BIT Numerical Mathematics (2025) 65:6
https://doi.org/10.1007/s10543-024-01046-5

BIT

RESEARCH PAPER

Exact dimension reduction for rough differential equations

Martin Redmann[1] · **Sebastian Riedel**[2]

Received: 5 July 2023 / Accepted: 2 December 2024

Abstract

In this paper, practically computable low-order approximations of potentially high-dimensional differential equations driven by geometric rough paths are proposed and investigated. In particular, equations are studied that cover the linear setting, but we allow for a certain type of dissipative nonlinearity in the drift as well. In a first step, a linear subspace is found that contains the solution space of the underlying rough differential equation (RDE). This subspace is associated to covariances of linear Ito-stochastic differential equations which is shown exploiting a Gronwall lemma for matrix differential equations. Orthogonal projections onto the identified subspace lead to a first exact reduced order system. Secondly, a linear map of the RDE solution (quantity of interest) is analyzed in terms of redundant information meaning that state variables are found that do not contribute to the quantity of interest. Once more, a link to Ito-stochastic differential equations is used. Removing such unnecessary information from the RDE provides a further dimension reduction without causing an error. The resulting reduced order rough equation can be solved numerically much faster than the original system. Therefore, our approach provides enormous savings in computing time and is hence beneficial from the practical point of view. Finally, we discretize a linear parabolic rough partial differential equation and a rough wave equation in space. The resulting large-order RDEs are subsequently tackled with the exact reduction techniques studied in this paper. We illustrate the enormous complexity reduction potential in the corresponding numerical experiments.

Keywords Rough differential equations · Model order reduction · Galerkin projections · Non-Markovian processes

Mathematics Subject Classification 60G33 · 60H10 · 60L20 · 60L50 · 65C30 · 93A15

✉ Sebastian Riedel
sebastian.riedel@fernuni-hagen.de

Martin Redmann
martin.redmann@mathematik.uni-halle.de

[1] Institute of Mathematics, Martin Luther University Halle-Wittenberg, Theodor-Lieser-Str. 5, 06120 Halle (Saale), Germany

[2] Faculty of Mathematics and Computer Science, Chair of Applied Stochastics, Fern Universität in Hagen, 58084 Hagen, Germany

Published online: 08 January 2025

Introduction

Rough paths theory is a powerful tool in stochastic analysis that allows to study stochastic ordinary differential equations pathwise. Invented by Lyons in the 90s [25], the theory found applications in a variety of fields, cf. [13] for an overview. As already conjectured in Lyons' seminal work [25], the theory has a vast potential to study stochastic *partial* differential equations (SPDEs), too. Nowadays, there exist numerous approaches to these *rough partial differential equations* (RPDEs). Parabolic equations with roughness in time were studied, e.g., via semigroup theory [15, 16], with (stochastic) viscosity theory [6, 7, 12], and with a Feynman-Kac approach [11]. Note that this is by far not an exhaustive review of the existing literature, the interesting reader may consult [13, Chapter 12] for a more extensive overview of approaches to rough-in-time RPDEs. Roughness in space of parabolic SPDEs, e.g., in the presence of space-time white noise, was also investigated with rough paths theory [17, 21]. This line of thinking culminated in Hairer's solution to the KPZ-equation [18] and his seminal theory of regularity structures [19]. We are not trying to summarize the vast literature built on regularity structures here and refer, once again, to [13] for a (non-exhaustive) overview. However, when it comes to actually solve rough SPDEs numerically, much less work can be found (let us, however, mention [2, 9, 20] here).

A standard approach to solve a deterministic (time and space dependent) PDE is to discretize in space and hence to approximate the solution by a high-dimensional system of ordinary differential equations (ODEs). For a RPDE, this strategy results in a system of rough ODEs. Solving these equations numerically is a notoriously difficult problem due to the high dimension of the system, especially if many system evaluations are required. Such computationally challenging situations occur for instance in an optimal control context or if a Monte-Carlo method is used. One common approach in PDE and SPDE theory to escape the curse of dimensionality is to use *model order reduction* (MOR). We refer to [1, 3] for a comprehensive overview on various projection-based MOR techniques for deterministic equations and to [4, 28] for a system-theoretic ansatz to tackle high-dimensional stochastic ODEs. The basic observation is that many equations contain redundancies that lead to the fact that the solution described by the system essentially evolves in a subspace (or manifold) of much lower dimension. MOR aims to identify these subspaces (or manifolds) on which the dynamics of the equations are essentially acting. Subsequently, one transforms the initial high-dimensional (stochastic) ODE to a (stochastic) ODE of lower order that describes the evolution in this smaller space (or manifold). For many equations, MOR can lead to a drastic dimension reduction while keeping a high accuracy. In fact, MOR is nowadays a standard procedure and widely used in practice.

The contribution of this work is to make an important first step towards establishing MOR in the context of rough differential equations (RDEs). More precisely, we will study the exact dimension reduction for a linear RDE driven by a geometric rough path $\mathbf{W}$, i.e., an equation of the form

$$dx(t) = Ax(t)\,dt + N(x(t))\,K^{\frac{1}{2}}\,d\mathbf{W}(t), \quad x(0) = x_0 \in \mathbb{R}^n$$

with state space dimension n being large. In fact, we can even allow for a nonlinear drift term, cf. Sect. 1. Our first main result is Theorem 2.5 that identifies an operator P on $\mathbb{R}^n$ having the property that every $x(t)$ lies in the image of P. Interestingly, P is explicit and given by

$$P = \int_0^\infty \mathbb{E}[x_B(t)x_B(t)^\top]\,dt,$$

where x_B solves the corresponding Ito stochastic differential equation

$$dx_B(t) = Ax_B(t)\,dt + N(x_B(t))\,K^{\frac{1}{2}}\,dB(t), \quad x_B(0) = x_0 \in \mathbb{R}^n.$$

This means that we found an approximation for the solution space of the RDE that is independent of the driver $\mathbf{W}$. In addition, this approximation can practically be computed in high dimensions by solving a matrix equation. Now, given that the image of P is low-dimensional, we can construct an exact reduced system that is much cheaper to evaluate for each $\mathbf{W}$. Therefore, we significantly save computational time when solving the reduced RDE for many different drivers.

To prove this main theorem, we first approximate $\mathbf{W}$ by smooth rough paths $\mathbf{W}^\epsilon$ and study the corresponding smooth equations. One key ingredient to make the comparison is a Gronwall-type lemma for matrix differentials, cf. Lemma 2.4. Once the statement of Theorem 2.5 is proved for the smooth rough paths $\mathbf{W}^\epsilon$, one can safely pass to the limit using the continuity property of RDEs. The eigenvalue decomposition of P now leads to a dimension reduced equation by using a standard procedure, cf. the discussion after Theorem 2.5. If the quantity of interest is given by $y(t) = Cx(t)$ for a matrix C, we can potentially reduce the dimension even further, cf. Theorem 3.2. In Sect. 4, we apply both theorems and perform MOR for two discretized linear RPDEs. For a rough heat equation and the quantity of interest being the average temperature on the domain, we can reduce the dimension of the discretized equation from $n = 100$ to $r = 33$ with practically no reduction error. In fact, even a reduction to $r = 5$ yields an error below one percent. Secondly, we use our approach for finding a low-order approximation of a finite difference discretization corresponding to a formal rough wave equation. In this spatially discrete 1000-dimensional model, more than half of the variables could be removed without an error. Furthermore, keeping solely one percent of the states causes an error of approximately one percent. This underlines the enormous potential of MOR for RPDEs.

Notation and basic definitions

Continuous functions $W\colon [0,T] \to \mathbb{R}^d$ will be called *paths*. Let $\alpha \in (0,1]$. If the α-Hölder seminorm

$$\sup_{s<t} \frac{\|W(t) - W(s)\|_2}{|t-s|^\alpha}$$

is finite, we say $W \in C^\alpha$. Here and throughout the rest of the paper, $\|\cdot\|_2$ denotes the Euclidean norm. In the following, we recall some basic definitions of rough paths theory. For a more comprehensive overview, we refer the reader to [13, 14, 26]. If $W\colon [0,T] \to \mathbb{R}^d$ is sufficiently smooth, we can define the n-times iterated integrals

$$\mathbb{W}_{s,t}^{(n)} := \int_{s\le t_1<\ldots<t_n\le t} dW(t_1)\otimes\cdots\otimes dW(t_n) \in (\mathbb{R}^d)^{\otimes n} = \bigotimes_{k=1}^{n}\mathbb{R}^d.$$

