Y_s^{0,\mathcal{U},1}\right)_{0\le s\le T},$$

which is understood in the sense that the Brownian trajectory B is independent of $\mathcal{U}$. At time $s = 0$ we sample i.i.d. copies $\mathcal{U}_1, ..., \mathcal{U}_M$ of $\mathcal{U}$. We then construct a collection of "training paths" $\mathcal{D}_M^{tr}$, consisting of independent realizations

$$\mathcal{D}_M^{tr} := \left\{\left(X_s^{0,\mathcal{U}_m;m}, Y_s^{0,\mathcal{U}_m,1;m}\right)_{0\le s\le T} \mid m = 1, \dots, M\right\}, \tag{4.5}$$

again based on independent realizations of the Brownian motion B. Next consider for a fixed time t, $0 \le t \le T$, the vector $\mathcal{Y}^{(t)} \in \mathbb{R}^M$, where

$$\mathcal{Y}_m^{(t)} := g\left(X_T^{0,\mathcal{U}_m;m}\right)\frac{Y_T^{0,\mathcal{U}_m,1;m}}{Y_t^{0,\mathcal{U}_m,1;m}}. \tag{4.6}$$

Now let $\psi_1, ..., \psi_K$ be a set of basis functions on $\mathbb{R}^n$ and define define a matrix $\mathcal{M}^{(t)} \in \mathbb{R}^{M\times K}$ by

$$\mathcal{M}_{mk}^{(t)} := \psi_k\left(X_t^{0,\mathcal{U}_m;m}\right),$$

In the next step we solve the least squares problem

$$\widehat{\gamma}^{(t)} := \operatorname*{arg\,min}_{\gamma\in\mathbb{R}^K}\frac{1}{M}\sum_{m=1}^{M}\left(\mathcal{Y}_m^{(t)} - \sum_{k=1}^{K}\mathcal{M}_{mk}^{(t)}\gamma_k\right)^2 = \left(\left(\mathcal{M}^{(t)}\right)^\top\mathcal{M}^{(t)}\right)^{-1}\left(\mathcal{M}^{(t)}\right)^\top\mathcal{Y}^{(t)}. \tag{4.7}$$

This gives an approximation

$$\widehat{u}(t,x) = \widehat{u}(t,x;\mathcal{D}_M^{tr}) := \sum_{k=1}^{K}\widehat{\gamma}_k^{(t)}\psi_k(x) \tag{4.8}$$

of u. Thus, with one and the same sample (4.5) we may so get for different times t and states x an approximate solution $\widehat{u}(t,x)$. Let us first consider the particular case $t = 0$, where we have

$$\mathcal{M}_{mk}^{(0)} := \psi_k\left(\mathcal{U}_m\right), \quad \mathcal{Y}_m^{(0)} = g\left(X_T^{0,\mathcal{U}_m;m}\right)Y_T^{0,\mathcal{U}_m,1;m}$$

and then (4.7) reads

$$\widehat{\gamma}^{(0)} = \frac{1}{M}\left(\frac{1}{M}\left(\mathcal{M}^{(0)}\right)^\top\mathcal{M}^{(0)}\right)^{-1}\left(\mathcal{M}^{(0)}\right)^\top\mathcal{Y}^{(0)}. \tag{4.9}$$

Instead of the inverted random matrix in (4.9) we may turn over to a so called pseudo-regression estimator where the matrix entries

$$\left[\frac{1}{M}\left(\mathcal{M}^{(0)}\right)^{\top}\mathcal{M}^{(0)}\right]_{kl} = \frac{1}{M}\sum_{m=1}^{M}\psi_k\left(\mathcal{U}_m\right)\psi_l\left(\mathcal{U}_m\right)$$

are replaced by their limits as $M \to \infty$, i.e., by the scalar products

$$\mathcal{G}_{kl} := \langle \psi_k, \psi_l \rangle := \int_{\mathcal{D}} \psi_k(z)\,\psi_l(z)\,\mu(dz).$$

That is, we may also consider the estimate

$$\tilde{u}(0,x) := \tilde{u}(0,x;\mathcal{D}_M^{tr}) := \sum_{k=1}^{K}\tilde{\gamma}_k^{(0)}\psi_k(x) \quad \text{with} \tag{4.10}$$

$$\tilde{\gamma}^{(0)} := \frac{1}{M}\mathcal{G}^{-1}\left(\mathcal{M}^{(0)}\right)^{\top}\mathcal{Y}^{(0)}. \tag{4.11}$$

The interesting point is that in (4.10) we may freely choose both the initial measure, and the set of basis functions. So by a suitable choice of basis functions (ψ_k) and initial measure μ, we may arrange the matrix $\mathcal{G}$ to be known explicitly, or even that $\mathcal{G} = \mathrm{Id}$ (the identity matrix), thus simplifying the regression procedure significantly from a computational point of view. Indeed, the cost of computing (4.7) in (4.8) is of order MK^2 while the cost of computing (4.11) is only of order MK.

It should be emphasized that the function estimates (4.8) and (4.10) are random as they depend on the simulated training paths (4.5). In the next section we study mean-squares-estimation errors in a suitable sense for the particular case (4.10), and for the general case (4.8), respectively.

4.1.1 Error analysis

For the error analysis of the pseudo-regression method (4.10) we could basically refer to Anker et al. [2], where pseudo regression is applied in the context of global solutions for random PDEs. For the convenience of the reader, however, let us here recap the analysis in condensed form, consistent with the present context and terminology. For the formulation of the theorem and its proof below, let us abbreviate (cf. (4.6) and (4.10))

$$\begin{aligned}
\mathcal{V} &:= g(X_T^{0,\mathcal{U}})Y_T^{0,\mathcal{U},1}, \quad v(z) := u(0,z), \quad \tilde{v}(z) := \tilde{u}(0,z),\\
\mathcal{V}^{(m)} &:= g(X_T^{0,\mathcal{U}^{(m)};m})Y_T^{0,\mathcal{U}^{(m)},1;m}, \quad \mathcal{M} := \mathcal{M}^{(0)}, \quad \mathcal{Y} := \mathcal{Y}^{(0)}, \quad \tilde{\gamma} := \tilde{\gamma}^{(0)}.
\end{aligned}$$

Theorem 4.1. *Suppose that*

$$\begin{aligned}
&|v(z)| \le A \quad \textit{and} \quad \mathrm{Var}\left[\mathcal{V}|\mathcal{U}=z\right] < \sigma^2, \quad \textit{for all } z \in \mathcal{D},\\
&0 < \underline{\lambda_{\min}} \le \lambda_{\min}\left(\mathcal{G}^K\right) \le \lambda_{\max}\left(\mathcal{G}^K\right) \le \overline{\lambda_{\max}}, \quad \textit{for all } K = 1,2,...,
\end{aligned}$$

where $\lambda_{\min}\left(\mathcal{G}^K\right)$, *and* $\lambda_{\max}\left(\mathcal{G}^K\right)$, *denote the smallest, respectively largest, eigenvalue of the positive symmetric matrix* $\mathcal{G}^K := \left(\mathcal{G}_{ij}\right)_{1\le i,j\le K}$. *Then it holds,*

$$E\int_{\mathcal{D}}|\tilde{v}(z)-v(z)|^2\,\mu(dz) \le \frac{\overline{\lambda_{\max}}}{\underline{\lambda_{\min}}}\left(\sigma^2+A^2\right)\frac{K}{M}+ \\ + \inf_{w\in\operatorname{span}\{\psi_1,\dots,\psi_K\}}\int_{\mathcal{D}}|w(z)-v(z)|^2\,\mu(dz). \tag{4.12}$$

Proof. Let v^K be the projection of v on to the linear span of $\psi_1,\dots,\psi_K$, i.e.,

$$v^K = \operatorname*{arg\,inf}_{w\,\in\,\operatorname{span}\{\psi_1,\dots,\psi_K\}}\int_{\mathcal{D}}|w(z)-v(z)|^2\,\mu(dz). \tag{4.13}$$

Then, with $\gamma^\circ := (\gamma_1^\circ,\dots,\gamma_K^\circ)^\top \in \mathbb{R}^K$ defined by

$$v^K = \sum_{k=1}^K \gamma_k^\circ \psi_k, \tag{4.14}$$

and $\alpha \in \mathbb{R}^K$ defined by $\alpha_k := \langle \psi_k, v\rangle$, it follows straightforwardly by taking scalar products that

$$\gamma^\circ = \mathcal{G}^{-1}\alpha. \tag{4.15}$$

By the rule of Pythagoras it follows that,

$$E\int_{\mathcal{D}}|\tilde{v}(z)-v(z)|^2\,\mu(dz) = E\int_{\mathcal{D}}\left|\tilde{v}(z)-v^K(z)\right|^2\mu(dz)+\int_{\mathcal{D}}\left|v^K(z)-v(z)\right|^2\mu(dz). \tag{4.16}$$

With $\psi := (\psi_1,\dots,\psi_K)^\top$ it holds by (4.15) that,

$$\begin{aligned}
E\int_{\mathcal{D}}\left|\tilde{v}(z)-v^K(z)\right|^2\mu(dz) &= \int_{\mathcal{D}}E\left|\tilde{\gamma}^\top\psi(z)-\gamma^{\circ\top}\psi(z)\right|^2\mu(dz)\\
&= \int_{\mathcal{D}}E\left|\left(\frac{1}{M}\mathcal{Y}^\top\mathcal{M}-\alpha^\top\right)\mathcal{G}^{-1}\psi(z)\right|^2\mu(dz)\\
&= \int_{\mathcal{D}}E\left[\left(\frac{1}{M}\mathcal{Y}^\top\mathcal{M}-\alpha^\top\right)\mathcal{G}^{-1}\psi(z)\psi^\top(z)\mathcal{G}^{-1}\left(\frac{1}{M}\mathcal{M}^\top\mathcal{Y}-\alpha\right)\right]\mu(dz)\\
&= E\left[\left(\frac{1}{M}\mathcal{Y}^\top\mathcal{M}-\alpha^\top\right)\mathcal{G}^{-1}\left(\frac{1}{M}\mathcal{M}^\top\mathcal{Y}-\alpha\right)\right],
\end{aligned}$$

since

$$\int_{\mathcal{D}}\left[\psi(z)\psi^\top(z)\right]_{kl}\mu(dz) = \langle\psi_k,\psi_l\rangle = \mathcal{G}_{kl}.$$

We thus have that

$$0 \le E\int_{\mathcal{D}}\left|\tilde{v}(z)-v^K(z)\right|^2\mu(dz) \le \frac{1}{\underline{\lambda_{\min}}}E\left|\frac{1}{M}\mathcal{M}^\top\mathcal{Y}-\alpha\right|^2 = \frac{1}{\underline{\lambda_{\min}}}\sum_{k=1}^K \operatorname{Var}\left[\frac{1}{M}\mathcal{M}^\top\mathcal{Y}\right]_k,$$

using that

$$E\left[\frac{1}{M}\mathcal{M}^\top \mathcal{Y}\right]_k = \frac{1}{M}E\sum_{m=1}^{M}\psi_k(\mathcal{U}^{(m)})\mathcal{V}^{(m)}$$
$$= E\left(\psi_k(\mathcal{U}^{(1)})E\left[\mathcal{V}^{(1)}|\mathcal{U}^{(1)}\right]\right)$$
$$= \langle\psi_k, v\rangle = \alpha_k.$$

Now, by observing that

$$\mathrm{Var}\left[\frac{1}{M}\mathcal{M}^\top \mathcal{Y}\right]_k = \mathrm{Var}\left(\frac{1}{M}\sum_{m=1}^{M}\psi_k(\mathcal{U}^{(m)})\mathcal{V}^{(m)}\right)$$
$$= \frac{1}{M}\mathrm{Var}\left(\psi_k(\mathcal{U}^{(1)})\mathcal{V}^{(1)}\right)$$
$$= \frac{1}{M}E\,\mathrm{Var}\left[\psi_k(\mathcal{U}^{(1)})\mathcal{V}^{(1)}|\mathcal{U}^{(1)}\right] + \frac{1}{M}\mathrm{Var}\,E\left[\psi_k(\mathcal{U}^{(1)})\mathcal{V}^{(1)}|\mathcal{U}^{(1)}\right]$$
$$= \frac{1}{M}E\left(\psi_k^2(\mathcal{U}^{(1)})\mathrm{Var}\left[\mathcal{V}^{(1)}|\mathcal{U}^{(1)}\right]\right) + \frac{1}{M}\mathrm{Var}\,\psi_k(\mathcal{U}^{(1)})v\left(\mathcal{U}^{(1)}\right)$$
$$\leq \frac{\sigma^2 + A^2}{M}\mathcal{G}^K_{kk},$$

one has

$$\frac{1}{\underline{\lambda_{\min}}}\sum_{k=1}^{K}\mathrm{Var}\left[\frac{1}{M}\mathcal{M}^\top \mathcal{Y}\right]_k \leq \frac{\sigma^2 + A^2}{M\underline{\lambda_{\min}}}\mathrm{tr}\left(\mathcal{G}^K\right) \leq \frac{\sigma^2 + A^2}{M\underline{\lambda_{\min}}}K\overline{\lambda_{\max}},$$

and then (4.12) follows. □

4.2 Spatio-temporal resolution obtained by regression

If we want to approximate $u(t,x)$ in space and time, we can perform regression on a given set of trajectories for different time points t. Let us fix a time grid $(t_1,\ldots,t_L)$ with $0 < t_1 < t_2 < \ldots < t_L < T$ and consider regression problems

$$\widehat{\gamma}^{(l)} := \underset{\gamma\in\mathbb{R}^K}{\arg\min}\frac{1}{M}\sum_{m=1}^{M}\left(\mathcal{Y}_m^{(l)} - \sum_{k=1}^{K}\mathcal{M}_{mk}^{(l)}\gamma_k\right)^2$$
$$= \left(\left(\mathcal{M}^{(l)}\right)^\top \mathcal{M}^{(l)}\right)^{-1}\left(\mathcal{M}^{(l)}\right)^\top \mathcal{Y}^{(l)}, \quad l = 1,\ldots,L, \tag{4.17}$$

where

$$\mathcal{M}_{mk}^{(l)} := \psi_k\left(X_{t_l}^{0,\mathcal{U}_m;m}\right)$$

and

$$\mathcal{Y}_m^{(l)} := g\left(X_T^{0,\mathcal{U}_m;m}\right)\frac{Y_T^{0,\mathcal{U}_m,1;m}}{Y_{t_l}^{0,\mathcal{U}_m,1;m}}, \quad l = 1,\ldots,L.$$