Note that $\mathbb{W}_{s,t}^{(1)} = W(t) - W(s)$. For some fixed N, we call $\mathbf{W} = (\mathbf{W}_{s,t})_{0\le s<t\le T}$ given by

$$\mathbf{W}_{s,t} = (1, \mathbb{W}_{s,t}^{(1)}, \ldots, \mathbb{W}_{s,t}^{(N)}) \in \bigoplus_{n=0}^{N}(\mathbb{R}^d)^{\otimes n}$$

with $(\mathbb{R}^d)^{\otimes 0} = \mathbb{R}$ the *canonical lift* of W. The space $\bigoplus_{n=0}^{N}(\mathbb{R}^d)^{\otimes n}$ is called *truncated tensor algebra of level* N. Let

$$\mathbf{W}_{s,t} = (1, \mathbb{W}_{s,t}^{(1)}, \ldots, \mathbb{W}_{s,t}^{(N)}), \quad \tilde{\mathbf{W}}_{s,t} = (1, \tilde{\mathbb{W}}_{s,t}^{(1)}, \ldots, \tilde{\mathbb{W}}_{s,t}^{(N)})$$

be two two-parameter functions with values in $\bigoplus_{n=0}^{N}(\mathbb{R}^d)^{\otimes n}$. Then, we set

$$\varrho_\alpha(\mathbf{W}, \tilde{\mathbf{W}}) := \sum_{n=1}^{N}\sup_{s<t}\frac{\|\mathbb{W}_{s,t}^{(n)} - \tilde{\mathbb{W}}_{s,t}^{(n)}\|}{|t-s|^{n\alpha}}.$$

Let $W \in C^\alpha$ and $N \le \frac{1}{\alpha} < N+1$. A two-parameter function

$$\mathbf{W}_{s,t} = (1, \mathbb{W}_{s,t}^{(1)}, \ldots, \mathbb{W}_{s,t}^{(N)})$$

with $\mathbb{W}_{s,t}^{(1)} = W(t) - W(s)$ is called a *geometric α-Hölder rough path associated to* W if there exists a sequence of smooth paths W^ϵ for which the canonical lifts $\mathbf{W}^\epsilon$ satisfy

$$\varrho_\alpha(\mathbf{W}, \mathbf{W}^\epsilon) \to 0$$

as $\epsilon \to 0$. It can be shown [13, 14] that the set of all geometric rough paths constitutes a complete separable metric space with the metric ϱ_α. An α-Hölder path $x\colon [0,T] \to \mathbb{R}^n$ is called a *solution to the rough differential equation*

$$dx(t) = b(x(t))\,dt + \sigma(x(t))\,d\mathbf{W}(t), \quad x(0) = x_0 \tag{0.1}$$

if $x(0) = x_0$ and for any approximating sequence $\mathbf{W}^\epsilon$ to $\mathbf{W}$, the solutions x^ϵ to

$$dx^\epsilon(t) = b(x^\epsilon(t))\,dt + \sigma(x^\epsilon(t))\,dW^\epsilon(t), \quad x^\epsilon(0) = x_0$$

converge in α-Hölder metric to x. Conditions on b and σ under which (0.1) has a unique global-in-time solution can be found in [14, Chapter 10]. In particular, it is shown in [14, Section 10.7] that linear equations have unique solutions globally in time.

1 Setting

Let $\mathbf{W}$ be a geometric rough path associated to a path $W \in C^\alpha$ that takes values in $\mathbb{R}^d$. By definition, there exists a sequence of smooth paths W^ϵ such that their canonical lifts $\mathbf{W}^\epsilon$ satisfy $\mathbf{W}^\epsilon \to \mathbf{W}$ ($\epsilon \to 0$) w.r.t. the rough path metric. In this paper, we will only assume that there exist left-continuous functions $\dot{W}^\epsilon \in L^2_T$, i.e., $\|\dot{W}^\epsilon\|^2_{L^2_T} := \int_0^T \|\dot{W}^\epsilon(v)\|_2^2 dv < \infty$, so that

$$W^\epsilon(t) = W^\epsilon(0) + \int_0^t \dot{W}^\epsilon(s)\, ds \tag{1.1}$$

for all $\epsilon > 0$. We consider the following rough differential equation

$$dx(t) = [Ax(t) + f(x(t))]\, dt + N(x(t))\, K^{\frac{1}{2}}\, d\mathbf{W}(t), \quad x(0) = x_0, \tag{1.2a}$$

$$y(t) = Cx(t), \quad t \in [0, T], \tag{1.2b}$$

with $A \in \mathbb{R}^{n\times n}$, $C \in \mathbb{R}^{p\times n}$, $N : \mathbb{R}^n \to \mathbb{R}^{n\times d}$ is a linear mapping defined by $N(x) = \begin{bmatrix} N_1 x & \dots & N_d x \end{bmatrix}$ for $x \in \mathbb{R}^n$ and given that $N_1, \dots, N_d \in \mathbb{R}^{n\times n}$. Moreover, we interpret the symmetric positive semidefinite matrix $K = (k_{ij})_{i,j=1,\dots,d}$ as a covariance matrix and assume the nonlinearity to be of the form $f(x) = xg(x)$, where g is a scalar function satisfying $g(x) \le 0$ for all $x \in \mathbb{R}^n$. This setting covers interesting cases like the cubic function $x \mapsto x - x\|x\|_2^2$ which we can make part of the drift in (1.2a) by setting $g(x) = -\|x\|_2^2$. Note, however, that the classical results on rough differential equations found, e.g., in [14] can not be applied here to see that (1.2a) has a unique global-in-time solution since the drift may have superlinear growth. Instead, we can argue as follows: We first consider the corresponding equation without drift, i.e.,

$$dx(t) = N(x(t))\, K^{\frac{1}{2}}\, d\mathbf{W}(t). \tag{1.3}$$

The solution to (1.2a) can be obtained by a suitable flow decomposition of (1.3), cf. [32, Section 2]. Since (1.3) is a linear equation, we can use the bounds in [14, Section 10.7] to see that all solution trajectories with initial conditions in a ball with given radius $R > 0$ lie in a compact set $K_R \subset \mathbb{R}^n$. Therefore, for any $R > 0$, we can replace the linear vector fields in (1.3) by smooth vector fields having compact support by just redefining them to be zero outside K_R. Note that $b(x) = Ax + xg(x)$ satisfies [32, Condition (4.2) and (4.3)]. Therefore, we can argue as in [32, Theorem 4.3] to see that the solution to (1.2a) exists globally in time.

We introduce the Lyapunov operator

$$\mathcal{L}(X) := AX + XA^\top + \sum_{i,j=1}^{d} N_i X N_j^\top k_{ij} \tag{1.4}$$

for a simpler notation below, where X is an $n \times n$ matrix.

2 Approximating solution spaces based on a Gronwall lemma

Starting with a high-dimensional system (1.2), the goal of this section is to reduce the dimension and hence the computational complexity. This means that we aim to find a reduced system that has exactly the same dynamics, but is cheaper to evaluate for each potential driver $\mathbf{W}$. Then, we solve for $x(t)$, $t \in [0, T]$, from a low dimensional RDE instead of solving (1.2) directly. Here, the main obstacle is to identify and compute an approximation for the solution space of (1.2a). Below, this is done by studying covariance functions of Ito stochastic differential equations. We will study matrix inequalities that have to be understood in terms of definiteness. In particular, we write $M_1 \leq M_2$ for two matrices M_1 and M_2 if $M_2 - M_1$ is a positive semidefinite matrix. Let us first derive such a matrix inequality for a quadratic form of the solution of (1.2a) in case the rough driver is replaced by its smooth approximation.

Lemma 2.1 *Let $x^\epsilon(t)$, $t \in [0, T]$, satisfy*

$$dx^\epsilon(t) = [Ax^\epsilon(t) + f\left(x^\epsilon(t)\right)]dt + N\left(x^\epsilon(t)\right) K^{\frac{1}{2}} dW^\epsilon(t), \quad x^\epsilon(0) = x_0, \tag{2.1}$$

given that W^ϵ is absolutely continuous with representation in (1.1) and left-continuous $\dot{W}^\epsilon \in L_T^2$. Then, the quadratic form $X^\epsilon(t) = x^\epsilon(t)x^\epsilon(t)^\top \in \mathbb{R}^{n\times n}$ satisfies

$$\dot{X}^\epsilon(t) \leq \mathcal{L}(X^\epsilon(t)) + X^\epsilon(t) \left\| \dot{W}^\epsilon(t) \right\|_2^2, \quad X^\epsilon(0) = x_0 x_0^\top, \tag{2.2}$$

for all $t \in [0, T]$ in which W^ϵ is differentiable.