This would give us a decomposition

$$\widehat{u}(t_l,x) = \widehat{u}(t_l,x;\mathcal{D}_M^{tr}) := \sum_{k=1}^{K}\widehat{\gamma}_k^{(l)}\psi_k(x), \quad l = 1,\ldots,L.$$

of u. Furthermore, the coefficients $\widehat{\gamma}_k^{(l)}$ can be interpolated to provide us with the approximation of the form (4.1). The convergence analysis of the estimates (4.17) is more involved and follows from the general theory of non-parametric regression, see Section 11 in [19] Assume that

(A1) $\max_{l=1,\dots,L} \sup_{z\in\mathbb{R}^n} \mathrm{Var}\left[g(X_T^{0,\mathcal{U}}) \frac{Y_T^{0,\mathcal{U},1}}{Y_{t_l}^{0,\mathcal{U},1}} \middle| X_{t_l}^{0,\mathcal{U}} = z \right] \le \sigma^2 < \infty,$

(A2) $\max_{l=1,\dots,L} \sup_{x\in\mathbb{R}^n} |u(t_l,x)| \le A < \infty,$

for some positive constants σ and A. Then we denote by $\overline{u}$ a truncated regression estimate, which is defined as follows:

$$\overline{u}(t,x) := T_A\widehat{u}(t,x) := \begin{cases} \widehat{u}(t,x) & \text{if } |\widehat{u}(t,x)| \le A, \\ A \operatorname{sgn} \widehat{u}(t,x) & \text{otherwise.} \end{cases}$$

Under (A1)–(A2) we have the following L^2-upper bound (see Theorem 11.3 in [19])

$$\begin{aligned} (4.18)\quad E\|\overline{u}(t_l,\cdot) - u(t_l,\cdot)\|^2_{L^2(\mathrm{P}_{X_{t_l}})} \;\le\; & \tilde{c}\left(\sigma^2 + A^2(\log M + 1)\right)\frac{K}{M} \\ & + 8 \inf_{f\in\mathrm{span}\{\psi_1,\dots,\psi_K\}} \|u(t_l,\cdot) - f(\cdot)\|^2_{L^2(\mathrm{P}_{X_{t_l}})}, \end{aligned}$$

for all $l = 1,\dots,L$, where $\tilde{c} > 0$ is a universal constant. Note that the use of the measure $\mathrm{P}_{X_{t_l}}$ in (4.18) is essential and $\mathrm{P}_{X_{t_l}}$ can not be in general replaced by an arbitrary measure μ as in the case of pseudo-regression algorithm.

Instead of linear regression, we could use a nonlinear one. Let us fix a nonlinear class of functions Ψ_M and define

$$\widehat{u}(t_l,x) = \operatorname*{arg\,min}_{\psi\in\Psi_M} \frac{1}{M}\sum_{m=1}^{M} \left(\mathcal{Y}_m^{(t_l)} - \psi\left(X_{t_l}^{0,\mathcal{U}_m;m}\right)\right)^2.$$

Under a stronger assumption that $|\mathcal{Y}^{(t_l)}| \le A$ with probability 1 for all $l = 1,\dots,L$ and a constant $A > 0$, we get (see Theorem 11.5 in [19])

$$\begin{aligned} (4.19)\qquad E\|\overline{u}(t_l,\cdot) - u(t_l,\cdot)\|^2_{L^2(\mathrm{P}_{X_{t_l}})} \;\le\; & (c_1 + c_2\log M)\frac{V_{\Psi_M}}{M} \\ & + 2 \inf_{f\in\Psi_M} \|u(t_l,\cdot) - f(\cdot)\|^2_{L^2(\mathrm{P}_{X_{t_l}})}, \end{aligned}$$

for all $l = 1,\dots,L$, where the constants c_1, c_2 depend on A_t, V_{Ψ_M} is the Vapnik-Chervonenkis dimension of Ψ_M and $\overline{u}$ is a truncated version of $\widehat{u}$. The advantage of using nonlinear classes consists in their ability to significantly reduce the approximation errors $\inf_{f\in\Psi_M} \|u(t_l,\cdot) - f(\cdot)\|^2_{L^2(\mathrm{P}_{X_{t_l}})}$, while keeping the complexity V_{Ψ_M} comparable to the linear classes. One popular choice of Ψ_M is neural networks.

4.3 Rates of convergence

There are several ways to choose the basis functions $\psi_1, \dots, \psi_K$. In this section we consider the so-called piecewise polynomial partitioning estimates and present L^2-upper bounds for the corresponding projection errors

$$\inf_{f \in \mathrm{span}\{\psi_1,\dots,\psi_K\}} \|u(t,\cdot) - f(\cdot)\|^2_{L^2(\varrho)} =: \|u(t,\cdot) - \underline{u}(t,\cdot)\|^2_{L^2(\varrho)}, \tag{4.20}$$

for some fixed $t \geq 0$ and some generic measure ϱ on $\mathbb{R}^n$. For instance, in (4.18) t and ϱ may taken to be t_l and $\mathrm{P}_{X_{t_l}}$, $l = 1, ..., L$, respectively, and in (4.12) we may take $t = 0$ and ϱ equal to μ. The piecewise polynomial partitioning estimate of u works as follows: We fix some $q \in \mathbb{N}$ that denotes the maximal degree of polynomials involved in our basis functions. Next fix some $R > 0$ and a uniform partition of $[-R, R]^n$ into S^n cubes $C_1, \dots, C_{S^n}$. That is, $[-R, R]$ is partitioned into S subintervals with equal length. Further, consider the set of basis functions $\psi_{j,1}, \dots, \psi_{j,c_{q,n}}$ with $j \in \{1, \dots, S^n\}$ and $c_{q,n} := \binom{q+n}{n}$ such that $\psi_{j,1}(x), \dots, \psi_{j,c_{q,n}}(x)$ are polynomials with degree less than or equal to q for $x \in C_j$, and $\psi_{j,1}(x) = \ldots = \psi_{j,c_{q,n}}(x) = 0$ for $x \notin C_j$. Then we consider the least squares projection estimate $\underline{u}(t,x)$ for $x \in \mathbb{R}^n$, based on $K = S^n c_{q,n} = O(S^n q^n)$ basis functions. Let us define the operator D^α as

$$D^\alpha f(x) := \frac{\partial^{|\alpha|} f(x)}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}},$$

for any real-valued function f, $\alpha \in \mathbb{N}_0^n$ and $|\alpha| = \alpha_1 + \ldots + \alpha_n$. For $r \in \mathbb{N}_0$ and $L_f : \mathbb{R}^n \to \mathbb{R}_+$ we say that a function $f \colon \mathbb{R}^n \to \mathbb{R}$ is $(r+1, L_f)$-*smooth w.r.t. the (Euclidean) norm* $|\cdot|$ whenever, for all α with $|\alpha| = \sum_{i=1}^n \alpha_i = r$ and all $R > 0$, we have

$$\left|D^\alpha f(x) - D^\alpha f(y)\right| \leq L_f(x)|x - y|, \quad x \in \mathbb{R}^n, \quad |y - x|_\infty \leq 1,$$

i.e., the function $D^\alpha f$ is locally Lipschitz with the Lipschitz function L_f with respect to the norm $|\cdot|$ on $\mathbb{R}^n$. Let us make the following assumptions.

(A3) The function $u(t,\cdot)$ is $(q+1, L_u)$-smooth with

$$\int_{\mathbb{R}^n} L_u^2(x)\, \varrho(dx) \leq C_u^2 < \infty$$

for some constant $C_u > 0$.

(A4) It holds

$$\int_{\{|z|_\infty > R\}} u^2(t,z)\, \varrho(dz) \leq B_\nu R^{-\nu}$$

for some $\nu > 0$ all $R > 0$.

The following result holds.

Lemma 4.2. *Suppose that (A3) and (A4) hold, then*

$$\|u(t,\cdot) - \underline{u}(t,\cdot)\|^2_{L^2(\varrho)} \lesssim \frac{C_u^2}{[(q+1)!]^2 n} \left(\frac{Rn}{S}\right)^{2(q+1)} + B_\nu R^{-\nu}, \tag{4.21}$$

where $\lesssim$ stands for inequality up to an absolute constant.

Remark 4.3. Notice that the terms on the right-hand-side of (4.21) are of order

$$\left(\frac{R}{S}\right)^{2(q+1)} + R^{-\nu}, \tag{4.22}$$

provided that we only track R and S and ignore the remaining parameters, such as q and $\kappa_p(\delta, \mathbf{W})$. Let us assume that both terms in (4.22) are of the same order. Then we get $R = O(S^{\frac{2(q+1)}{\nu+2(q+1)}})$ and thus $R^{-\nu} = O(S^{-\frac{2\nu(q+1)}{\nu+2(q+1)}})$. Together with the fact that the overall number of basis functions K is of order S^n, we have $R^{-\nu} = O(K^{-\frac{2\nu(q+1)}{n(\nu+2(q+1))}})$. Thus there is a constant $D > 0$ such that

$$\|u(t,\cdot) - \underline{u}(t,\cdot)\|^2_{L^2(\varrho)} \leq \frac{D}{K^\kappa}.$$

with $\kappa = \frac{2\nu(q+1)}{n(\nu+2(q+1))}$.

The following result is based on Corollary 3.3 and gives sufficient conditions for (A3) and (A4) to hold.

Corollary 4.4. *Let $u(t,x,\mathbf{W})$ be as above. Assume that g is $q+1$-times differentiable (in x) and that there are constants $\zeta_1, \zeta_2 \geq 0$ such that*

$$|D^l g(x)| \leq \zeta_1 e^{\zeta_2 |x|} \tag{4.23}$$

for all $x \in \mathbb{R}^n$ and $l = 1, \dots, q+1$. Assume that σ is bounded, $(4+q)$-times differentiable with bounded derivatives, b and β are bounded, $(3+q)$-times differentiable with bounded derivatives, and let $K_1 > 0$ be a bound for their norms, i.e.

$$\|\sigma\|_{C_b^{3+q}} \vee \|\widehat{b}\|_{C_b^{3+q}} \vee \|\beta\|_{C_b^{3+q}} \leq K_1 \tag{4.24}$$

with $\widehat{b}$ denoting the Stratonovich corrected drift as given in (3.6)*. Suppose that*

$$\int e^{2\zeta_2|x|} \varrho(dx) < \infty,$$

then (A3) holds with

$$C_u \leq D_1 \exp\left(\left(\frac{\log\left(2\kappa_p(\delta,\mathbf{W})\right)}{\delta}\right)^{p/2}\right) + D_2.$$

for some constants $D_1 = D_1(q, K_1, \zeta_1, \zeta_2)$ and $D_2 = D_2(q, K_1, \zeta_1, \zeta_2)$. Moreover, (A4) holds for some $\nu > 0$ and B_ν depending on K_1, T, ζ_1, ζ_2.

Using to the parameter allocations in Remark 4.3 we end up with the following convergence rates for the regression procedures proposed in Section 4.1 and Section 4.2, respectively.

Corollary 4.5. *Suppose that the conditions* (4.23) *and* (4.24) *are satisfied. Moreover assume that*

$$\int e^{2\zeta_2|x|} \mu(dx) < \infty,$$

then under assumptions of Theorem 4.1, the latter reads,

$$E\int_{\mathcal{D}}|\tilde{v}(z)-v(z)|^2\,\mu(dz)\le D_3\frac{K}{M}+\frac{D_4}{K^\kappa},$$

for some constants $D_3, D_4 > 0$.

Corollary 4.6. *Suppose that the conditions* (4.23) *and* (4.24) *are satisfied. Moreover assume that*

$$\int e^{2\zeta_2|x|}\mathrm{P}_{X_{t_l}}(dx)<\infty,\quad l=1,\dots,L,$$

then under assumptions (A1) and (A2)

$$E\|\overline{u}(t_l,\cdot)-u(t_l,\cdot)\|^2_{L^2(\mathrm{P}_{X_{t_l}})}\le D_5(\log M+1)\frac{K}{M}+\frac{D_6}{K^\kappa},$$

for some constants $D_5, D_6 > 0$.

5 Simplified Euler scheme for rough differential equations

For the computation of the optimal coefficients γ in (4.7) and (4.11) it is required to construct the vector $\mathcal{Y}$ with components $\mathcal{Y}_m$ defined in (4.6) that depends on paths of the solution to equation (2.4). For that reason, we introduce an Euler scheme which allows us to numerically solve (2.4).

As in Section 3, we consider the hybrid Stratonovich-rough differential equation with $0 \le t \le r \le T$:

$$dX_r^{t,x}=\Big[b(X_r^{t,x})-a(X_r^{t,x})\Big]dt+\sigma(X_r^{t,x})\circ dB_r+\beta(X_r^{t,x})d\mathbf{W}_r,\quad X_t^{t,x}=x, \tag{5.1}$$

where $a(\cdot)=\frac{1}{2}\sum_{i=1}^m D\sigma_i(\cdot)\cdot\sigma_i(\cdot)$ is the Itô-Stratonovich correction and σ_i is the ith column of σ. Again, the above hybrid equation is defined as an RDE driven by the joint rough path of B and $\mathbf{W}$. This geometric joined rough path $\mathbf{Z}^{\mathrm{g}}$ is given as in (2.5) but $\mathbb{B}^{\mathrm{Itô}}$ is substituted by the Stratonovich integral $\mathbb{B}^{\mathrm{Str}}$. Below, we set $\widehat{b}(\cdot):=b(\cdot)-a(\cdot)$ and $V(\cdot):=\begin{bmatrix}\sigma(\cdot) & \beta(\cdot)\end{bmatrix}$ for the simplicity of the notation.

First of all, let $t=r_1<r_2<\ldots<r_{\bar{n}}=T$ be an equidistant time grid with step size h. In the numerical experiments later on the path W will be specified as a trajectory of a fractional Brownian motion with Hurst index $\frac{1}{3}<H\le\frac{1}{2}$. For this situation the following scheme provides a meaningful approximation $\bar{X}_{r_k}$ of $X_{r_k}^{t,x}$:

$$\bar{X}_{r_{k+1}}=\bar{X}_{r_k}+\widehat{b}(\bar{X}_{r_k})h+\sum_{\ell=1}^{3}\frac{1}{\ell!}V_{i_1}\dots V_{i_\ell}I(\bar{X}_{r_k})\Delta_k Z^{i_1}\dots\Delta_k Z^{i_\ell}, \tag{5.2}$$

where V_i is the ith column of V, $I(x)=x$, $V_iV_j(x)=DV_j(x)\;V_i(x)$ and $\Delta_k Z^i=Z^i_{r_{k+1}}-Z^i_{r_k}$. Notice that we use Einstein's summation convention in (5.2) which we indicate by the upper indices for the components of Z.

This simplified Euler scheme was first introduced in [11] and also investigated in [5]. In the following, we state a result from [5] on the strong order of convergence to (5.2).

Theorem 5.1. *Let W be a d-dimensional, continuous, centered Gaussian process with independent components. Moreover, we assume that each component W^i, $i \in \{1, \dots, d\}$, has stationary increments with a concave variance function*

$$\sigma_i^2(\tau) := E\left|W_{t+\tau}^i - W_t^i\right|^2, \quad t, \tau \geq 0,$$

where $\sigma_i^2(\tau) = O\left(\tau^{\frac{1}{\rho}}\right)$ as $\tau \to 0$ for some $\rho \in [1, 2)$. Let X be the solution to (5.1) and $\bar{X}$ be its approximation based on (5.2), where $Z_t = \binom{W_t(\omega)}{B_t}$ for fixed $\omega \in \Omega$. Then, for almost all paths of W and for any $1 \leq p < \infty$, there is a constant $\tilde{C}$ such that

$$\left| E \max_{k=1,\dots,\bar{n}} \left|X_{r_k} - \bar{X}_{r_k}\right|^p \right|^{\frac{1}{p}} \leq \tilde{C} h^{\frac{1}{\rho} - 0.5 - \delta},$$

where h is the time step of the Euler method and $\delta > 0$ is arbitrary small.

Proof. This theorem is a consequence of [5, Theorem 1] together with [11, Theorem 1.1]. □

Remark 5.2.

- Theorem 5.1 covers the case of W being a fractional Brownian motion with Hurst index $\frac{1}{4} < H \leq \frac{1}{2}$ ($\frac{1}{\rho} = 2H$).
- An almost sure rate for the scheme in (5.2) is proved in [11, Theorem 1.1] in case Z is a fractional Brownian motion.

6 Numerical examples

We illustrate the methods by some numerical examples. First we study examples involving linear vector fields, for which the rough differential equation has an explicit solution. This allows for easy comparison with a reliable reference solution. Later on, we consider an example with non-linear vector fields without readily available reference values. All examples take place in a two- or three-dimensional state space, and we assume that the driving Brownian motion is one-dimensional (i.e., the PDE fails to be elliptic), whereas the rough driver is two-dimensional in order to rule out trivial cases.

6.1 Numerical examples with linear vector fields

Let us investigate a particular example for the RPDE (2.1). We set $c, \gamma \equiv 0$ such that by Theorem 2.1 the corresponding regular solution is simply represented by

$$u(t, x; \mathbf{W}) = E\left[g(X_T^{t,x})\right], \quad (t, x) \in [0, T] \times \mathbb{R}^n, \tag{6.1}$$

where $X_\cdot^{t,x}$ is the solution to (5.1) with initial time t and initial value x. Below, we from now on assume that

$$b(x) = Ax, \quad \sigma_i(x) = C_i x \quad \text{and} \quad \beta_j(x) = N_j x, \tag{6.2}$$

for $i = 1, \dots, m$, $j = 1, \dots, d$, $x \in \mathbb{R}^n$ and where all coefficients A, C_i, N_j are $n \times n$ matrices.

6.1.1 Explicit solutions to linear RDEs

We can find an explicit representation for the resulting linear RDE (compare 5.1) by introducing the fundamental solution Φ to the linear system. Using the Einstein convention, we formally define Φ as the $\mathbb{R}^{n\times n}$-valued process satisfying

$$\Phi_r = I + \int_0^r \left(A - \frac{1}{2}\sum_{i=1}^m C_i^2\right)\Phi_s ds + \int_0^r C_i \Phi_s \circ dB_s^i + \int_0^r N_j \Phi_s d\mathbf{W}_s^j. \tag{6.3}$$

For $t \le r$ we can easily see that the following identity holds:

$$\Phi_r \Phi_t^{-1} = I + \int_t^r \left(A - \frac{1}{2}\sum_{i=1}^m C_i^2\right)\Phi_s \Phi_t^{-1} ds + \int_t^r C_i \Phi_s \Phi_t^{-1} \circ dB_s^i + \int_t^r N_j \Phi_s \Phi_t^{-1} d\mathbf{W}_s^j.$$

Consequently, equation (5.1) with the linear coefficients (6.2) is represented as

$$X_r^{t,x} = \Phi_r \Phi_t^{-1} x, \quad 0 \le t \le r \le T. \tag{6.4}$$

Case of commuting matrices We now point out a case, in which Φ is given explicitly. Let all matrices A, C_i and N_j commute, then we have

$$\Phi_r = f(r, B_r^i, W_r^j) := \exp\left((A - \frac{1}{2}\sum_{i=1}^m C_i^2)r + C_i B_r^i + N_j W_r^j\right). \tag{6.5}$$

Using the classical chain rule for geometric rough paths

$$df(r, B_r^i, W_r^j) = \frac{\partial}{\partial t} f(r, B_r^i, W_r^j) dr + \frac{\partial}{\partial b_i} f(r, B_r^i, W_r^j) \circ dB_r^i + \frac{\partial}{\partial w_j} f(r, B_r^i, W_r^j) d\mathbf{W}_r^j,$$

we indeed see that f solves (6.3) taking into account that

$$\exp\left(\sum_{i=1}^q A_i\right) = \prod_{i=1}^q \exp(A_i) \quad \text{and} \quad A_j \exp(A_i) = \exp(A_i) A_j \tag{6.6}$$

for commuting matrices $A_1, \dots, A_q$.

Case of nilpotent matrices We know from the above considerations that the fundamental matrix Φ is given by (6.5) if all matrices commute, i.e., the rough path structure does not enter the solution at all. For that reason, we investigate another case with an explicit solution. Let us again look at the linear RDE which is of the form:

$$dX_r = \sum_{i=1}^{d_1} A_i X_r d\mathbf{Z}_r^{g,i}, \quad r \in [t,T], \tag{6.7}$$

where $d_1 = m + d$, $\mathbf{Z}^g$ is the geometric joint rough path of B and W, $A_i = C_i$, $A_{j+m} = N_j$ for $i = 1,\dots,m$ and $j = 1,\dots,d$. For simplicity, we assume to have a zero drift, i.e., $A - \frac{1}{2}\sum_{i=1}^m C_i^2 = 0$.

The Chen-Strichartz formula, see [28], provides a general solution formula in terms of a infinite series in the general case, involving higher order iterated integrals of the driving rough path.[3] For simplicity, we shall only provide the solution in the step-2 nilpotent case, i.e., we assume that

$$\forall i,j,l: \ [[A_i,A_j],A_l] = 0, \tag{6.8}$$

where $[A,B] := AB - BA$ denotes the usual commutator of matrices.

Lemma 6.1. *For $1 \le i \ne j \le d_1$ let*

$$a_{t,r}^{ij} := \frac{1}{2}\int_t^r Z_{t,s}^i dZ_s^j - \frac{1}{2}\int_t^r Z_{t,s}^j dZ_s^i$$

denote the area *swapped by the paths Z^i and Z^j, where the integrals are, of course, understood in the sense of the rough path $\mathbf{Z}^g$. Then, we have*

$$X_r^{t,x} = \exp\left(\sum_{i=1}^{d_1} A_i Z_{t,r}^i - \sum_{1\le i<j\le d_1} [A_i,A_j] a_{t,r}^{ij}\right) x.$$

Remark 6.2. The unusual minus sign in Lemma 6.1 comes from the fact that the linear vector field $y \mapsto -[A,B]y$ is the Lie bracket of the linear vector fields $y \mapsto Ay$ and $y \mapsto By$. In the more general formulation involving general vector fields, the minus sign above, therefore, turns into a plus sign.

Sketch of proof of Lemma 6.1. Formally, suppose that the paths $t \mapsto Z_t^i$ are actually smooth, so that (6.7) can be replaced by the non-autonomous ODE

$$\dot{X}_t = A(t)X_t, \quad A(t) := \sum_{i=1}^{d_1} A_i \dot{Z}_t^i.$$

(Here, $A(t)$ is considered a time dependent vector field.) The Chen-Strichartz formula (also known as "generalized Baker-Campbell-Hausdorff-Dynkin formula") [28,

[3] The Chen-Strichartz formula is usually given for the smooth case, but one can repeat the proof for the rough case, see, for instance, [3] for the Brownian case in a free setting.

formula (G.C-B-H-D)] involves n-fold Lie brackets of the vector fields $A(s_j)$ for different times s_j, $j = 1, \dots, n$. Note that

$$[A(s_1), A(s_2)](x) = -\sum_{1 \le i < j \le d_1} \left(\dot{Z}^i_{s_1} \dot{Z}^j_{s_2} - \dot{Z}^j_{s_1} \dot{Z}^i_{s_2} \right) [A_i, A_j] \cdot x,$$

while all Lie brackets of terms involving two or more Lie brackets vanish. The result is then obtained by inserting into the formula. □

6.1.2 Numerical example with commuting matrices

For the following numerical considerations we now assume that $m = 1$, $n = d = 2$, $T = 1$ and $g(x) = \exp\left(-0.5 \|x\|^2\right)$.

We specify the matrices in (6.2) of the linear system. We introduce a matrix V which satisfies the property $V^{-1} = V$:

$$V = \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}.$$

Using V, we then set

$$A = V \begin{pmatrix} 0.5 & 0 \\ 0 & 4.5 \end{pmatrix} V = \begin{pmatrix} 2.5 & 2 \\ 2 & 2.5 \end{pmatrix}, \quad C_1 = C = V \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} V = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix},$$
$$N_1 = V \begin{pmatrix} 0.5 & 0 \\ 0 & 1.5 \end{pmatrix} V = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \quad \text{and} \quad N_2 = V \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} V = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}.$$

Due to their special structure, all these matrices commute. Furthermore, we observe that $A = \frac{1}{2} C^2$ such that the drift is zero. Hence, the corresponding fundamental solution is of the following simple form

$$\begin{aligned} \Phi_r &= V \exp\left(\begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} B_r + \begin{pmatrix} 0.5 & 0 \\ 0 & 1.5 \end{pmatrix} W^1_r + \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} W^2_r \right) V \\ &= V \begin{pmatrix} \exp\left(B_r + 0.5 W^1_r + 3 W^2_r \right) & 0 \\ 0 & \exp\left(3 B_r + 1.5 W^1_r + W^2_r \right) \end{pmatrix} V. \end{aligned}$$

Inserting (6.4) into (6.1) with the above fundamental matrix and using numerical integration to estimate the expected value delivers an "exact" solution u of the underlying RPDE. The goal of this section is to compare the exact solution with the solution that is obtained from the pseudo-regression procedure from Section 4.

We conduct the experiments for two different paths of W (see Figure 1). In both cases we choose W^1 and W^2 to be fixed paths of independent scalar fractional Brownian motions with Hurst index $H = 0.4$.

Simulations for the first path (left picture in Figure 1) We compute a numerical approximation $\tilde{u}$ of the RPDE solution u based on the pseudo-regression

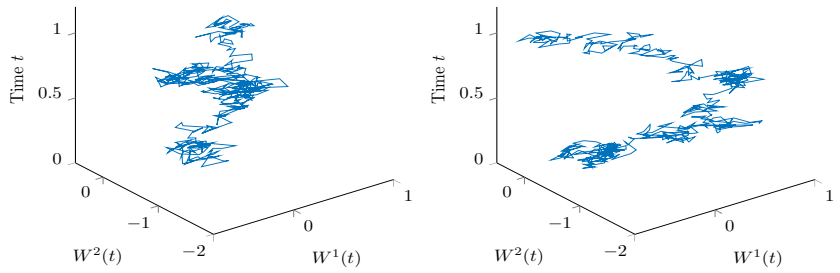

Figure 1: Two paths of a fractional Brownian motion with Hurst index $H = 0.4$.

procedure, see Theorem 4.1, where for every fixed $t \in [0,1]$ the approximation $\tilde{u}(t,\cdot)$ is derived according to (4.10) and (4.11). We start with the left driver in Figure 1. Within the numerical approximation we encounter three different errors. The regression error itself depends on the number of basis functions K and the number of samples M that we use to approximate the expected value with respect to the probability measure μ of the initial data. In order to generate the paths of (5.1) that we require for the regression approach, we need to apply the Euler scheme from Section 5. The error in this discretization depends on the step size h which is our third parameter.

We choose μ to be the uniform measure on $[0,1]^2$ and zero elsewhere. Based on this we choose Legendre polynomials on $[0,1]^2$ as an ONB $(\psi_i)_{i=1,\dots,K}$ of $L^2(\mathbb{R}^2,\mu)$. To be more precise, we consider the Legendre polynomials up to a certain fixed order p for every spatial direction and then take into account the total tensor product between the basis functions of different spatial variables, such that $K = (p+1)^n$.