Proof We obtain by the product rule that

$$\begin{aligned} x^\epsilon(t)x^\epsilon(t)^\top &= x_0 x_0^\top + \int_0^t dx^\epsilon(v)x^\epsilon(v)^\top + \int_0^t x^\epsilon(v)dx^\epsilon(v)^\top \\ &= x_0 x_0^\top + \int_0^t \left[Ax^\epsilon(v)x^\epsilon(v)^\top + f\left(x^\epsilon(v)\right)x^\epsilon(v)^\top + x^\epsilon(v)x^\epsilon(v)^\top A^\top \right. \\ &\quad \left. + x^\epsilon(v) f\left(x^\epsilon(v)\right)^\top \right] dv + \int_0^t \left[N\left(x^\epsilon(v)\right) K^{\frac{1}{2}} \dot{W}^\epsilon(v) x^\epsilon(v)^\top \right. \\ &\quad \left. + x^\epsilon(v)\dot{W}^\epsilon(v)^\top K^{\frac{1}{2}} N\left(x^\epsilon(v)\right)^\top \right] dv. \end{aligned}$$

Now that $t \mapsto x^\epsilon(t)x^\epsilon(t)^\top$ is absolutely continuous, we can take the derivative which exists almost everywhere in points, where W^ϵ can be differentiated. Subsequently,

given two matrices M_1 and M_2 of suitable dimension, we exploit that $M_1 M_2^\top + M_2 M_1^\top \leq M_1 M_1^\top + M_2 M_2^\top$. In particular, we set $M_1 = N(x^\epsilon(v)) K^{\frac{1}{2}}$, $M_2 = x^\epsilon(v) \dot{W}^\epsilon(v)^\top$ and use that $f(x)x^\top = xf(x)^\top = xx^\top g(x) \leq 0$. This yields for almost all $t \in [0, T]$ that

$$\frac{d}{dt} x^\epsilon(t) x^\epsilon(t)^\top \leq A x^\epsilon(t) x^\epsilon(t)^\top + x^\epsilon(t) x^\epsilon(t)^\top A^\top + N\left(x^\epsilon(t)\right) K N\left(x^\epsilon(t)\right)^\top + x^\epsilon(t) \dot{W}^\epsilon(t)^\top \dot{W}^\epsilon(t) x^\epsilon(v)^\top.$$

The result follows by $N(x) K N(x)^\top = \sum_{i,j=1}^d N_i x x^\top N_j^\top k_{ij}$. □

We now find a (stochastic) representation for the respective equality in (2.2) based on a quadratic form of the solution of a linear Ito-stochastic differential equation.

Lemma 2.2 *Let B be a d-dimensional standard Brownian motion and $x_B(t)$, $t \geq 0$, be the solution to the following Ito-stochastic differential equation*

$$dx_B(t) = A x_B(t) dt + N(x_B(t)) K^{\frac{1}{2}} dB(t), \quad x_B(0) = x_0, \tag{2.3}$$

Then, $Z(t) = \mathbb{E}[x_B(t) x_B(t)^\top]$, $t \geq 0$, solves

$$Z(t) = x_0 x_0^\top + \int_0^t \mathcal{L}(Z(v)) dv. \tag{2.4}$$

Moreover, given the left-continuous $\dot{W}^\epsilon \in L_T^2$ from (1.1), the function $\bar{X}^\epsilon(t) = \exp\left\{\int_0^t \left\|\dot{W}^\epsilon(v)\right\|_2^2 dv\right\} Z(t)$, $t \in [0, T]$, solves the following matrix identity:

$$\bar{X}^\epsilon(t) = x_0 x_0^\top + \int_0^t \mathcal{L}(\bar{X}^\epsilon(v)) + \bar{X}^\epsilon(v) \left\|\dot{W}^\epsilon(v)\right\|_2^2 dv. \tag{2.5}$$

Proof Ito's product rule yields

$$x_B(t) x_B(t)^\top = x_0 x_0^\top + \int_0^t dx_B(v) x_B(v)^\top + \int_0^t x_B(v) dx_B(v)^\top + \int_0^t N(x_B(v)) K N(x_B(v))^\top dv.$$

We insert (2.3) above, take the expected value and utilize $N(x) K N(x)^\top = \sum_{i,j=1}^d N_i x x^\top N_j^\top k_{ij}$. Since the Ito-integral has mean zero, we obtain

$$\mathbb{E}[x_B(t) x_B(t)^\top] = x_0 x_0^\top + \int_0^t A \mathbb{E}[x_B(v) x_B(v)^\top] dv + \int_0^t \mathbb{E}[x_B(v) x_B(v)^\top] A^\top dv + \int_0^t \sum_{i,j=1}^d N_i \mathbb{E}[x_B(v) x_B(v)^\top] N_j k_{ij} dv$$

giving us the first part of the claim. Applying the product rule to $\bar{X}^\epsilon(t) = \exp\left\{\int_0^t \left\|\dot{W}^\epsilon(v)\right\|_2^2 dv\right\} Z(t)$ and taking (2.4) into account, we see that the second part of the result follows. □

Remark 2.3 First, we observe that mean square asymptotic stability, i.e., $\mathbb{E}\,\|x_B(t)\|_2^2 \to 0$ for all $x_0 \in \mathbb{R}^n$ as $t \to \infty$ is equivalent to $\mathbb{E}[x_B(t)x_B(t)^\top] \to 0$. For that reason, Lemma 2.2 tells us that mean square asymptotic stability is equivalent to the asymptotic stability of (2.4). It is well known that this is equivalent to

$$\lambda(\mathcal{L}) \subset \mathbb{C}_- = \{z \in \mathbb{C} : \mathrm{Re}(z) < 0\},$$

where $\lambda(\cdot)$ denotes the spectrum of an operator, and that the decay of the solution of (2.4) to zero is exponential. We refer to [8, 23, 28] for additional algebraic characterizations and for a further discussion on second moment exponential stability of (2.3). Let us further point out that this stability concept is stronger than almost sure exponential stability in the linear case, see [27, Theorem 4.2].

In the next step, a relation between solutions of (2.2) and (2.5) is pointed out. The following lemma can be interpreted as Gronwall type result for matrix differential inequalities/equations. We generalize arguments exploited in [29] in the corresponding proof.

Lemma 2.4 *Suppose that $\dot{W}^\epsilon \in L_T^2$ is left-continuous. Given an (absolutely) continuous $X^\epsilon(t)$, $t \in [0, T]$, satisfying (2.2) and $\bar{X}^\epsilon(t)$, $t \in [0, T]$, being the solution to the matrix integral equation (2.5). Then, we have that $X^\epsilon(t) \le \bar{X}^\epsilon(t)$ for all $t \in [0, T]$.*

Proof We introduce $Y := \bar{X}^\epsilon - X^\epsilon$ and the time-dependent Lyapunov operator $\mathcal{L}_t(Y) := \mathcal{L}(Y) + Y\left\|\dot{W}^\epsilon(t)\right\|_2^2$. From the integrated version of (2.2) and (2.5), we find that

$$Y(t) - Y(s) \ge \int_s^t \mathcal{L}_v(Y(v))dv, \quad s \le t. \tag{2.6}$$

We define $D(t) := Y(t) - \int_0^t \mathcal{L}_v(Y(v))dv$ and consider a perturbed integral equation

$$Y_\gamma(t) = \gamma I + \int_0^t [\mathcal{L}_v(Y_\gamma(v)) + \gamma I]dv + D(t) \tag{2.7}$$

with parameter $\gamma \ge 0$. By construction, we observe that $Y_0(t) = Y(t)$ for all $t \in [0, T]$. Moreover, it holds that $\lim_{\gamma \to 0} Y_\gamma(t) = Y(t)$ for all $t \in [0, T]$.

Below, let us assume that Y_γ is not positive definite for $\gamma > 0$ meaning that $\tilde{z}^\top Y_\gamma(\tilde{t})\tilde{z} \le 0$ for some $\tilde{z} \ne 0$ and $\tilde{t} > 0$. Y_γ is positive definite at $t = 0$ as $Y_\gamma(0) = \gamma I$. This is equivalent to all the eigenvalues of this matrix being positive. Now that Y_γ is continuous and takes values in the space of symmetric matrices, there exist continuous and real functions $\lambda_1, \ldots, \lambda_n$ such that $\lambda_1(t), \ldots, \lambda_n(t)$ represent the eigenvalues of $Y_\gamma(t)$ for each fixed $t \in [0, T]$, see [5, Corollary VI.1.6]. By assumption, at least one of these eigenvalue functions crosses or touches zero, while starting with a positive

value. Let λ_i be the one that reaches zero first at some $t_0 \in (0, \tilde{t}]$, i.e., t_0 is the smallest point of time with $\lambda_i(t_0) = 0$. Since we have $\lambda_i(t_0) = 0$ while all the other eigenvalues are nonnegative, Y_γ turns from a positive definite into a positive semidefinite matrix at this t_0 meaning that

$$z_0^\top Y_\gamma(t_0) z_0 = 0 \quad \text{and} \quad z_0^\top Y_\gamma(t) z_0 > 0, \quad 0 \le t < t_0, \tag{2.8}$$

for some $z_0 \neq 0$ while $z^\top Y_\gamma(t_0) z \ge 0$ for all $z \in \mathbb{R}^n$. Now, $\mathcal{L}_t$ is a Lyapunov operator for fixed $t \ge 0$ and hence resolvent positive, see Appendix A. The relation $0 = z_0^\top Y_\gamma(t_0) z_0 = \langle Y_\gamma(t_0), z_0 z_0^\top \rangle_F$ consequently implies $0 \le \langle \mathcal{L}_{t_0}(Y_\gamma(t_0)), z_0 z_0^\top \rangle_F = z_0^\top \mathcal{L}_{t_0}(Y_\gamma(t_0)) z_0$ according to Theorem A.2. As $\dot{W}^\epsilon$ is left-continuous, the same holds for $t \mapsto z_0^\top \mathcal{L}_t(Y_\gamma(t)) z_0$. For that reason, there exists a $\delta > 0$ such that $z_0^\top \mathcal{L}_v(Y_\gamma(v)) z_0 > -\gamma \|z_0\|_2^2$ for all $v \in (t_0 - \delta, t_0]$. Let $s, t \in (t_0 - \delta, t_0]$ with $s \le t$. Then,

$$\begin{aligned} z_0^\top Y_\gamma(t) z_0 - z_0^\top Y_\gamma(s) z_0 &= \int_s^t z_0^\top \mathcal{L}_v(Y_\gamma(v)) z_0 + \gamma \|z_0\|_2^2 \, dv + z_0^\top (D(t) - D(s)) z_0 \\ &\ge z_0^\top (D(t) - D(s)) z_0. \end{aligned}$$