In Figure 2 we plot the regression solution $\tilde{u}$ on $[0,1]^2$ for three different time points. Here, we use $K = 36$ Legendre polynomials, $M = 10^6$ samples and a step size $h = 2^{-9}$ of the Euler scheme (5.2). We observe in our simulations that $\tilde{u}$ is a very good approximation for the reference solution u for these fixed parameters. The plots for u look exactly as in Figure 2. Since there is no visible difference between u and $\tilde{u}$, we omit the pictures for u.

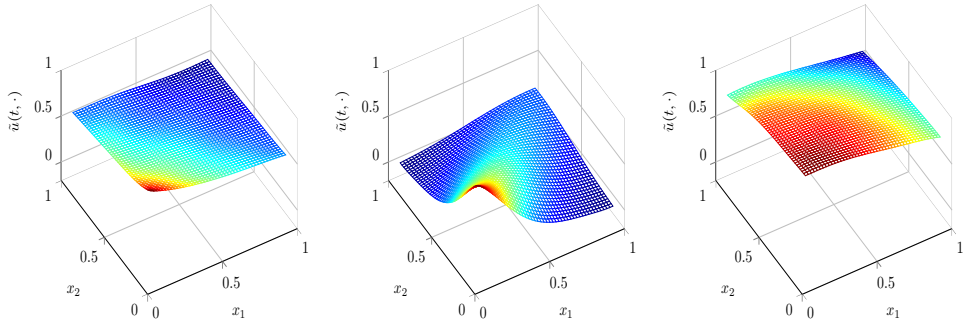

Figure 2: Pseudo-regression solution $\tilde{u}(t,\cdot)$, $t = 0, 0.84, 0.99$, of the RPDE driven by the left path in Figure 1. The parameters are $K = 36$, $h = 2^{-9}$, $M = 10^6$.

Below, we investigate how sensitive the pseudo-regression approach is in every single parameter. Therefore, we always fix two parameters and vary the remaining third one. All the errors are measured in $L^2([0,1]^2)$, i.e., we compute $\|u(t,\cdot) - \tilde{u}(t,\cdot)\|_{L^2([0,1]^2)}$, $t \in [0,1]$, or the corresponding relative error. In Figure 3, the absolute and relative errors are shown for different step sizes $h = 2^{-7}, 2^{-8}, 2^{-9}$. If we compare the curves with the largest step size with the ones having the smallest step size, we can see that there is a remarkable difference. We observe that the error is most of the times twice and sometimes up to three time larger when using a four times larger step size. This can lead to an relative approximation error of more than 10%. This implies that a small step size h is recommended in order to ensure a small error. This is not surprising since the order of convergence in h is worst out of all parameters.

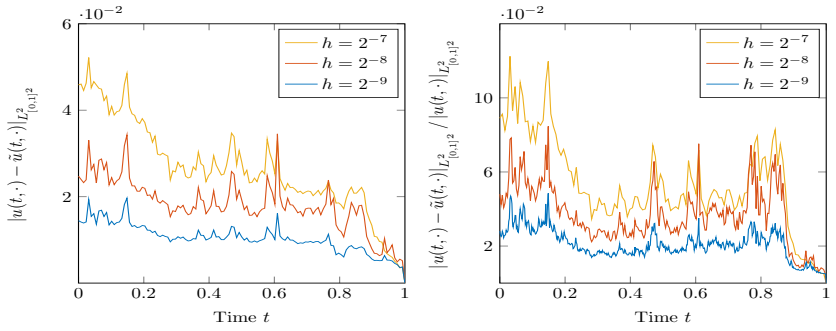

Figure 3: Absolute and relative error between RPDE and pseudo-regression solution. The parameters are $K = 36$, $M = 10^6$ and $h = 2^{-7}, 2^{-8}, 2^{-9}$.

Now we fix the number of basis functions and the step size of the Euler method. For different numbers of samples $M = 10^4, 10^5, 10^6$, we find the errors in Figure 4. We see that it does not really matter whether 10^6 or 10^5 samples are used, whereas 10^4 samples are probably too few, since the relative error can be up to 9%.

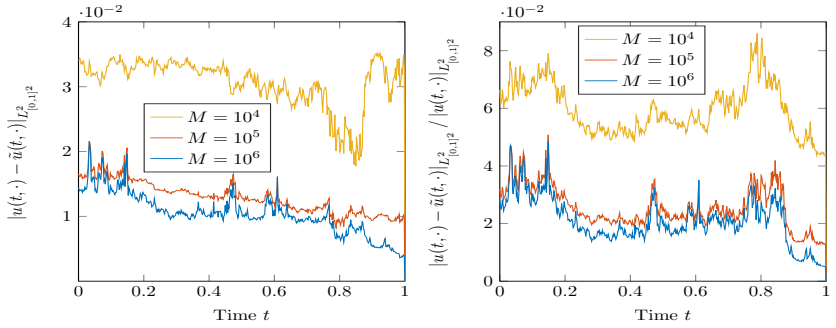

Figure 4: Absolute and relative error between RPDE and pseudo-regression solution. The parameters are $K = 36$, $M = 10^4, 10^5, 10^6$ and $h = 2^{-9}$.

It remains to analyze the error in the number of basis functions. In Figure 2, the solution looks relatively flat, such that it is not surprising that the parameter K only plays a minor role. Since there is barely a difference when varying K, we state the logarithmic errors in Figure 5 for $K = 16, 81$. The error for $K = 36$ lies between the

curves in Figure 5 and is omitted because it would have been hard to distinguish between the plots if it would have been included. Even in the logarithmic scale there is almost no difference in the errors. Moreover, we observe that for most of the time points, an additional error is caused by taking too many polynomials. Thus, the approach is not at all sensitive in the parameter K for this problem with the left driving path in Figure 1. This is not true for every driving path as the following experiment will show. Using the right path in Figure 1 as the driver instead will lead to a much larger error if we choose the same parameters as before.

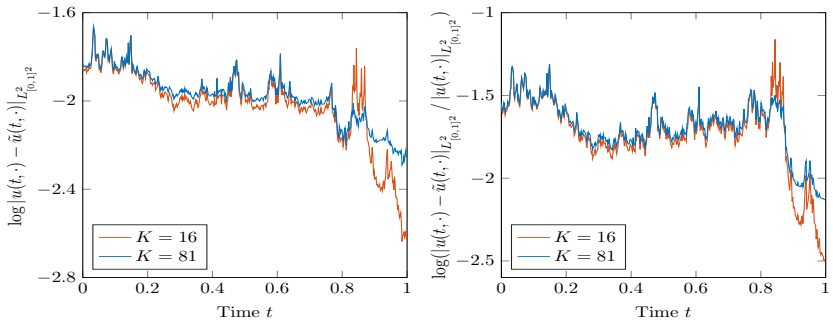

Figure 5: Logarithmic absolute and relative error between RPDE and pseudo-regression solution. The parameters are $K = 16, 81$, $M = 10^6$ and $h = 2^{-9}$.

Simulations for the second path (right picture in Figure 1) We conduct a second experiment with the same example as above. We only change the driving path, i.e., the left path in Figure 1 is replaced by the right one. This leads to a very large relative L^2-error for $M = 10^6$, $h = 2^{-9}$ and $K = 36$, see Figure 7. In the worst case ($t = 0.48$) the relative error is almost 80%. The reason for this can be seen in Figure 6, where $u(0.48, \cdot)$ (left) is compared with $\tilde{u}(0.48, \cdot)$ (right). The exact solution in this worst case is close to be a delta function which is generally hard to approximate. The pseudo-regression solution clearly looks differently which shows that the parameter K depends on the underlying driving path. If we increase the number of polynomials to $K = 121$, we can reduce the error in Figure 7 but still many more basis functions would be required to obtain a small relative error which is still large at every time point, where u is close to be a delta function.

6.1.3 Numerical example with nilpotent matrices

In Subsection 6.1.2 an example has been considered that does not depend on the complete rough path but only on the path. Therefore, we investigate another case, where still a reference solution can be derived but this time it depends on the full rough path.

Let us consider an scenario that fits the framework (6.8). We set $b, c, \gamma \equiv 0$ in system (2.1) with terminal time $T = 1$ and terminal value $g(x) = \exp\left(-0.5\,\|x\|^2\right)$.

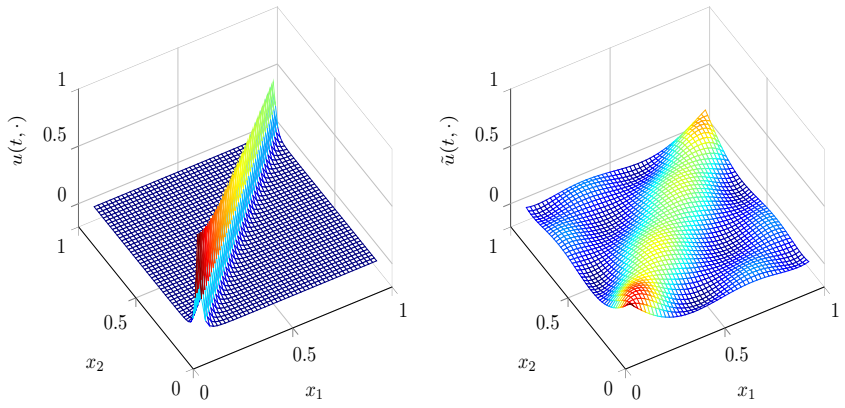

Figure 6: Exact solution $u(t,\cdot)$ (left) and pseudo-regression solution $\tilde{u}(t,\cdot)$ (right), $t = 0.48$, of the RPDE driven by the right path in Figure 1 (coefficients as in (6.2)). The regression parameters are $K = 36$, $h = 2^{-9}$, $M = 10^6$.

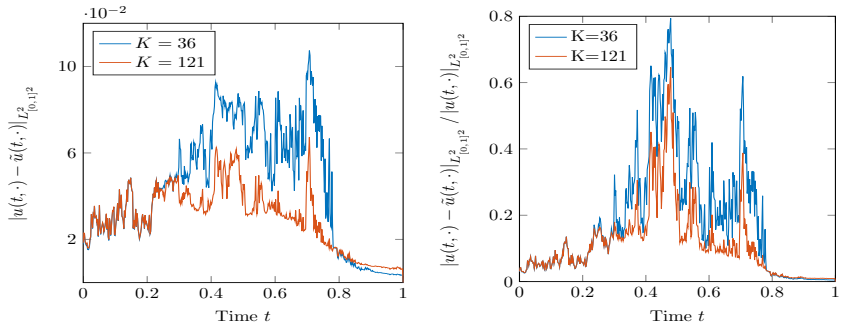

Figure 7: Absolute and relative error between RPDE and pseudo-regression solution with parameters $M = 10^6$, $h = 2^{-9}$ and $K = 36, 121$.

Moreover, we define

$$\sigma(x) = A_1 x, \quad \beta_1(x) = A_2 x \quad \text{and} \quad \beta_2(x) = A_3 x, \tag{6.9}$$

where we assume $m = 1$, $d = 2$ and a three-dimensional space variable $x \in \mathbb{R}^3$. Again, by Theorem 2.1, the solution to (2.1) has the following stochastic representation:

$$u(t,x;\mathbf{W}) = E\left[g(X_1^{t,x})\right], \quad (t,x) \in [0,1] \times \mathbb{R}^3, \tag{6.10}$$

where $X_1^{t,x}$ satisfies the RDE

$$dX_r = A_1 X_r dB_r + A_2 X_r \mathbf{W}_r^1 + A_3 X_r \mathbf{W}_r^2, \quad X_t = x, \quad r \in [t,1]. \tag{6.11}$$

Now we fix the coefficients such that (6.8) is fulfilled:

$$A_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

Notice that $A_1^2 = 0$ implies that the Itô-Stratonovich correction term is zero in (6.11) such that we automatically obtain a geometric driver in the equation. The driving path is again a fractional Brownian motion with Hurst index $H = 0.4$. Now, since A_1 commutes with A_2 and A_3 it can be seen that $[A_1, A_2] = [A_1, A_3] = 0$. Furthermore, we observe that $[A_2, A_3] = A_1$ which by Lemma 6.1 leads to the following solution representation:

$$X_r^{t,x} = \exp\left(A_1(B_{t,r} - a_{t,r}^{12}) + A_2 W_{t,r}^1 + A_3 W_{t,r}^2\right) x,$$

where the term

$$a_{t,r}^{12} = \frac{1}{2}\left(\int_t^r W_{t,s}^1 dW_s^2 - \int_t^r W_{t,s}^2 dW_s^1\right)$$

is approximated numerically by using piece-wise linear approximations to W^1 and W^2 on a very fine time grid. Consequently, we have

$$X_r^{t,x} = \exp\begin{pmatrix} 0 & W_{t,r}^1 & B_{t,r} - a_{t,r}^{12} \\ 0 & 0 & W_{t,r}^2 \\ 0 & 0 & 0 \end{pmatrix} x =: \exp\left(D_{t,r}\right) x.$$

The matrix $D_{t,r}$ is nilpotent with index 3, so that $\exp(D_{t,r}) = I + D_{t,r} + \frac{1}{2} D_{t,r}^2$ which then leads to

$$X_r^{t,x} = \begin{pmatrix} 1 & W_{t,r}^1 & B_{t,r} - a_{t,r}^{12} + 0.5 W_{t,r}^1 W_{t,r}^2 \\ 0 & 1 & W_{t,r}^2 \\ 0 & 0 & 1 \end{pmatrix} x.$$

Inserting this into (6.10) and estimating the expected value with numerical integration provides the exact solution u of the underlying RPDE.

The probability measure within the regression approach has the same structure as before but it is now defined on $\mathbb{R}^3$, i.e., we choose μ to be the uniform measure on $[0,1]^3$ and zero elsewhere. So, the ONB of $L^2(\mathbb{R}^3, \mu)$ is given by Legendre polynomials on $[0,1]^3$. In Figure 8, the absolute and the relative error in $L^2([0,1]^3)$ between the exact and the pseudo-regression solution is considered. The algorithm also works very well in this case since the relative error is less than 1% for $K = 64$ basis polynomials, $M = 10^6$ samples and a Euler step size of $h = 2^{-9}$.

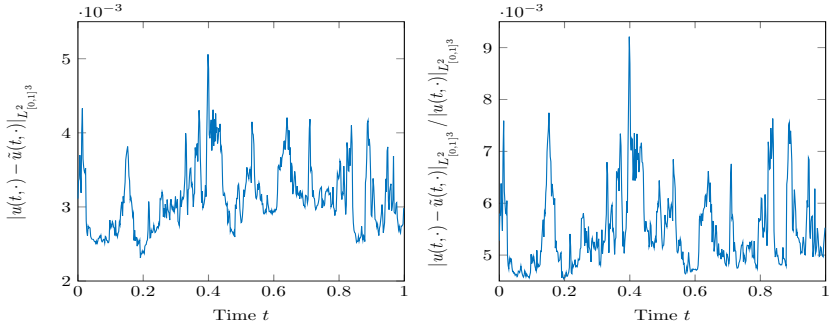

Figure 8: Absolute and relative error between the RPDE with coefficients as in (6.9) and the pseudo-regression solution. The parameters are $K = 64$, $M = 10^6$ and $h = 2^{-9}$.

The pseudo-regression $\tilde{u}$ for several fixed time points is illustrated in Figure 9. In the cubic domain the color represents the corresponding function value of $\tilde{u}(t,\cdot)$ which is relatively large if the color is red and relatively small if the color is blue. We omit the plots for the exact solution since there is no visible difference to the regression solution.

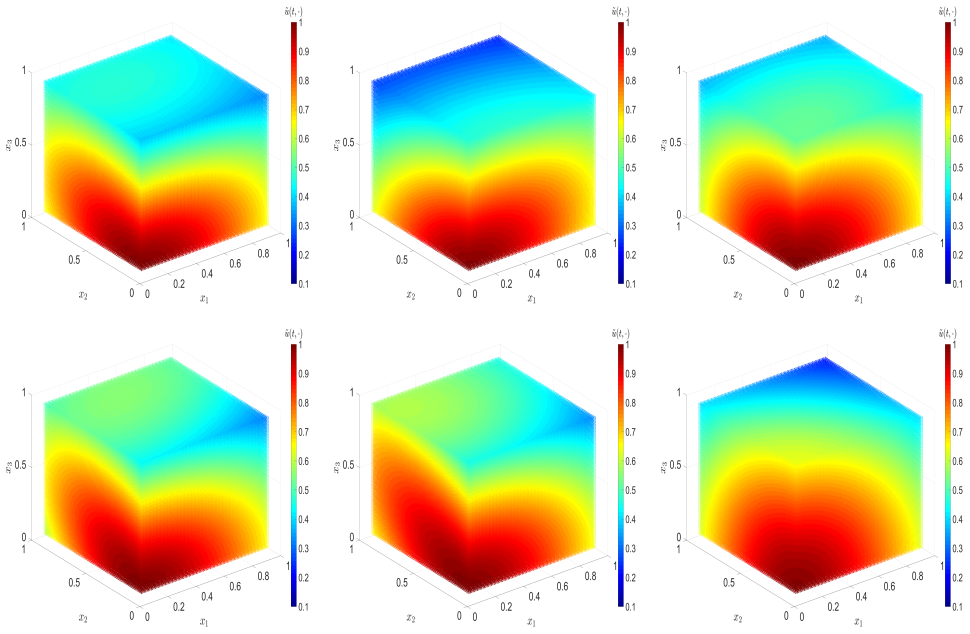

Figure 9: The pseudo-regression solution $\tilde{u}(t,\cdot)$, $t = 0, 0.14, 0.43, 0.61, 0.78, 0.99$, of the RPDE driven a path of a fBm. The parameters are $K = 64$, $h = 2^{-9}$, $M = 10^6$.

We conclude this section by discussing an alternative to the pseudo-regression, that is the stochastic regression, where an approximation $\widehat{u}(t,\cdot)$ to $u(t,\cdot)$ is derived based on (4.7) instead of (4.11). Within the stochastic regression the Euler scheme (5.2) has to be used only once, whereas we run (5.2) for every fixed t when computing the pseudo-regression solution $\tilde{u}(t,\cdot)$. Consequently, $\widehat{u}$ can be computationally cheaper than $\tilde{u}$ if the Euler method is very expensive in terms of computational time. We determine the solution of the stochastic regression $\widehat{u}$ with the same parameters as before and compare it with the exact solution in Figure 10. In this context, we modify our basis, i.e., we use $\tilde{\psi}_i = g\psi_i$, where ψ_i are again Legendre polynomials ($i = 1, \dots, K$). This compensation is required since ψ_i takes very large values outside $[0,1]^3$. Now, we evaluate the basis functions at samples of the solution to (5.1) in order to compute $\mathcal{M}^{(t)}$ in (4.7). Since the paths of the solution to (5.1) leave $[0,1]^3$ quite frequently, we would encounter a very large variance and hence a large error when using the non-modified basis. With the basis $(\tilde{\psi}_i)_{i=1,\dots,K}$ we see that the error in Figure 10 is relatively small but it is clearly larger than the error of the pseudo-regression in Figure 8.

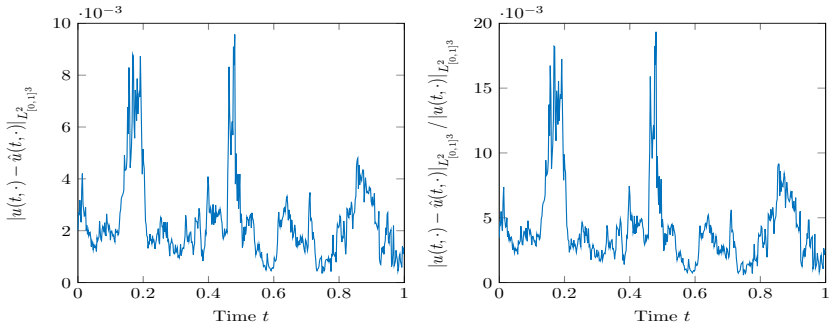

Figure 10: Absolute and relative error between the RPDE with coefficients as in (6.9) and the stochastic regression solution $\widehat{u}$. The parameters are as in Figure 8, i.e., $K = 64$, $M = 10^6$ and $h = 2^{-9}$.

6.2 Numerical example with non-linear vector fields

We conclude the numerical section with an example which has no reference solution. We start with a similar setting as in Section 6.1, i.e., we assume that $c, \gamma \equiv 0$. Hence, the solution of the underlying RPDE is given by the expected value in (6.1) but here non-linear vector fields enter equation (5.1). We define them as follows:

(6.12)
$$b(x) = \begin{pmatrix} \sin(x_1+2x_2) \\ \sin(2x_1+x_2) \end{pmatrix}, \quad \sigma(x) = \begin{pmatrix} x_1+\sin(x_2) \\ 2x_1+0.5\cos(x_2) \end{pmatrix}, \quad \beta(x) = \begin{pmatrix} 0.3\cos(x_1+x_2) & 0.2(x_1+x_2) \\ \sin(x_2)-0.5x_1 & \sin(x_1x_2)-0.5x_2 \end{pmatrix},$$

where we suppose to have a scalar Brownian motion B ($m = 1$), a two-dimensional spaces variable $x = \left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right) \in \mathbb{R}^2$ as well as a two dimensional rough path $\mathbf{W}$ ($n = d = 2$). Moreover, the terminal time and the terminal value of the RPDE are $T = 1$ and $g(x) = \exp\left(-0.5\,\|x\|^2\right)$, respectively. We fix the probability measure μ like in Subsection 6.1.2 such that the ONB is again represented by Legendre polynomials on $[0,1]^2$. We apply the pseudo-regression approach to this case and illustrate the resulting solution $\tilde{u}$ in Figure 11 for $K = 36, h = 2^{-8}$ and $M = 10^6$. Although there is no reference solution to determine the exact error, we expect the approximation to be good because $\tilde{u}$ is relatively flat in space and does not show an extreme behavior like in Figure 6.

A Proof of Theorem 3.1

A.1 Controlled p-variation paths

Recall that a function

$$\omega\colon \{0 \le s \le t \le T\} \to [0,\infty)$$

is called *control* if it is continuous, $\omega(t,t) = 0$ for all $t \in [0,T]$ and if it is superadditive, i.e.

$$\omega(s,u) + \omega(u,t) \le \omega(s,t)$$

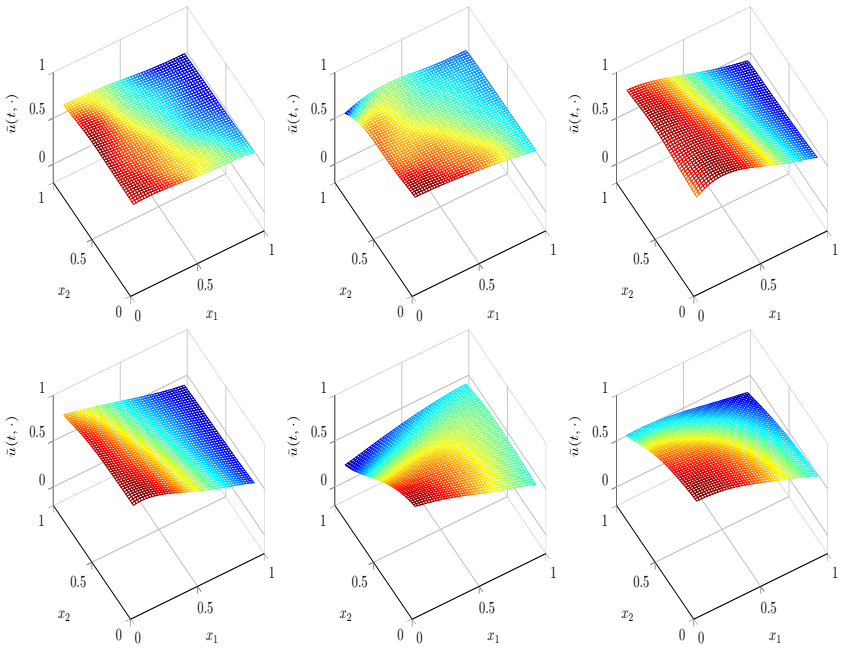

Figure 11: The pseudo-regression solution $\tilde{u}(t,\cdot)$, $t = 0, 0.33, 0.57, 0.73, 0.87, 0.99$, of the RPDE (vector fields as in (6.12)) driven by a path of a fBm with $H = 0.4$. The regression parameters are $K = 36, h = 2^{-8}$ and $M = 10^6$.

for all $s \le u \le t$. Examples of a control include $\omega(s,t) = |t-s|$ or $\omega(s,t) = \|\mathbf{Z}\|^p_{p\text{-var};[s,t]}$ provided $\mathbf{Z}$ is a p-rough path. We say that the *p-variation of $\mathbf{Z}$ is controlled by a control function ω* if $|\mathbf{Z}_{s,t}|^p \le \omega(s,t)$ holds for all $s \le t$. For arbitrary control functions, the quantity $N_\alpha(\omega;[s,t])$ is defined exactly as in (3.2) by replacing $\|\mathbf{Z}\|^p_{p\text{-var};[u,v]}$ by $\omega(u,v)$.

The following definition generalizes the notion of a *controlled path* from Hölder- to p-variation rough paths.

Definition A.1. Let U and W be normed spaces. Let $Z\colon [0,T] \to U$ be a path whose p-variation is controlled by a control function ω. We say that a path $y\colon [0,T] \to W$ is *controlled by Z and ω* if there exists a path $y'\colon [0,T] \to L(U,W)$ whose p-variation is controlled by ω so that for R^y given implicitly by the relation

$$y_{s,t} = y'_s Z_{s,t} + R^y_{s,t},$$

we have

$$\|R^y\|_{p/2-\omega;[0,T]} := \sup_{0\le s<t\le T} \frac{|R^y_{s,t}|}{\omega(s,t)^{\frac{2}{p}}} < \infty.$$

We will usually not explicitly mention the control ω and just say that y is controlled by Z. We denote by $\mathscr{D}^p_Z([0,T],W)$ the space of controlled p-rough paths. We will call a function y' with the given property a *Gubinelli-derivative* of y (with respect to Z).

It is an (admittedly lengthy) exercise to show that all classical estimates proven for Hölder rough paths can be generalized to p-rough paths and their controlled functions in the sense above for $p \in [2,3)$ when replacing $|t-s|$ by $\omega(s,t)$ in these estimates. Indeed, the results follow by using an appropriate version of the *Sewing Lemma* for control functions which was proven recently, even for discontinuous control functions, in [18, Theorem 2.2]. For instance, the corresponding results for rough integrals are summarized in the following theorem.

Theorem A.2. *Let $U, W, \widehat{W}$ be finite dimensional vector spaces and $\mathbf{Z} = (Z, \mathbb{Z})$ be a p-rough path with p-variation controlled by ω, $p \in [2,3)$. Let $y \in \mathscr{D}_Z^p([0,T], L(W, \widehat{W}))$ and $z \in \mathscr{D}_Z^p([0,T], W)$. Then*

$$\int_s^t y_u \, dz_u = (\mathcal{I}\Xi)_{s,t}\,, \qquad \Xi_{u,v} = y_u z_{u,v} + y'_u z'_u \mathbb{Z}_{u,v}$$

exists as an abstract integral (cf. [14, Lemma 4.2 and p. 49 eq. (4.6)]), and there is a constant C depending only on p such that the estimate

$$\begin{aligned}&\left| \int_s^t y_u \, dz_u - y_s z_{s,t} - y'_s z'_s \mathbb{Z}_{s,t} \right| \\ &\quad \le C \left(\|x\|_{p-\omega;[s,t]} \|R^z\|_{p/2-\omega;[s,t]} + \|\mathbb{Z}\|_{p/2-\omega;[s,t]} \|y' z'\|_{p-\omega;[s,t]} \right) \omega(s,t)^{3/p}\end{aligned}$$

holds for every $s < t$. In particular, the map $t \mapsto \int_0^t y_u \, dz_u$ is itself a controlled p-rough path, both controlled by z with derivative y, and by Z with derivative yz'.