From (2.6), we obtain that $D(t) - D(s) \ge 0$. Consequently, we know that $z_0^\top Y_\gamma(s) z_0 \le z_0^\top Y_\gamma(t) z_0$, i.e., $v \mapsto z_0^\top Y_\gamma(v) z_0$ is increasing on $(t_0 - \delta, t_0]$ which contradicts (2.8). Therefore, $Y_\gamma(t)$ is positive definite for all $t \in [0, T]$ and $\gamma > 0$. Taking the limit of $\gamma \to 0$, we obtain $Y(t) \ge 0$ for all $t \in [0, T]$ which concludes the proof. □

As a consequence of Gronwall Lemma 2.4, the following theorem can be established that provides information on the solution space of the considered rough differential equation.

Theorem 2.5 *Suppose that x is the solution of (1.2a) on $[0, T]$ with a driver $\mathbf{W}$. Then, it holds that*

$$x(t) \in \operatorname{im}[P_T], \quad t \in [0, T], \tag{2.9}$$

where $P_T = \int_0^T \mathbb{E}[x_B(t) x_B(t)^\top] dt$ with x_B solving the Ito-stochastic differential equation (2.3). If (2.3) further is mean square asymptotically stable, that is, $\mathbb{E}\|x_B(t)\|_2^2 \to 0$ as $t \to \infty$, the limit $P := \lim_{T \to \infty} P_T$ exists. Then, P_T can be replaced by P in (2.9).

Proof Let $z \in \ker[P_T]$ and let x^ϵ be the approximation of x defined by (2.1). Then,

$$\int_0^T \langle z, x^\epsilon(t) \rangle_2^2 dt = z^\top \int_0^T x^\epsilon(t) x^\epsilon(t)^\top dt \, z.$$

By Lemma 2.1, we observed that $x^\epsilon(t) x^\epsilon(t)^\top$ is a continuous solution to (2.2). Hence, it can be bounded from above by the solution $\bar{X}^\epsilon$ to (2.5) using Lemma 2.4. By Lemma

2.2, it is known that $\bar{X}^\epsilon(t) = \exp\left\{\int_0^t \left\|\dot{W}^\epsilon(v)\right\|_2^2 dv\right\} \mathbb{E}[x_B(t)x_B(t)^\top]$. Consequently, we have

$$\begin{aligned}\int_0^T \langle z, x^\epsilon(t)\rangle_2^2 dt &\leq \exp\left\{\int_0^T \left\|\dot{W}^\epsilon(v)\right\|_2^2 dv\right\} z^\top \int_0^T \mathbb{E}[x_B(t)x_B(t)^\top]dt\, z \\ &= \exp\left\{\int_0^T \left\|\dot{W}^\epsilon(v)\right\|_2^2 dv\right\} z^\top P_T z = 0. \end{aligned} \tag{2.10}$$

Since x^ϵ is continuous it follows that $\langle z, x^\epsilon(t)\rangle_2^2 = 0$, $t \in [0, T]$. Taking the limit as $\epsilon \to 0$, we find $\langle z, x(t)\rangle_2^2 = 0$ for all $t \in [0, T]$. This means that $x(t)$ is orthogonal to $\ker[P_T]$. By the symmetry of P_T the orthogonal complement of this kernel is $\mathrm{im}[P_T]$, so that the first claim follows. If the Ito-stochastic differential equation is mean square asymptotically stable, it decays exponentially fast to zero, see Remark 2.3. This implies exponential convergence of $\mathbb{E}[x_B(t)x_B(t)^\top]$ to zero. In this case, P exists and it holds that $z^\top P_T z \leq z^\top P z$ for all $z \in \mathbb{R}^n$. Now, choosing $z \in \ker[P]$, the second claim follows from (2.10). This concludes the proof. □

Remark 2.6 It can be shown that the solution x_B of (2.3) takes values in $\mathrm{im}[P_T]$ with probability one, see, e.g., [32]. Therefore, Theorem 2.5 tells us that $\mathrm{im}[P_T]$ does not only characterize the solution space of the Ito equation, but is much more universal: it can be used to describe the solution space of a linear RDE driven by a general geometric rough path. Note that to apply Theorem 2.5 in practice, we need to either compute P_T or P in order to find the superset for the solution space of the RDE. One might think of an empirical approach due to the stochastic representation of P_T or P, i.e., the Ito stochastic differential equation (2.3) can be simulated and subsequently one averages a quadratic form of x_B in ω and in time t. However, P_T or P can be computed exactly most of the time using the matrix identity (2.4).

We can now consider the eigenvalue decomposition of $\mathcal{P} \in \{P_T, P\}$ given by

$$\mathcal{P} = \begin{bmatrix} V_{\mathcal{P}} & \star \end{bmatrix} \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_{\mathcal{P}}^\top \\ \star \end{bmatrix} = V_{\mathcal{P}} \Lambda V_{\mathcal{P}}^\top,$$

where Λ is the diagonal matrix of non-zero eigenvalues of $\mathcal{P}$ and the matrix $V_{\mathcal{P}}$ of associated eigenvectors provides an orthonormal basis for $\mathrm{im}[\mathcal{P}]$. Therefore, we can find a reduced order function $x_r(t) \in \mathbb{R}^r$, with r being the number of non-zero eigenvalues, giving us $x(t) = V_{\mathcal{P}} x_r(t)$. Inserting this identity into (1.2) and multiplying the resulting equation with $V_{\mathcal{P}}^\top$ from the left leads to

$$\begin{aligned} dx_r(t) &= [V_{\mathcal{P}}^\top A V_{\mathcal{P}} x_r(t) + V_{\mathcal{P}}^\top f\,(V_{\mathcal{P}} x_r(t))]dt \\ &\quad + V_{\mathcal{P}}^\top N\,(V_{\mathcal{P}} x_r(t))\, K^{\frac{1}{2}} d\mathbf{W}(t), \quad x_r(0) = V_{\mathcal{P}}^\top x_0, \qquad (2.11a) \\ y(t) &= C V_{\mathcal{P}} x_r(t), \quad t \in [0, T]. \qquad (2.11b) \end{aligned}$$

It is important to notice that solving for x_r is potentially much cheaper than solving for the solution x of (1.2). In addition, the exact approximation is independent of the

driver $\mathbf{W}$. Consequently, the computational benefit is the larger the more often the system needs to be evaluated for different drivers, e.g., when running a Monte-Carlo simulation in case of a random rough path $\mathbf{W}$.

In the remainder of this section, we will briefly sketch the difficulties when trying to extend the here presented dimension reduction techniques to nonlinear dynamics. Let us consider a general nonlinear rough system given existence and uniqueness of a global solution, meaning that we study (1.2a) while assuming that f and N are general, but sufficiently nice nonlinearities. In this case, Theorem 2.5 could be reformulated for the operator $P_{\mathbf{W},T} := \int_0^T x(t)x(t)^\top dt$ in the sense that $x(t) \in \mathrm{im}[P_{\mathbf{W},T}]$, $t \in [0, T]$. This expression describes the solution space in the general framework. However, note that $P_{\mathbf{W},T}$ depends on the driver $\mathbf{W}$ and is not accessible unless x itself is computed. Therefore, we are lead to seek for an $\mathbf{W}$-independent matrix-valued function $Z(t)$ approximating $x(t)x(t)^\top$, $t \in [0, T]$, in the sense that the image of $P_T = \int_0^T Z(t)dt$ contains the image of $P_{\mathbf{W},T}$ and is computable in practice. This is the key idea of Theorem 2.5, where Z is the covariance function of an Ito stochastic differential equation in the linear case. In general, Z would still be the solution to a matrix ODE like in (2.4) and hence be related to an Ito equation, but the associated Lyapunov operator $\mathcal{L}$ would be different. It turns out that we would have to choose a linear $\mathcal{L}$ for which Lemma 2.2 holds true, but the existence of this operator is generally an open question. This approach can be seen as a *linearization of the Lyapunov operator* corresponding to the nonlinear system. However, the linearization is hard to construct in practice.[1]

3 Redundancies in the quantity of interest

Instead of looking at an approximation for the solution space of the state variable, let us now point out which states in x can be removed from the dynamics without an effect on y defined in (1.2b). This allows us to reduce the dimension of (2.11) further. Here, we assume a purely linear system, i.e., $f \equiv 0$ in (1.2a). Let $\mathcal{Z}$ denote the solution to

$$\mathcal{Z}(t) = C^\top C + \int_0^t \mathcal{L}^*(\mathcal{Z}(v))dv. \tag{3.1}$$

which can be interpreted as the dual equation of (2.4), where

$$\mathcal{L}^*(X) := A^\top X + XA + \sum_{i,j=1}^{d} N_i^\top X N_j k_{ij}. \tag{3.2}$$

[1] A more promising approach might be to approximate the solution x of a nonlinear RDE by a linear map of the (truncated) signature of $\mathbf{W}$ Such an approximation can always be constructed with an arbitrary accuracy due to the universal approximation theorem, see [[31], Theorem 3.1]. Since the (truncated) signature can be shown to be the solution to a linear RDE as considered in this paper, a model reduction technique to the signature equation would then lead to a linear reduced order model for the original nonlinear dynamics. However, the drawback is that the linear map of the signature is generally unknown, so that the problem is transferred to finding an accurate linearization first.

Here, $\mathcal{L}^*$ is the adjoint operator of $\mathcal{L}$ with respect to the Frobenius inner product $\langle \cdot, \cdot \rangle_F$. As $\mathcal{Z} = \mathcal{Z}(\cdot, C^\top C)$ is linear in its initial state, we obtain $\mathcal{Z}(t, C^\top C) = \sum_{\ell=1}^{p} = \mathcal{Z}(t, c_\ell^\top c_\ell)$, where c_ℓ is the ℓth row of C. By Lemma 2.2, we know that $\mathcal{Z}(t, c_\ell^\top c_\ell) = \mathbb{E}[\boldsymbol{x}_B(t)\boldsymbol{x}_B(t)^\top]$, where $\boldsymbol{x}_B$ solves the Ito-stochastic differential equation

$$d\boldsymbol{x}_B(t) = A^\top \boldsymbol{x}_B(t)dt + N^*(\boldsymbol{x}_B(t))\, K^{\frac{1}{2}} dB(t), \quad \boldsymbol{x}_B(0) = c_\ell^\top, \tag{3.3}$$

with $N^*(x) := \left[N_1^\top x \ \ldots \ N_d^\top x\right]$. This stochastic representation implies that $\mathcal{Z}(t)$ is a positive semidefinite matrix for all fixed t. This is exploited in the next lemma.

Lemma 3.1 *Let us define $Q_T := \int_0^T \mathcal{Z}(t)dt$, where $\mathcal{Z}$ solves (3.1). Suppose that $z \in \ker[Q_T]$. Then, we have*

$$Q_T A z = 0, \quad Cz = 0, \quad (K \otimes Q_T)\left[N_1^\top \ldots N_d^\top\right]^\top z = 0. \tag{3.4}$$

If the solution of (3.3) satisfies $\mathbb{E}\,\|\boldsymbol{x}_B(t)\|_2^2 \to 0$ as $t \to \infty$, the limit $Q := \lim_{T\to\infty} Q_T$ exists and (3.4) holds when replacing Q_T by its limit.

Proof Using (3.1) for $t = T$ and the linearity of $\mathcal{L}^*$, we obtain

$$\begin{aligned}\mathcal{Z}(T) &= C^\top C + \mathcal{L}^*(Q_T) = C^\top C + A^\top Q_T + Q_T A + \sum_{i,j=1}^{d} N_i^\top Q_T N_j k_{ij} \\ &= C^\top C + A^\top Q_T + Q_T A + \left[N_1^\top \ldots N_d^\top\right](K \otimes Q_T)\left[N_1^\top \ldots N_d^\top\right]^\top, \end{aligned} \tag{3.5}$$

since $\sum_{i,j=1}^{d} N_i^\top k_{ij} Q_T N_j = \left[N_1^\top \ldots N_d^\top\right](K \otimes Q_T)\left[N_1^\top \ldots N_d^\top\right]^\top$. Suppose that $z \in \ker[Q_T]$. Then, we have $0 = z^\top Q_T z = \int_0^T \left\| \mathcal{Z}(t)^{\frac{1}{2}} z \right\|_2^2 dt$ exploiting that $\mathcal{Z}(t)$ is positive semidefinite. As $\mathcal{Z}$ is continuous, we obtain that $\mathcal{Z}(t)z = 0$ for all $t \in [0, T]$. Now, we can multiply (3.5) with $z^\top$ from the left and z from the right yielding

$$0 = z^\top C^\top C z + z^\top \left[N_1^\top \ldots N_d^\top\right](K \otimes Q_T)\left[N_1^\top \ldots N_d^\top\right]^\top z. \tag{3.6}$$

$K \otimes Q_T$ is a positive semidefinite matrix, because K and Q_T are positive semidefinite. Hence, both summands on the right-hand side of (3.6) must be zero. Therefore, we have $Cz = 0$ and $(K \otimes Q_T)\left[N_1^\top \ldots N_d^\top\right]^\top z = 0$. With this knowledge, we multiply (3.5) only with z from the right resulting in $Q_T A z = 0$. Finally, $\mathbb{E}\,\|\boldsymbol{x}_B(t)\|_2^2 \to 0$ is equivalent to $\mathbb{E}[\boldsymbol{x}_B(t)\boldsymbol{x}_B(t)^\top] \to 0$ as $t \to \infty$. In particular, this convergence is exponential, see Remark 2.3. Therefore, $\mathcal{Z}$ converges exponentially fast to zero yielding the existence of $Q = \int_0^\infty \mathcal{Z}(t)dt$. Taking the limit of $T \to \infty$ in (3.5), this Q satisfies $0 = C^\top C + \mathcal{L}^*(Q)$, so that the above arguments can be used to proof the same result for Q instead of Q_T. □

Notice that mean square asymptotic stability of (3.3) exploited in Lemma 3.1 is equivalent to the same type of stability in (2.3) since $\lambda(\mathcal{L}^*) = \lambda(\mathcal{L})$, see Remark 2.3.

Let us introduce $\mathcal{Q} \in \{Q_T, Q\}$. Since $\mathcal{Q}$ is positive semidefinite, we can find an associated orthogonal basis for $\mathbb{R}^n$ consisting of eigenvectors (q_k) of $\mathcal{Q}$. We define the matrix $V_{\mathcal{Q}} := [\, q_1 \cdots q_r \,]$, where the columns of this matrix are the eigenvectors corresponding to the non zero eigenvalue of $\mathcal{Q}$. The remaining eigenvectors $q_{r+1}, \ldots, q_n$ form a basis for $\ker[\mathcal{Q}]$. We set $V_{\mathcal{Q}}^{\perp} := [\, q_{r+1} \cdots q_n \,]$. We can find processes $\boldsymbol{x}_r$ and $\tilde{\boldsymbol{x}}$, so that $x(t) = V_{\mathcal{Q}} \boldsymbol{x}_r(t) + V_{\mathcal{Q}}^{\perp} \tilde{\boldsymbol{x}}(t)$ which implies that $\boldsymbol{x}_r(t) = V_{\mathcal{Q}}^{\top} x(t)$. As a consequence of Lemma 3.1, we obtain that $y(t) = Cx(t) = CV_{\mathcal{Q}} \boldsymbol{x}_r(t)$. Now, the differential equation associated to $\boldsymbol{x}_r$ is obtained by

$$d\boldsymbol{x}_r(t) = V_{\mathcal{Q}}^{\top} dx(t) = V_{\mathcal{Q}}^{\top} Ax(t)dt + V_{\mathcal{Q}}^{\top} N\,(x(t))\, K^{\frac{1}{2}} d\mathbf{W}(t), \quad \boldsymbol{x}_r(0) = V_{\mathcal{Q}}^{\top} x_0.$$

By Lemma 3.1, we have $V_{\mathcal{Q}}^{\top} Ax(t) = V_{\mathcal{Q}}^{\top} A(V_{\mathcal{Q}} \boldsymbol{x}_r(t) + V_{\mathcal{Q}}^{\perp} \tilde{\boldsymbol{x}}(t)) = V_{\mathcal{Q}}^{\top} AV_{\mathcal{Q}} \boldsymbol{x}_r(t)$. Moreover, given that the covariance matrix K is invertible, we can multiply the last identity of (3.4) with $K^{-1} \otimes I$ providing $\mathcal{Q} N_i z = 0$ for $z \in \ker[\mathcal{Q}]$ and all $i = 1, \ldots, d$. This can now be exploited to obtain that $V_{\mathcal{Q}}^{\top} N\,(x(t)) = V_{\mathcal{Q}}^{\top} N\,(V_{\mathcal{Q}} \boldsymbol{x}_r(t))$. Let us summarize the above considerations in the following theorem.