Proof. A combination of [14, Theorem 4.10 and Remark 4.11] generalized to p-rough paths. □

In the next lemmas, we prepare some estimates for rough integrals and solutions to rough differential equations we are going to use at the end of this section. In the following, $U, W, \widehat{W}, W_1, W_2, \dots$ will denote finite dimensional vector spaces, and $\mathbf{Z}$ will be a fixed weakly geometric p-rough path, $p \in [2,3)$, with values in $U \oplus (U \otimes U)$, controlled by a control function ω. $C \ge 0$ will denote a generic constant whose actual value may change from line to line and which might depend on the parameters specified before.

Lemma A.3. *Let $V \colon W \to L(U, W)$ and let $y \colon [0,T] \to W$ be a solution to*

$$y_t = x + \int_0^t V(y_s) \, d\mathbf{Z}_s. \tag{A.1}$$

Assume that

$$\|y\|_{p-\omega} \vee \|\mathbf{Z}\|_{p-\omega} \le 1 \quad \text{and} \quad \|V\|_{C_b^2} \le 1.$$

Then there are constants C and α depending only on p such that

$$\|R^y\|_{p/2-\omega} \le C(1 + N_\alpha(\omega; [0,T])).$$

Proof. For $\alpha > 0$, set

$$\|R^y\|_{p/2-\omega;\alpha} := \sup_{0\le s<t\le T;|t-s|\le\alpha} \frac{|R^y_{s,t}|}{\omega(s,t)^{2/p}}.$$

Choose $s < t$ such that $\omega(s,t) \le \alpha$. Then we have, using the estimate in Theorem A.2,

$$\begin{aligned}
|R^y_{s,t}| &= \left|\int_s^t V(y_u)\, d\mathbf{Z}_u - V(y_s)Z_{s,t}\right| \\
&\le \left|\int_s^t V(y_u)\, d\mathbf{Z}_u - V(y_s)Z_{s,t} - DV(y_s)V(y_s)\mathbb{Z}_{s,t}\right| + |DV(y_s)V(y_s)\mathbb{Z}_{s,t}| \\
&\le C\alpha^{1/p}\omega(s,t)^{2/p}(\|R^y\|_{p/2-\omega;\alpha} + \|DV(y.)V(y.)\|_{p-\omega}) + \omega(s,t)^{2/p}.
\end{aligned}$$

Using boundedness of V and its derivatives, one can check that

$$\|DV(y.)V(y.)\|_{p-\omega} \le C\|y\|_{p-\omega} \le C.$$

Hence we obtain

$$\|R^y\|_{p/2-\omega;\alpha} \le C\alpha^{1/p}\|R^y\|_{p/2-\omega;\alpha} + C.$$

Choosing α such that $C\alpha^{1/p} \le 1/2$, we obtain

$$\|R^y\|_{p/2-\omega;\alpha} \le 2C.$$

Now choose (τ_n) such that $0 = \tau_0 < \tau_1 < \ldots < \tau_N < \tau_{N+1} = T$ with $\omega(\tau_i, \tau_{i+1}) \le \alpha$ and $N = N_\alpha(\omega; [0,T])$. Let $s < t$ be arbitrary. Choose i and j such that $s \in [\tau_{i-1}, \tau_i)$ and $t \in (\tau_j, \tau_{j+1}]$. Then

$$\begin{aligned}
\frac{|R^y_{s,t}|}{\omega(s,t)^{2/p}} &\le \omega(s,t)^{-2/p}(|R^y_{s,\tau_i}| + |R^y_{\tau_i,\tau_{i+1}}| + \ldots + |R^y_{\tau_j,t}| \\
&\quad + |V(y_{\tau_i}) - V(y_s)||Z_{\tau_i,\tau_{i+1}}| + \ldots + |V(y_{\tau_j}) - V(y_s)||Z_{\tau_j,t}|) \\
&\le \frac{|R^y_{s,\tau_i}|}{\omega(s,\tau_i)^{2/p}} + \ldots + \frac{|R^y_{\tau_j,t}|}{\omega(\tau_j,t)^{2/p}} + \frac{|V(y_{\tau_i}) - V(y_s)||Z_{\tau_i,\tau_{i+1}}|}{\omega(s,\tau_i)^{1/p}\omega(\tau_i,\tau_{i+1})^{1/p}} + \ldots + \frac{|V(y_{\tau_j}) - V(y_s)||Z_{\tau_j,t}|}{\omega(s,\tau_j)^{1/p}\omega(\tau_j,t)^{1/p}} \\
&\le 2C(N_\alpha(\omega;[0,T]) + 1) + C(N_\alpha(\omega;[0,T]) + 1). \qquad \square
\end{aligned}$$

Lemma A.4. *Let y be a solution to* (A.1)*. Consider*

$$\zeta_t = \zeta_0 + \int_0^t \nu(y_s)(d\mathbf{Z}_s)z_s \in L(\widehat{W}, W)$$

where $\nu\colon W \to L(U, L(W_1, W))$ is bounded, twice differentiable with bounded derivatives and $z\colon [0,T] \to L(\widehat{W}, W_1)$ is controlled by Z. Assume that

$$\|y\|_{p-\omega} \vee \|\mathbf{Z}\|_{p-\omega} \le 1 \quad \textit{and} \quad \|\nu\|_{C_b^2} \le 1.$$

Then there is a constant C and some $\alpha > 0$ depending on p such that

$$\begin{aligned}\|\zeta\|_{p-\omega;[s,t]} \le C\omega(s,t)^{2/p}\Big(&\|z\|_{\infty;[s,t]}(1+N_\alpha(\omega;[0,T])) + \|z\|_{\infty;[s,t]} + \|z'\|_{\infty;[s,t]} \\ &+ \|z\|_{p-\omega;[s,t]} + \|z'\|_{p-\omega;[s,t]} + \|R^z\|_{p/2-\omega;[s,t]}\Big) \\ &+ C\omega(s,t)^{1/p}(\|z\|_{\infty;[s,t]} + \|z'\|_{\infty;[s,t]}) + C\|z\|_{\infty;[s,t]}\end{aligned}$$

and

$$\begin{aligned}\|R^\zeta\|_{p/2-\omega;[s,t]} \le C\omega(s,t)^{1/p}\Big(&\|z\|_{\infty;[s,t]}(1+N_\alpha(\omega;[0,T])) + \|z\|_{\infty;[s,t]} + \|z'\|_{\infty;[s,t]} \\ &+ \|z\|_{p-\omega;[s,t]} + \|z'\|_{p-\omega;[s,t]} + \|R^z\|_{p/2-\omega;[s,t]}\Big) + C(\|z\|_{\infty;[s,t]} + \|z'\|_{\infty;[s,t]})\end{aligned}$$

for all $s \le t$.

Proof. Note first that the path $t \mapsto y_t$ is controlled by Z, and the path $t \mapsto \nu(y_t)z_t \in L(U, L(\widehat{W}, W))$ is controlled by Z as composition with a smooth function [14, Lemma 7.3]. Moreover, its Gubinelli derivative is given by

$$(\nu(y_t)z_t)' = (\nu(y_t))'z_t + \nu(y_t)z_t' = D\nu(y_t)(y_t')z_t + \nu(y_t)z_t' = D\nu(y_t)V(y_t)z_t + \nu(y_t)z_t'$$

where we used (p-variation versions of) [14, Lemma 7.3] in the first and second equality and [14, Theorem 8.4] in the third. We start to prove the claimed estimate for R^ζ. The Gubinelli derivative of ζ is given by $\zeta_t' = \nu(y_t)z_t$ and

$$R^\zeta_{s,t} = \int_s^t \nu(y_u)(d\mathbf{Z}_u)z_u \; - \nu(y_s)(z_s)Z_{s,t}.$$

Hence we can estimate

$$\begin{aligned}|R^\zeta_{s,t}| &\le \left|\int_s^t \nu(y_u)(d\mathbf{Z}_u)z_u - \nu(y_s)(z_s)Z_{s,t} - (\nu(y_s)z_s)'\mathbb{Z}_{s,t}\right| + \left|D\nu(y_s)V(y_s)z_s\mathbb{Z}_{s,t} + \nu(y_s)z_s'\mathbb{Z}_{s,t}\right| \\ &\le C\left(\|R^{\nu(y_\cdot)z_\cdot}\|_{p/2-\omega;[s,t]} + \|D\nu(y_\cdot)V(y_\cdot)z_\cdot\|_{p-\omega;[s,t]} + \|D\nu(y_\cdot)z_\cdot'\|_{p-\omega;[s,t]}\right)\omega(s,t)^{3/p} \\ &\quad + C(\|z\|_{\infty;[s,t]} + \|z'\|_{\infty;[s,t]})\omega(s,t)^{2/p}\end{aligned}$$

where we used Theorem A.2 and that V, ν and all its derivatives are bounded. We have

$$\begin{aligned}|R^{\nu(y_\cdot)z_\cdot}_{s,t}| &= |\nu(y_t)z_t - \nu(y_s)z_s - (\nu(y_s))'z_sZ_{s,t} - \nu(y_s)z_s'Z_{s,t}| \\ &\le |(\nu(y_t) - \nu(y_s) - (\nu(y_s))'Z_{s,t})z_t| + |\nu(y_s)(z_t - z_s - z_s'Z_{s,t})| + |(\nu(y_s))'(z_t - z_s)Z_{s,t}| \\ &\le C(\|R^{\nu(y_\cdot)}\|_{p/2-\omega;[s,t]}\|z\|_\infty + \|R^z\|_{p/2-\omega;[s,t]} + \|z\|_{p-\omega;[s,t]})\omega(s,t)^{2/p}\end{aligned}$$

using $(\nu(y_s))' = D\nu(y_s)V(y_s)$ and boundedness of the vector fields an their derivatives. As in [14, Lemma 7.3], we can see that

$$\begin{aligned}\|R^{\nu(y_\cdot)}\|_{p/2-\omega;[s,t]} &\le C\left(\|y\|^2_{p-\omega;[s,t]} + \|R^y\|_{p/2-\omega;[s,t]}\right) \\ &\le C(1+N_\alpha(\omega;[0,T]))\end{aligned}$$

where the second estimate follows from Lemma A.3. Thus

$$\|R^{\nu(y_\cdot)z_\cdot}\|_{p-\omega;[s,t} \le C(\|z\|_\infty(1+N_\alpha(\omega;[0,T])) + \|R^z\|_{p/2-\omega;[s,t]} + \|z\|_{p-\omega;[s,t]}).$$

Using the Lipschitz bounds for V, ν and its derivatives, we can easily see that

$$\|D\nu(y_\cdot)V(y_\cdot)z_\cdot\|_{p-\omega;[s,t]} \le C(\|z\|_{\infty;[s,t]} + \|z\|_{p-\omega;[s,t]}) \quad \text{and}$$
$$\|D\nu(y_\cdot)z'_\cdot\|_{p-\omega;[s,t]} \le C(\|z'\|_{\infty;[s,t]} + \|z'\|_{p-\omega;[s,t]}).$$

Therefore, we find that

$$\begin{aligned}\|R^\zeta\|_{p/2-\omega;[s,t]} \le C\omega(s,t)^{1/p}&(\|z\|_{\infty;[s,t]}(1+N_\alpha(\omega;[0,T])) + \|z\|_{\infty;[s,t]} + \|z'\|_{\infty;[s,t]}\\ &+ \|z\|_{p-\omega;[s,t]} + \|z'\|_{p-\omega;[s,t]} + \|R^z\|_{p/2-\omega;[s,t]}) + C(\|z\|_{\infty;[s,t]} + \|z'\|_{\infty;[s,t]}).\end{aligned}$$

For ζ, we have

$$\begin{aligned}\frac{|\zeta_{s,t}|}{\omega(s,t)^{1/p}} &\le \omega(s,t)^{1/p}\frac{|R^\zeta_{s,t}|}{\omega(s,t)^{2/p}} + \frac{|\nu(y_s)(z_s)Z_{s,t}|}{\omega(s,t)^{1/p}}\\ &\le \omega(s,t)^{1/p}\|R^\zeta\|_{p/2-\omega;[s,t]} + C\|z\|_{\infty;[s,t]}\end{aligned}$$

for all $s < t$ and the claim follows. □

Lemma A.5. *Let $A\colon [0,T] \to L(U, L(W,W))$ be controlled by Z, and let $\mathbf{Z}$ be weakly geometric. Consider a solution $\Phi\colon [0,T] \to L(W,W)$ to*

$$\Phi_t = \Phi_0 + \int_0^t A(s)(d\mathbf{Z}_s)\Phi_s \in L(W,W).$$

Then Liouville's formula holds:

$$\det(\Phi_t) = \det(\Phi_0)\exp\left(\operatorname{Tr}\int_0^t A(s)\,d\mathbf{Z}_s\right).$$

In particular, if $\det(\Phi_0) \neq 0$, Φ_t is invertible for every $t \ge 0$. In this case, the inverse $\Psi_t := \Phi_t^{-1}$ solves the equation

$$\Psi_t = \Phi_0^{-1} - \int_0^t \Psi_s A(s)(d\mathbf{Z}_s) \in L(W,W).$$

Proof. Assume first that Z is smooth. In this case, Liouville's formula is well-known, cf. [1, (11.4) Proposition]. The statement about Ψ follows from the identity

$$\frac{d\Phi_t^{-1}}{dt} = -\Phi_t^{-1}\frac{d\Phi_t}{dt}\Phi_t^{-1},$$

which is true for matrices depending smoothly on t. The general case follows by approximation of $\mathbf{Z}$ with smooth rough paths and continuity of the rough integral. □

Lemma A.6. *Let y be a solution to* (A.1)*. Consider a solution $\Phi\colon [0,T] \to L(W,W)$ to*

$$\Phi_t = \Phi_0 + \int_0^t \nu(y_u)(d\mathbf{Z}_u)\Phi_u \in L(W,W)$$

where $\nu\colon W \to L(U, L(W, W))$ *is bounded, twice differentiable and has bounded derivatives. Assume that*

$$\|y\|_{p-\omega} \vee \|\mathbf{Z}\|_{p-\omega} \leq 1 \quad \textit{and} \quad \|\nu\|_{C_b^2} \leq 1.$$

Then Φ *is controlled by* Z *and there are constants* C *and* $\alpha > 0$ *depending on* p *such that*

$$\|R^\Phi\|_{p/2-\omega} \leq C(1 + |\Phi_0|)\exp(CN_\alpha(\omega; [0, T])) \quad \textit{and}$$
$$\|\Phi\|_\infty + \|\Phi\|_{p-\omega} + \|\Phi'\|_\infty + \|\Phi'\|_{p-\omega} \leq C(1 + |\Phi_0|)\exp(CN_\alpha(\omega; [0, T])).$$

The same estimate holds true for any solution $\Psi\colon [0, T] \to L(W, W)$ *to*

$$\Psi_t = \Psi_0 - \int_0^t \Psi_u \nu(y_u)(d\mathbf{Z}_u) \in L(W, W)$$

when we replace Φ_0 *by* Ψ_0.

Proof. Note that Φ is controlled by Z with derivative

$$(\Phi_t)' = \nu(y_t)\Phi_t. \tag{A.2}$$

Using boundedness of ν and its derivative and our assumptions on ω, this implies that

$$\|\Phi'\|_{\infty;[s,t]} \leq C\|\Phi\|_{\infty;[s,t]} \quad \text{and} \quad \|\Phi'\|_{p-\omega;[s,t]} \leq C(\|\Phi\|_{p-\omega;[s,t]} + \|\Phi\|_{\infty;[s,t]})$$

for all $s < t$, therefore it is enough to bound Φ to obtain bounds for Φ'. Let K be a constant such that

$$\|\Phi\|_{p-\omega;[0,T]} + \|\Phi\|_{\infty;[0,T]} \leq K.$$

Let $\alpha > 0$ and choose $s < t$ such that $\omega(s, t) \leq \alpha$. Using Lemma A.4, we have for sufficiently small α

$$\|R^\Phi\|_{p/2-\omega;[s,t]} \leq C\alpha^{1/p}\|R^\Phi\|_{p/2-\omega;[s,t]} + C\alpha^{1/p}K(1 + N_\alpha(\omega; [0, T])) + CK.$$

Choosing α smaller if necessary, we may assume that $C\alpha^{1/p} \leq 1/2$ and we therefore obtain

$$\|R^\Phi\|_{p/2-\omega;[s,t]} \leq 2C\alpha^{1/p}K(1 + N_\alpha(\omega; [0, T])) + 2CK \leq CK(1 + N_\alpha(\omega; [0, T])).$$

Using the same strategy as at the end of the proof of Lemma A.3, we can conclude that

$$\|R^\Phi\|_{p/2-\omega;[0,T]} \leq CK(1 + N_\alpha(\omega; [0, T]))^2.$$

Using the results about linear rough differential equations in [15, Section 5], we see that we can choose

$$K = C(1 + |\Phi_0|)\exp(CN_\alpha(\omega; [0, T]))$$

and the claim follows for R^Φ. The estimates for Φ can either be obtained by a direct calculation similar to the one performed in Lemma A.4, but also follow from the results proven for linear rough differential equations in [15, Section 5]. The estimates for Ψ can be obtained in exactly the same way. □

Lemma A.7. *Let y be a solution to* (A.1)*. Let $\zeta\colon [0,T] \to L(\widehat{W}, W)$ be of the form*

$$\zeta_t = \int_0^t \widehat{\nu}(y_u)(d\mathbf{Z}_u)\widehat{z}_u$$

for some $\widehat{\nu}\colon W \to L(U, L(W_1, W))$ and $\widehat{z}\colon [0,T] \to L(\widehat{W}, W_1)$. Assume that $\widehat{z}$ is controlled by Z. Consider

$$z_t = \Phi_t\left(z_0 + \int_0^t \Psi_u \, d\zeta_u\right), \quad z_0 \in L(\widehat{W}, W)$$

with Φ, Ψ as in Lemma A.6 where we assume in addition that $\Phi_0 = \Psi_0 = \mathrm{Id}$. Assume that

$$\|y\|_{p-\omega} \vee \|\mathbf{Z}\|_{p-\omega} \le 1 \quad \textit{and} \quad |\widehat{\nu}|_{C_b^2} \le 1.$$

Let $\kappa \ge 1$ be a constant such that

$$|\widehat{z}|_\infty + |\widehat{z}|_{p-\omega} + |\widehat{z}'|_\infty + |\widehat{z}'|_{p-\omega} + |R^{\widehat{z}}|_{p/2-\omega} \le \kappa.$$

Then z is controlled by Z and there are constants $C > 0$ and $\alpha > 0$ depending only on p such that

$$\begin{aligned}&\|z\|_\infty + \|z\|_{p-\omega} + \|z'\|_\infty + \|z'\|_{p-\omega} + \|R^z\|_{p/2-\omega} \\ &\quad \le C\kappa(1 + |z_0|)(1 + \omega(0,T)^{1/p} + \omega(0,T)^{4/p}) \exp(CN_\alpha(\omega; [0,T])).\end{aligned}$$

Proof. It is clear that z is controlled by Z, and the Gubinelli derivative is given by

$$\begin{aligned} z_t' &= \Phi_t'\left(z_0 + \int_0^t \Psi_u \, d\zeta_u\right) + \Phi_t(\Psi_t \zeta_t') = \nu(y_t)\Phi_t\left(z_0 + \int_0^t \Psi_u \, d\zeta_u\right) + \widehat{\nu}(y_t)\widehat{z}_t \\ &= \nu(y_t) z_t + \widehat{\nu}(y_t)\widehat{z}_t \end{aligned} \tag{A.3}$$

where we used Theorem A.2, $\zeta_t' = \widehat{\nu}(y_t)\widehat{z}_t$ and the fact that $\Psi_t = (\Phi_t)^{-1}$. For $s < t$, we therefore obtain

$$\begin{aligned} R^z_{s,t} &= z_{s,t} - z_s' Z_{s,t} \\ &= \Phi_t \int_s^t \Psi_u \, d\zeta_u + \Phi_{s,t} z_s - \nu(y_s)(Z_{s,t}) z_s - \widehat{\nu}(y_s)(Z_{s,t})\widehat{z}_s \\ &= \Phi_t\left(\int_s^t \Psi_u \, d\zeta_u - \Psi_s \zeta_{s,t} + \Psi_s R^\zeta_{s,t}\right) + \Phi_t \Psi_s \zeta_s' Z_{s,t} + \Phi_{s,t} z_s - \Phi_s' Z_{s,t} z_s - \widehat{\nu}(y_s)(Z_{s,t})\widehat{z}_s \\ &= \Phi_t\left(\int_s^t \Psi_u \, d\zeta_u - \Psi_s \zeta_{s,t} + \Psi_s R^\zeta_{s,t}\right) + R^\Phi_{s,t} z_s + (\Phi_t - \Phi_s)\Psi_s \widehat{\nu}(y_s)(Z_{s,t})\widehat{z}_s \end{aligned}$$

and by the triangle inequality,

$$\begin{aligned} |R^z_{s,t}| \le &\left|\Phi_t\left(\int_s^t \Psi_u \, d\zeta_u - \Psi_s \zeta_{s,t} + \Psi_s R^\zeta_{s,t}\right)\right| + \left|R^\Phi_{s,t}\left(z_0 + \int_0^s \Psi_u \, d\zeta_u\right)\right| \\ &+ |(\Phi_t - \Phi_s)\Psi_s \widehat{\nu}(y_s)(Z_{s,t})\widehat{z}_s|. \end{aligned} \tag{A.4}$$

For the first term on the right hand side in (A.4), we can use the estimate in Theorem A.2 to see that

$$\begin{aligned}&\left|\Phi_t\left(\int_s^t \Psi_u\, d\zeta_u - \Psi_s\zeta_{s,t} + \Psi_s R^{\zeta}_{s,t}\right)\right| \\ &\le C\|\Phi\|_{\infty;[s,t]}\omega(s,t)^{2/p}\Big((\|R^{\zeta}\|_{p/2-\omega;[s,t]} + \|\Psi'\zeta'\|_{p-\omega;[s,t]})\omega(s,t)^{1/p} + \|\Psi'\|_{\infty;[s,t]}\|\zeta'\|_{\infty;[s,t]} \\ &\qquad + \|\Psi\|_{\infty;[s,t]}\|R^{\zeta}\|_{p/2-\omega;[s,t]}\Big).\end{aligned}$$

Note first that

$$\|\Psi'\zeta'\|_{p-\omega;[s,t]} \le \|\Psi'\|_{p-\omega;[s,t]}\|\zeta'\|_{\infty;[s,t]} + \|\Psi'\|_{\infty;[s,t]}\|\zeta'\|_{p-\omega;[s,t]}$$

and

$$\|\zeta'\|_{\infty} + \|\zeta'\|_{p-\omega} \le C\kappa.$$

From Lemma A.4, it follows that

$$\|R^{\zeta}\|_{p/2-\omega;[s,t]} \le C\kappa\left(\omega(s,t)^{1/p}(1 + N_\alpha(\omega;[0,T])) + 1\right).$$

Using the estimates for Φ and Ψ in Lemma A.6, we therefore obtain

$$\left|\Phi_t\left(\int_s^t \Psi_u\, d\zeta_u - \Psi_s\zeta_{s,t} + \Psi_s R^{\zeta}_{s,t}\right)\right| \le C\kappa\omega(s,t)^{2/p}(1 + \omega(0,T)^{1/p} + \omega(0,T)^{2/p})\exp(CN_\alpha(\omega;[0,T]))$$

for all $s < t$. For the second summand in (A.4), we can again use Theorem A.2 to estimate

$$\begin{aligned}\left|R^{\Phi}_{s,t}\left(z_0 + \int_0^s \Psi_u\, d\zeta_u\right)\right| &\le C\omega(s,t)^{2/p}\|R^{\Phi}\|_{p/2-\omega;[s,t]} \\ &\quad\times\Big(|z_0| + \omega(0,s)^{3/p}(\|R^{\zeta}\|_{p/2-\omega;[0,s]} + \|\Psi'\zeta'\|_{p-\omega;[0,s]}) \\ &\quad + \|\zeta\|_{\infty;[0,s]} + \omega(0,s)^{2/p}|\widehat{z_0}|\Big)\end{aligned}$$

using $\Psi_0 = \mathrm{Id}$, $\Psi_0' = \nu(y_0)$ and $\zeta_0' = \widehat{\nu}(y_0)\widehat{z_0}$. With Lemma A.4, we obtain

$$\|\zeta\|_{\infty;[0,s]} \le \omega(0,T)^{1/p}\|\zeta\|_{p-\omega} \le C\kappa(\omega(0,T)^{1/p} + \omega(0,T)^{3/p})(1 + N_\alpha(\omega;[0,T])).$$

As above, we obtain the bound

$$\left|R^{\Phi}_{s,t}\left(z_0 + \int_0^s \Psi_u\, d\zeta_u\right)\right| \le C\kappa\omega(s,t)^{2/p}(1 + |z_0|)(1 + \omega(0,T)^{1/p} + \omega(0,T)^{4/p})\exp(CN_\alpha(\omega;[0,T]))$$

for all $s < t$. For the third term in (A.4), we have

$$\begin{aligned}|(\Phi_t - \Phi_s)\Psi_s\widehat{\nu}(y_s)(Z_{s,t})\widehat{z_s}| &\le \kappa\omega(s,t)^{2/p}\|\Phi\|_{p-\omega;[s,t]}\|\Psi\|_{\infty;[0,s]} \\ &\le C\kappa\omega(s,t)^{2/p}\exp(CN_\alpha(\omega;[0,T])).\end{aligned}$$

Using all these estimates in (A.4), we can conclude that

$$\|R^{z}\|_{p/2-\omega} \le C\kappa(1 + |z_0|)(1 + \omega(0,T)^{1/p} + \omega(0,T)^{4/p})\exp(CN_\alpha(\omega;[0,T])).$$

We proceed with z. For $s < t$,

$$|z_t - z_s| \le |\Phi_t| \left| \int_s^t \Psi_u \, d\zeta_u \right| + |\Phi_t - \Phi_s| \left| z_0 + \int_0^s \Psi_u \, d\zeta_u \right|$$

and as before, we obtain the estimate

$$\|z\|_{p-\omega} \le C\kappa(1 + |z_0|)(1 + \omega(0,T)^{1/p} + \omega(0,T)^{4/p}) \exp(CN_\alpha(\omega; [0,T])).$$

Similarly,

$$\|z\|_{\infty} \le C\kappa(1 + |z_0|)(1 + \omega(0,T)^{1/p} + \omega(0,T)^{4/p}) \exp(CN_\alpha(\omega; [0,T])).$$

From (A.3), we see that

$$\|z'\|_{\infty;[s,t]} \le C(\|z\|_{\infty;[s,t]} + \|\widehat{z}\|_{\infty;[s,t]})$$

and

$$\|z'\|_{p-\omega;[s,t]} \le C(\|z\|_{\infty;[s,t]} + \|z\|_{p-\omega;[s,t]} + \|\widehat{z}\|_{\infty;[s,t]} + \|\widehat{z}\|_{p-\omega;[s,t]})$$

for all $s < t$, therefore the same estimates hold for z'. This proves the claim. □

Lemma A.8. *Let $z \in \mathscr{D}_Z^p([0,T], L(W, W_1))$ and $\widehat{z} \in \mathscr{D}_Z^p([0,T], L(\widehat{W}, \widehat{W}_1))$. Then $z \otimes \widehat{z} \in \mathscr{D}_Z^p([0,T], L(W \otimes \widehat{W}, W_1 \otimes \widehat{W}_1))$ with derivative*

$$(z \otimes \widehat{z})'_s(u) = (z'_s u) \otimes \widehat{z}_s + z_s \otimes (\widehat{z}'_s u), \quad u \in U$$

and remainder given by

$$R^{z\otimes\widehat{z}}_{s,t} = z_{s,t} \otimes \widehat{z}_{s,t} + R^z_{s,t} \otimes \widehat{z}_s + z_s \otimes R^{\widehat{z}}_{s,t}.$$