Theorem 3.2 *Given $\mathcal{Q} \in \{Q_T, Q\}$ defined in Lemma 3.1, K being invertible and $f \equiv 0$. Then, we find a reduced order system with the same quantity of interest like (1.2). It is given by*

$$d\boldsymbol{x}_r(t) = V_{\mathcal{Q}}^{\top} AV_{\mathcal{Q}} \boldsymbol{x}_r(t)dt + V_{\mathcal{Q}}^{\top} N\,(V_{\mathcal{Q}} \boldsymbol{x}_r(t))\, K^{\frac{1}{2}} d\mathbf{W}(t), \quad \boldsymbol{x}_r(0) = V_{\mathcal{Q}}^{\top} x_0, \tag{3.7a}$$

$$y(t) = CV_{\mathcal{Q}} \boldsymbol{x}_r(t), \quad t \in [0, T], \tag{3.7b}$$

with $V_{\mathcal{Q}} := [\, q_1 \cdots q_r \,]$, where $q_1, \ldots, q_r$ are orthonormal eigenvectors of $\mathcal{Q}$ corresponding to all r non zero eigenvalues.

4 Numerical experiments

Let the regularity of $\mathbf{W} = (W, \mathbb{W})$ now be $\alpha \in (1/3, 1/2]$. As before, we assume that it can be approximate (w.r.t. the rough path metric) by the lift of W^{ϵ} with representation (1.1).

4.1 Linear rough PDEs and Feynman-Kac solutions

We aim to study the solution

$$[0, T] \times \mathbb{R}^m \ni (t, x) \mapsto u(t, x)$$

to the initial value problem

$$du = L(u)\, dt + \sum_{k=1}^{d} \Gamma_k(u)\, d\mathbf{W}_k, \quad u(0, \cdot) = g, \tag{4.1}$$

where L and $\Gamma = (\Gamma_1, \dots, \Gamma_d)$ are

$$\begin{aligned} Lh(\zeta) &:= \frac{1}{2}\operatorname{tr}\left(\sigma(\zeta)\sigma(\zeta)^\top D^2 h(\zeta)\right) + \langle b(\zeta), Dh(\zeta)\rangle + c(\zeta)h(\zeta), \\ \Gamma_k h(\zeta) &:= \langle \beta_k(\zeta), Dh(\zeta)\rangle + \gamma_k(\zeta)h(\zeta), \end{aligned}$$

for a suitable test function $h : \mathbb{R}^m \to \mathbb{R}$. For the Feynman-Kac approach, it will be convenient to apply the time change $t \mapsto T - t$ and to study the equivalent terminal value problem

$$-dv = L(v)\,dt + \sum_{k=1}^{d} \Gamma_k(v)\,d\overleftarrow{\mathbf{W}}_k, \quad v(T,\cdot) = g, \tag{4.2}$$

instead where $\overleftarrow{\mathbf{W}}$ denotes the time reversed rough path $\overleftarrow{\mathbf{W}}(t) = \mathbf{W}(T-t)$. If we replace $\overleftarrow{\mathbf{W}}$ by $\overleftarrow{W}^\epsilon$, then every bounded $v^\epsilon \in C^{1,2}([0,T]\times\mathbb{R}^m, \mathbb{R})$ solution to the PDE (driven by $\overleftarrow{W}^\epsilon$) has the Feynman-Kac representation

$$v^\epsilon(t,\zeta) = \mathbb{E}\left[g(x^\zeta(T))\exp\left(\int_t^T c(x^\zeta(s))\,ds + \int_t^T \gamma(x^\zeta(s))\,\dot{\overleftarrow{W}}^\epsilon(s)\,ds\right)\right], \tag{4.3}$$

where x^ζ is the solution to the Ito-stochastic differential equation

$$\begin{aligned} dx^\zeta(s) &= b\left((x^\zeta(s))\right)ds + \left[\sigma\left(x^\zeta(s)\right)\beta\left(x^\zeta(s)\right)\right]\begin{bmatrix} dB(s) \\ d\overleftarrow{W}^\epsilon(s)\end{bmatrix}, \quad t \le s \le T, \\ x^\zeta(t) &= \zeta, \end{aligned}$$

If $v^\epsilon \notin C^{1,2}([0,T]\times\mathbb{R}^m,\mathbb{R})$, we use (4.3) to define the solution of the PDE as long as the associated stochastic differential equation admits a unique solution. Now, given that the initial value g is continuous and bounded, [10, 11, 13] showed that for v^ϵ in (4.3), it holds that

$$v^\epsilon(t,\zeta) \to v(t,\zeta) := \mathbb{E}\left[g(x^\zeta(T))\exp\left(\int_t^T c(x^\zeta(s))\,ds + \int_t^T \gamma(x^\zeta(s))\,d\overleftarrow{\mathbf{W}}(s)\right)\right], \tag{4.4}$$

point-wise in time and space. Here, x^ζ is the solution to the rough differential equation

$$\begin{aligned} dx^\zeta(s) &= b\left((x^\zeta(s))\right)ds + \left[\sigma\left(x^\zeta(s)\right)\beta\left(x^\zeta(s)\right)\right]d\mathbf{Z}(s), \quad t \le s \le T, \\ x^\zeta(t) &= \zeta, \end{aligned} \tag{4.5}$$

with the joint rough path $\mathbf{Z} = (Z, \mathbb{Z})$

$$Z(t) := \begin{pmatrix} B(t) \\ \overleftarrow{W}(t) \end{pmatrix},$$
$$\mathbb{Z}_{s,t} := \begin{pmatrix} \int_s^t (B(v) - B(s)) \otimes dB(v) & \int_s^t (\overleftarrow{W}(v) - \overleftarrow{W}(s)) \otimes dB(v) \\ \int_s^t (B(v) - B(s)) \otimes d\overleftarrow{W}(v) & \overleftarrow{\mathbb{W}}_{s,t} \end{pmatrix},$$

where the stochastic integrals are understood as Ito-integrals. The limit v now defines the solution to (4.1) given that (4.5) has a unique solution.

4.2 Dimension reduction for spatially discretized rough heat equations

We specify the coefficients for our numerical experiments by setting $\sigma \equiv \sqrt{2}I$, $b \equiv 0$ and $c \equiv 0$ resulting in the rough heat equation

$$du(t,\zeta) = \Delta u(t,\zeta)\, dt + \sum_{k=1}^{d} \big(\langle \beta_k(\zeta), \nabla u(t,\zeta) \rangle + \gamma_k(\zeta) u(t,\zeta) \big) d\mathbf{W}_k(t), \quad u(0,\cdot) = g. \tag{4.6}$$

Instead of exploiting the Feynman-Kac representation in (4.4), we formally discretize (4.1) by a finite difference scheme. Moreover, we consider the bounded spatial domain $[0,1]^m$ (in contrast to the above Feynman-Kac theory). Here, we set additional boundary conditions which are assumed to be of Dirichlet type. Notice that equation (4.6) can then also be defined in the mild sense (for general non geometric drivers) when the transport term is absent, see [13, 16].

For simplicity let us set $m = 1$. Then, $h_\zeta := \frac{1}{(n+1)}$ is supposed to be the spatial step size parameter leading to a grid $\zeta_j = j h_\zeta$ for $j = 0, 1, \ldots, n+1$. Intuitively, we find that $x_j(t) \approx u(t, \zeta_j)$, where

$$\begin{aligned} dx_1(t) &= \frac{x_2(t) - 2x_1(t)}{h_\zeta^2} dt + \sum_{k=1}^{d} \Big(\beta_k(\zeta_1) \frac{x_2(t) - x_1(t)}{h_\zeta} + \gamma_k(\zeta_1) x_1(t) \Big) d\mathbf{W}_k(t), \\ dx_j(t) &= \frac{x_{j+1}(t) - 2x_j(t) + x_{j-1}(t)}{h_\zeta^2} dt \\ &\quad + \sum_{k=1}^{d} \Big(\beta_k(\zeta_j) \frac{x_{j+1}(t) - x_j(t)}{h_\zeta} + \gamma_k(\zeta_j) x_j(t) \Big) d\mathbf{W}_k(t), \\ dx_n(t) &= \frac{-2x_n(t) + x_{n-1}(t)}{h_\zeta^2} + \sum_{k=1}^{d} \Big(\beta_k(\zeta_n) \frac{-x_n(t)}{h_\zeta} + \gamma_k(\zeta_n) x_n(t) \Big) d\mathbf{W}_k(t) \end{aligned} \tag{4.7}$$

for $j \in \{2, \ldots, n-1\}$ taking into account that $u(t,0) = u(t,1) = 0$. The initial condition associated to (4.7) is $x(0) = \big(g(\zeta_1) \ \ldots \ g(\zeta_n)\big)^\top$. W shall now be 2-dimensional, where its components are paths of independent fractional Brownian motions with Hurst index $H = 0.4$. Further, let us set $n = 100$, $\gamma_1(\zeta) = 4\sin(\zeta)$, $\gamma_2(\zeta) = 4\cos(\zeta)$,

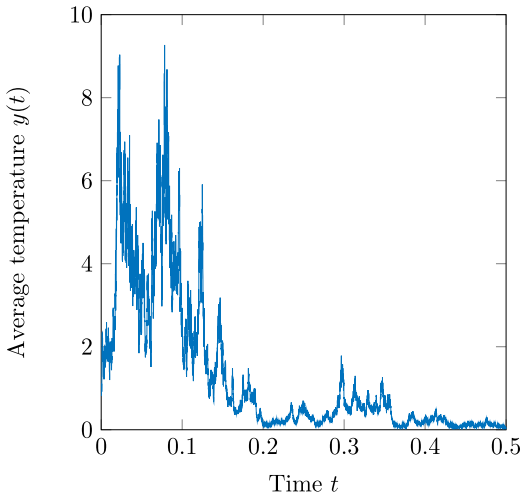

Fig. 1 Output y in (4.8) of (4.7) with driver in Fig. 2

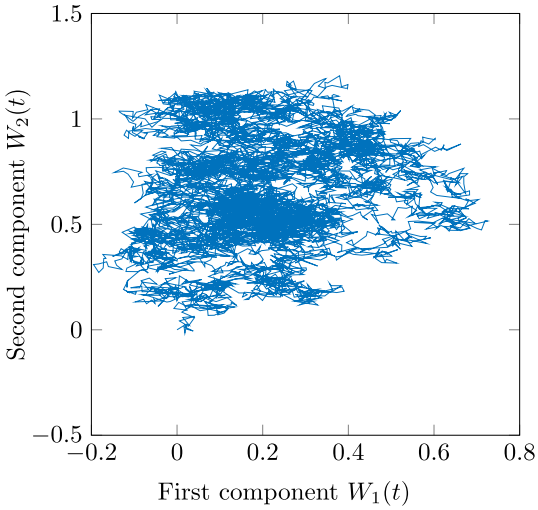

Fig. 2 Path of a 2D fractional Brownian motion used as driver

$\beta_1 \equiv 0.4, \beta_2 \equiv -0.2$ and $g(\zeta) = \mathrm{e}^{-2|\zeta-0.5|^2}, \zeta \in [0, 1]$. We fix $T = 0.5$ and introduce the quantity of interest

$$y(t) = \frac{1}{n}\sum_{j=1}^{n} x_j(t) \tag{4.8}$$

being the average temperature, i.e., $C = \frac{1}{n}\left[1 \ldots 1\right]$. We illustrate y in Fig. 1 given the driver depicted in Fig. 2.

Consequently, (4.7) together with (4.8) yield a system of the form (1.2) with $f \equiv 0$. Moreover, notice that (4.7) is a mean square asymptotically stable system given the above parameters. Therefore, P and Q, introduced in Theorem 2.5 and Lemma 3.1,

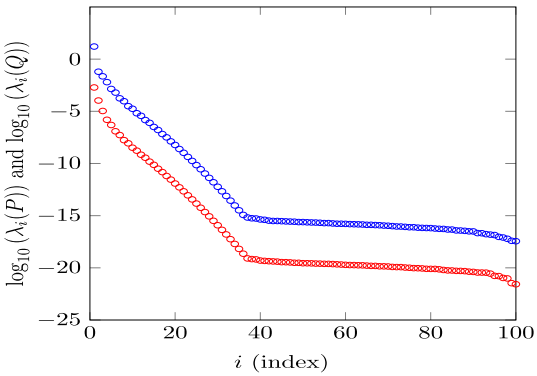

Fig. 3 Logarithmic eigenvalues of P (blue) and Q (red) (Color figure online)

exist and can be used to identify unnecessary information. In particular, P and Q can be computed much easier than P_T and Q_T. We obtain them from solving $0 = x_0 x_0^\top + \mathcal{L}(P)$ and $0 = C^\top C + \mathcal{L}^*(Q)$ which are the equations derived by taking the limit as $t \to \infty$ in (2.4) and (3.1). We observe from Fig. 3 that P and Q have many eigenvalues below machine precision that are numerically zero.

As a first step we remove the ones of P resulting in a reduced model (2.11) of order 35. Subsequently, the dimension of this system can be lowered further by applying the procedure of Sect. 3. Here, two eigenvalues below machine precision can be detected finally providing a model of dimension $r = 33$ in which we do not expect any reduction error. However, it is important to notice that there are several sources of numerical errors like, for instance, the time discretization leading to a non zero error in practice. For that reason, we denote the output of the reduced system by y_r and find a relative L_T^2-error $\frac{\|y-y_r\|_{L_T^2}}{\|y\|_{L_T^2}} = 1.5710\text{e}{-14}$ for $r = 33$. This can be assumed to be an exact approximation neglecting the other numerical errors. In addition, the logarithm of the point-wise error $\frac{|y(t)-y_r(t)|}{|y(t)|}$ for the same setting is shown in Fig. 5. Finally, we conducted experiments related to dimension reduction with a true error. In detail, in addition to the (numerical) zero eigenvalues, we neglect eigenspaces of P and Q that are associated to very small eigenvalues of which we have many, see Fig. 3. This is motivated by an observation in Ito-SDE settings, where those direction have a tiny influence on the dynamics, see, e.g, [28]. Fig. 4 depicts the relative L_T^2-errors for $r = 5, 7, 9, \ldots, 33$ in logarithmic scale. We observe a small error in each case, e.g., of order 1e−08 for an r around 20, Moreover, the deviation from the true output is below one percent even for $r = 5$. This illustrates that rough differential equations can have a very high reduction potential beyond truncating state variables that have no contribution (Fig. 5).

We conclude by explaining the time discretization used in order to obtain the simulation results. We implemented an implicit Runge–Kutta scheme for rough differential equations [22, 31] with "optimal" rate solely based on the increments of the driver. Here, the implicit nature is required due to the stiffness of (4.7). As a first step, we

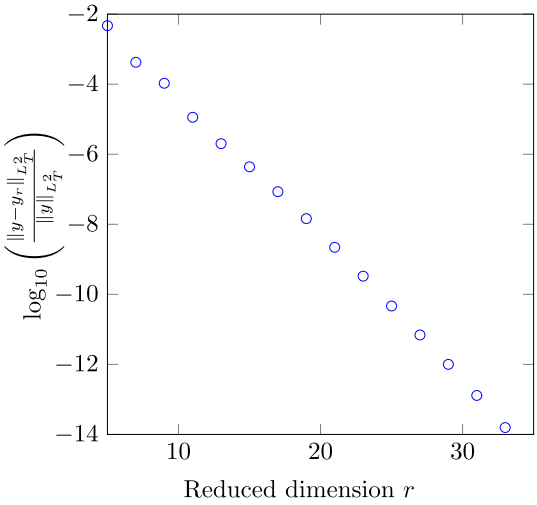

Fig. 4 Logarithmic relative L^2_T-error for different reduced dimensions $r = 5, 7, 9, \ldots, 33$

rewrite (1.2a) as

$$dx(t) = F\left(x(t)\right) d\tilde{\mathbf{W}}(t), \quad x(0) = x_0,$$

where $F(x) := \begin{bmatrix} Ax + f(x) & N(x)K^{\frac{1}{2}} \end{bmatrix}$ and $\tilde{W}(t) = \begin{bmatrix} t \\ W(t) \end{bmatrix}$. Given an equidistant partition $t_k = k\mathfrak{h}$ of $[0, T]$ with the step size $\mathfrak{h}$, we use the following scheme

$$\begin{aligned} Z_{k,i} &= z_k + \sum_{j=1}^{s} a_{ij} F(Z_{k,j})\big(\tilde{W}(t_{k+1}) - \tilde{W}(t_k)\big) \\ z_{k+1} &= z_k + \sum_{i=1}^{s} b_i F(Z_{k,i})\big(\tilde{W}(t_{k+1}) - \tilde{W}(t_k)\big), \end{aligned}$$

with $z_0 = x_0$ aiming that $z_k \approx x(t_k)$. In particular, Crouzeix's two stages ($s = 2$) and diagonally implicit method is exploited that has the following Butcher tableau

$$\begin{array}{c|cc} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \hline b_1 & b_2 \end{array} = \begin{array}{c|cc} \frac{1}{2} + \frac{\sqrt{3}}{6} & 0 \\ \frac{-\sqrt{3}}{3} & \frac{1}{2} + \frac{\sqrt{3}}{6} \\ \hline \frac{1}{2} & \frac{1}{2} \end{array}.$$

This method satisfies the optimality conditions provided in [31] and hence has a convergence order arbitrary close to $2H - 0.5$, where $\frac{1}{3} < H \leq \frac{1}{2}$ is the Hurst index of a fractional Brownian motion. Now, let us mention that all the above simulations have been conducted setting $\mathfrak{h} = 2^{-14}$.