Proof. Follows readily from a short calculation. □

Proof of Theorem 3.1. W.l.o.g, we may assume $s = 0$, otherwise we may replace $\mathbf{Z}$ by the time-shifted rough path $\mathbf{Z}_{s+\cdot}$ and solve the corresponding equation. Existence of the derivatives and their characterization as solutions to rough differential equations is a classical result, cf. [14, Section 8.9] and [17, Section 11.2]. It remains to prove the claimed bounds for $X^{(k)}$. Let us first assume that $\|V\|_{\mathcal{C}_b^{2+k}} \le 1/\kappa$ for some $\kappa \ge 1$ and that $\widehat{\omega}$ is some control function for which

$$\|\mathbf{Z}\|_{p-\widehat{\omega}} \vee \|X^x\|_{p-\widehat{\omega}} \le 1$$

holds (the precise choice of κ and $\widehat{\omega}$ will be made later). We claim that in this case, there are constants C, α and M depending on p and k such that

$$\begin{aligned} &\|X^{(k)}\|_\infty + \|X^{(k)}\|_{p-\widehat{\omega}} + \|(X^{(k)})'\|_\infty + \|(X^{(k)})'\|_{p-\widehat{\omega}} + \|R^{X^{(k)}}\|_{p/2-\widehat{\omega}} \\ &\quad \le C(1 + \widehat{\omega}(0,T)^{1/p} + \widehat{\omega}(0,T)^{M/p}) \exp(CN_\alpha(\widehat{\omega}; [0,T])) \end{aligned} \tag{A.5}$$

holds. We prove the claim by induction. For $k = 1$, $X_t^{(1)} =: \Phi_t$ solves

$$X_t^{(1)} = \mathrm{Id} + \int_0^t DV(X_s^x)(d\mathbf{Z}_s)X_s^{(1)}$$

and the bound (A.5) follows from Lemma (A.6). Let $k \geq 2$ and assume that our claim holds for all $l = 1, \dots, k-1$. It is easy to see that $X^{(k)}$ solves an inhomogeneous equation of the form

$$X_t^{(k)} = \zeta_t + \int_0^t DV(X_s^x)X_s^{(k)}\, d\mathbf{Z}_s \in L((\mathbb{R}^n)^{\otimes k}, \mathbb{R}^n) \tag{A.6}$$

where $\zeta\colon [0,T] \to L((\mathbb{R}^n)^{\otimes k}, \mathbb{R}^n))$ can be written as

$$\zeta_t = \sum_{\substack{2\leq l\leq k \\ i_1+\dots+i_l=k}} \lambda_{i_1,\dots,i_l} \int_0^t D^l V(X_s^x)(d\mathbf{Z}_s)(X_s^{(i_1)} \otimes \cdots \otimes X_s^{(i_l)}) =: \int_0^t \widehat{\nu}(X_s^x)(d\mathbf{Z}_s)\widehat{z}_s^k,$$

the $\lambda_{i_1,\dots,i_l}$ being integers which can be explicitly calculated using the Leibniz rule. Note that $\widehat{z}^k$ is controlled by Z by the induction hypothesis and Lemma A.8, therefore the integrals are well defined. Moreover, the estimate (A.5) holds for $\widehat{z}^k$ instead of $X^{(k)}$ again by the induction hypothesis and Lemma A.8. We also see that we can choose $\kappa \geq 1$ depending only on k to obtain $\|\widehat{\nu}\|_{C_b^2} \leq 1$. The equation (A.6) can be solved with the variation of constants method: making the ansatz $X_t^{(k)} = \Phi_t C_t$, $C_t \in L((\mathbb{R}^n)^{\otimes k}, \mathbb{R}^n))$, we can conclude that $X^{(k)}$ can be written as

$$X_t^{(k)} = \Phi_t \int_0^t \Phi_s^{-1} d\zeta_s.$$

The claim (A.5) for $X^{(k)}$ now follows from Lemma A.7. We proceed with deducing the bound (3.4) from (A.5). Note first that X^x also solves the equation

$$X_t^x = x + \int_0^t \tilde{V}(X_s^x)\, d\tilde{\mathbf{Z}}_s$$

where $\tilde{V} = V/(\kappa\|V\|_{C_b^{2+k}})$ and $\tilde{\mathbf{Z}} = (\kappa\|V\|_{C_b^{2+k}}Z, \kappa^2\|V\|_{C_b^{2+k}}^2\mathbb{Z})$. Clearly $\|\tilde{V}\|_{C_b^{2+k}} \leq 1/\kappa$, and

$$\widehat{\omega}(s,t) := \|\tilde{\mathbf{Z}}\|_{p-\mathrm{var};[s,t]}^p + \|X^x\|_{p-\mathrm{var};[s,t]}^p = \kappa^p\|V\|_{C_b^{2+k}}^p\|\mathbf{Z}\|_{p-\mathrm{var};[s,t]}^p + \|X^x\|_{p-\mathrm{var};[s,t]}^p$$

is a valid choice for $\widehat{\omega}$. Therefore, (A.5) holds for $X^{(k)}$ with this $\widehat{\omega}$. Next, [15, Corollary 3] implies that there is a constant depending on p such that

$$N_1(X^x;[0,T]) \leq C(N_1(\tilde{\mathbf{Z}};[0,T]) + 1).$$

Together with [15, Lemma 4], this implies that

$$\begin{aligned}\widehat{\omega}(0,T)^{1/p} &\leq \|\tilde{\mathbf{Z}}\|_{p-\mathrm{var};[0,T]} + \|X^x\|_{p-\mathrm{var};[0,T]} \leq N_1(\tilde{\mathbf{Z}};[0,T]) + N_1(X^x;[0,T]) + 2\\ &\leq C(N_1(\tilde{\mathbf{Z}};[0,T]) + 1) \leq C\exp(N_1(\tilde{\mathbf{Z}};[0,T])),\end{aligned}$$

therefore also

$$\widehat{\omega}(0,T)^{1/p} + \widehat{\omega}(0,T)^{M/p} \leq C\exp(CN_1(\tilde{\mathbf{Z}};[0,T])).$$

From [15, Lemma 3] and [5, Lemma 5], we see that

$$N_\alpha(\widehat{\omega};[0,T]) \le 2N_\alpha(\tilde{\mathbf{Z}};[0,T]) + 2N_\alpha(X^x;[0,T]) + 2 \le C(N_1(\tilde{\mathbf{Z}};[0,T]) + 1)$$

for a constant C depending on α and p, and therefore on p only. Using these estimates, (A.5) implies that there is a constant C depending on p and k such that

$$\|X^{(k)}\|_{p-\widehat{\omega}} \le C \exp\Big(CN_1(\tilde{\mathbf{Z}};[0,T])\Big)$$

holds. Using [15, Lemma 1 and Lemma 3], we see that

$$N_1(\tilde{\mathbf{Z}};[0,T]) \le \kappa^p \|V\|^p_{C_b^{2+k}}(2N_1(\mathbf{Z};[0,T]) + 1)$$

which shows that

$$\|X^{(k)}\|_{p-\widehat{\omega}} \le C \exp\Big(C\|V\|^p_{C_b^{2+k}}(N_1(\mathbf{Z};[0,T]) + 1)\Big).$$

Note that for every $s \le t$, the estimate for X^x in [17, Theorem 10.14] implies that

$$\begin{aligned}\widehat{\omega}(s,t) &= \|\tilde{\mathbf{Z}}\|^p_{p-\mathrm{var};[s,t]} + \|X^x\|^p_{p-\mathrm{var};[s,t]} \le \|\tilde{\mathbf{Z}}\|^p_{p-\mathrm{var};[s,t]} + C\|\tilde{\mathbf{Z}}\|^p_{p-\mathrm{var};[s,t]}(1 + \|\tilde{\mathbf{Z}}\|^p_{p-\mathrm{var};[0,T]}) \\ &\le C\kappa^p\|V\|^p_{C_b^{2+k}}\Big(1 + \kappa^p\|V\|^p_{C_b^{2+k}}\omega(0,T)\Big)\omega(s,t),\end{aligned}$$

therefore

$$\|X^{(k)}\|_{p-\omega} \le C\|V\|_{C_b^{2+k}}\Big(1 + \|V\|_{C_b^{2+k}}\omega(0,T)^{1/p}\Big)\|X^{(k)}\|_{p-\widehat{\omega}}$$

and we can use again the estimates above to conclude (3.4). The estimate (3.5) just follows from

$$\|X^{(k)}\|_\infty \le \|X^{(k)}\|_{p-\omega}\|\mathbf{Z}\|_{p-\mathrm{var}} + |X_0^{(k)}|. \qquad \square$$

References

[1] Herbert Amann. *Ordinary differential equations: an introduction to nonlinear analysis*, volume 13. Walter de Gruyter, 1990.

[2] Felix Anker, Christian Bayer, Martin Eigel, Marcel Ladkau, Johannes Neumann, and John Schoenmakers. SDE based regression for linear random PDEs. *SIAM Journal on Scientific Computing*, 39(3):A1168–A1200, 2017.

[3] Fabrice Baudoin. *An introduction to the geometry of stochastic flows*. World Scientific, 2004.

[4] Manuel Baumann, Peter Benner, and Jan Heiland. Space-time Galerkin POD with application in optimal control of semi-linear parabolic partial differential equations. *arXiv preprint arXiv:1611.04050*, 2016.

[5] Christian Bayer, Peter K. Friz, Sebastian Riedel, and John Schoenmakers. From rough path estimates to multilevel Monte Carlo. *SIAM Journal on Numerical Analysis*, 54(3):1449–1483, 2016.

[6] Christian Bayer, Ulrich Horst, and Jinniao Qiu. A functional limit theorem for limit order books with state dependent price dynamics. *Ann. Appl. Probab.*, 27(5):2753–2806, 10 2017.

[7] Michael Caruana, Peter K Friz, and Harald Oberhauser. A (rough) pathwise approach to a class of non-linear stochastic partial differential equations. *Annales de l'Institut Henri Poincare (C) Non Linear Analysis*, 28(1):27–46, 2011.

[8] Thomas Cass, Christian Litterer, and Terry Lyons. Integrability and tail estimates for Gaussian rough differential equations. *Ann. Probab.*, 41(4):3026–3050, 2013.

[9] Aurélien Deya, Massimiliano Gubinelli, Martina Hofmanova, and Samy Tindel. A priori estimates for rough PDEs with application to rough conservation laws. *arXiv preprint arXiv:1604.00437*, 2016.

[10] Aurélien Deya, Massimiliano Gubinelli, and Samy Tindel. Non-linear rough heat equations. *Probability Theory and Related Fields*, 153(1-2):97–147, 2012.

[11] Aurélien Deya, Andreas Neuenkirch, and Samy Tindel. A Milstein-type scheme without Lévy area terms for SDEs driven by fractional Brownian motion. *Ann. Inst. Henri Poincaré Probab. Stat.*, 48(2):518–550, 2012.

[12] Joscha Diehl, Peter Friz, and Wilhelm Stannat. Stochastic partial differential equations: a rough paths view on weak solutions via Feynman-Kac. *Annales de la faculté des sciences de Toulouse Sér. 6*, 26(4):911–947, 2017.

[13] Joscha Diehl, Harald Oberhauser, and Sebastian Riedel. A Lévy area between Brownian motion and rough paths with applications to robust nonlinear filtering and rough partial differential equations. *Stochastic processes and their applications*, 125:161–181, 2015.

[14] Peter Friz and Martin Hairer. *A course on rough paths*. Springer, 2014.

[15] Peter Friz and Sebastian Riedel. Integrability of (non-)linear rough differential equations and integrals. *Stochastic Analysis and Applications*, 31(2):336–358, 2013.

[16] Peter K Friz, Paul Gassiat, Pierre-Louis Lions, and Panagiotis E Souganidis. Eikonal equations and pathwise solutions to fully non-linear SPDEs. *Stochastics and Partial Differential Equations: Analysis and Computations*, 5(2):256–277, 2017.

[17] Peter K. Friz and Nicolas B. Victoir. *Multidimensional stochastic processes as rough paths*, volume 120 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010. Theory and applications.

[18] Peter K. Friz and Huilin Zhang. Differential equations driven by rough paths with jumps. arXiv, 2017.

[19] László Györfi, Michael Kohler, Adam Krzyżak, and Harro Walk. *A distribution-free theory of nonparametric regression*. Springer Series in Statistics. Springer-Verlag, New York, 2002.

[20] Martin Hairer. Solving the KPZ equation. *Annals of mathematics*, 178(2):559–664, 2013.

[21] Peter Kotelenez. *Stochastic ordinary and stochastic partial differential equations: transition from microscopic to macroscopic equations*, volume 58. Springer Science & Business Media, 2007.

[22] Raphael Kruse. *Strong and Weak Approximation of Semilinear Stochastic Evolution Equations*, volume 2093 of *Lecture Notes in Mathematics*. Springer, 2014.

[23] Terry Lyons. Rough paths, signatures and the modelling of functions on streams. *arXiv preprint arXiv:1405.4537*, 2014.

[24] Terry J Lyons. Differential equations driven by rough signals. *Revista Matemática Iberoamericana*, 14(2):215–310, 1998.

[25] G. N. Milstein and M. V. Tretyakov. *Stochastic numerics for mathematical physics*. Scientific Computation. Springer-Verlag, Berlin, 2004.

[26] GN Milstein and MV Tretyakov. Solving the Dirichlet problem for Navier–Stokes equations by probabilistic approach. *BIT Numerical Mathematics*, 52(1):141–153, 2012.

[27] Étienne Pardoux. Filtrage non linéaire et équations aux dérivées partielles stochastiques associées. In *Ecole d'Eté de Probabilités de Saint-Flour XIX 1989*, pages 68–163. Springer, 1991.

[28] Robert S Strichartz. The Campbell-Baker-Hausdorff-Dynkin formula and solutions of differential equations. *Journal of Functional Analysis*, 72(2):320–345, 1987.