4.3 Dimension reduction for spatially discretized rough wave equations

In this section, we provide another example, for which we apply our approach. We aim to consider a higher order system and illustrate the dimension reduction potential for non parabolic equations. In particular, we consider the following symbolic[2] damped rough wave equation

$$\frac{\partial^2}{\partial t^2}u(t,\zeta)+\alpha\frac{\partial}{\partial t}u(t,\zeta)=\frac{\partial^2}{\partial \zeta^2}u(t,\zeta)+\sum_{k=1}^{d}\gamma_k(\zeta)u(t,\zeta)\frac{\partial}{\partial t}\mathbf{W}_k(t), \tag{4.9a}$$

$$u(0,\cdot)=g,\quad \frac{\partial}{\partial t}u(0,\cdot)=\dot{g},\quad u(t,0)=u(t,1)=0, \tag{4.9b}$$

on a spatial domain $[0,1]$, for $t\in[0,1]$ and $\alpha>0$. We can rewrite (4.9) in order to obtain a system of the form like in (4.1). This is

$$d\begin{pmatrix}u(t,\zeta)\\ \frac{\partial}{\partial t}u(t,\zeta)\end{pmatrix}=\begin{pmatrix}0 & I\\ \frac{\partial^2}{\partial\zeta^2} & -\alpha I\end{pmatrix}\begin{pmatrix}u(t,\zeta)\\ \frac{\partial}{\partial t}u(t,\zeta)\end{pmatrix}dt \quad +\sum_{k=1}^{d}\begin{pmatrix}0 & 0\\ \gamma_k(\zeta) & 0\end{pmatrix}\begin{pmatrix}u(t,\zeta)\\ \frac{\partial}{\partial t}u(t,\zeta)\end{pmatrix}d\mathbf{W}_k(t),$$
$$(u(0,\cdot)\ \tfrac{\partial}{\partial t}u(0,\cdot))^\top=(g\ \dot{g})^\top.$$

However, notice that the above equation is not parabolic but hyperbolic. Such equations are generally harder to reduce in their dimension. Formally applying a finite difference scheme as in Sect. 4.2 with $\tilde{n}$ grid points and the interpretation that $x_j(t)\approx u(t,\zeta_j)$ and $\dot{x}_j(t)\approx\frac{\partial}{\partial t}u(t,\zeta_j)$, $j\in\{1,\dots,\tilde{n}\}$, leads to a system of order $n=2\tilde{n}$:

$$d\begin{pmatrix}x(t)\\ \dot{x}(t)\end{pmatrix}=\begin{pmatrix}0 & I\\ A & -\alpha I\end{pmatrix}\begin{pmatrix}x(t)\\ \dot{x}(t)\end{pmatrix}dt \qquad +\sum_{k=1}^{d}\begin{pmatrix}0 & 0\\ N_k & 0\end{pmatrix}\begin{pmatrix}x(t)\\ \dot{x}(t)\end{pmatrix}d\mathbf{W}_k(t),$$
$$(x(0)\ \dot{x}(0))^\top=(x_0\ \dot{x}_0)^\top, \tag{4.10}$$

where $A\in\mathbb{R}^{\tilde{n}\times\tilde{n}}$ is the finite difference discretization of the second derivative, $x(0)=\big(g(\zeta_1)\ \dots\ g(\zeta_{\tilde{n}})\big)^\top$, $\dot{x}(0)=\big(\dot{g}(\zeta_1)\ \dots\ \dot{g}(\zeta_{\tilde{n}})\big)^\top$, I is the $\tilde{n}\times\tilde{n}$ identity matrix and $N_k=\operatorname{diag}\big(\gamma_k(\zeta_1),\dots,\gamma_k(\zeta_{\tilde{n}})\big)$. We use the driver in Fig. 2 again, i.e., $d=2$. We pick the parameter $\gamma_1(\zeta)=4\sin(\zeta)$, $\gamma_2(\zeta)=4\cos(\zeta)$ and $g(\zeta)\equiv 0$, $\dot{g}(\zeta)=\mathrm{e}^{-2|\zeta-0.5|^2}$, $\zeta\in[0,1]$. In addition, we fix $\alpha=1$ and $\tilde{n}=500$ spatial grid points resulting in a system of order $n=1000$. The quantities of interest are the average position and velocity, i.e.,

$$y(t)=(y_1(t)\ y_2(t))^\top,\quad \text{where}\quad y_1(t)=\frac{1}{\tilde{n}}\sum_{j=1}^{\tilde{n}}x_j(t),\quad y_2(t)=\frac{1}{\tilde{n}}\sum_{j=1}^{\tilde{n}}\dot{x}_j(t). \tag{4.11}$$

[2] To our knowledge, there does not exist any solution theory yet that covers this equation.

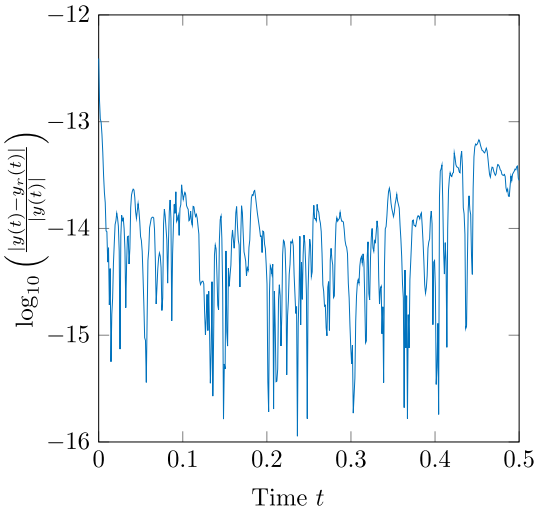

Fig. 5 Logarithmic relative error in time for reduced dimension $r = 33$

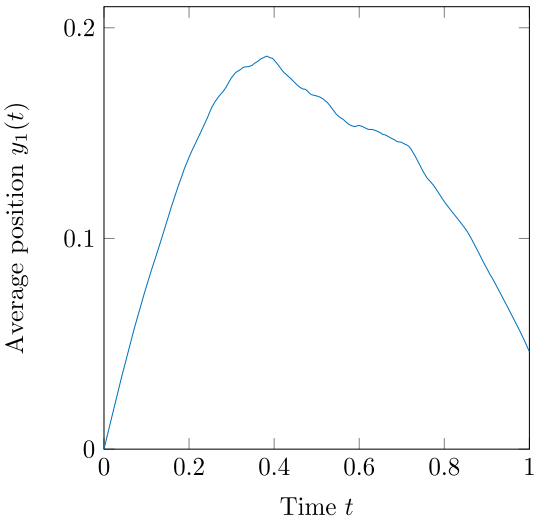

Fig. 6 Average position y_1 in (4.11) of (4.10) with driver in Fig. 2

For illustration purposes, we depict both variables y_1 and y_2 in Figs. 6 and 7. The position y_1 is a smooth function, whereas the velocity y_2 is as rough as the driver.

We exploit the observation of Sect. 4.2 that eigenspaces of P and Q related to non-zero eigenvalues can be removed in case they are small enough. The result of the inexact dimension reduction can be seen in Fig. 8, where the relative L^2_T-errors for reduced dimensions $r = 10, 20, 30, \ldots, 300$ are shown in logarithmic scale. We observe a small error in each case, i.e., deviations of less than one percent can easily be achieved for small r. However, notice that the reduction potential is lower than in the heat equation case. We encounter a (numerical) zero error for dimensions $r \geq 492$, meaning that we are able to truncate more than half of the variables without a change in the outputs y_1 and y_2 in (4.11).

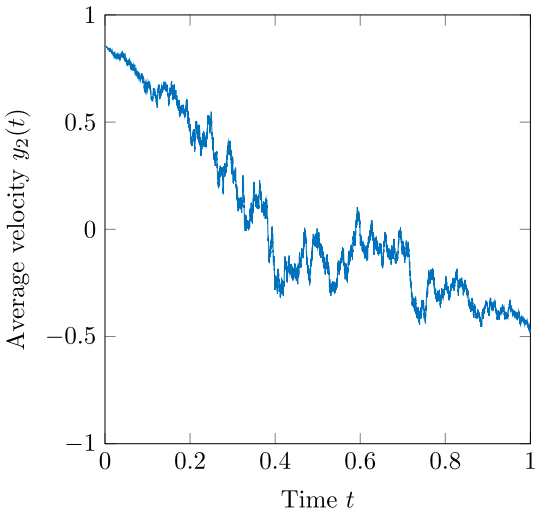

Fig. 7 Average velocity y_2 in (4.11) of (4.10) with driver in Fig. 2

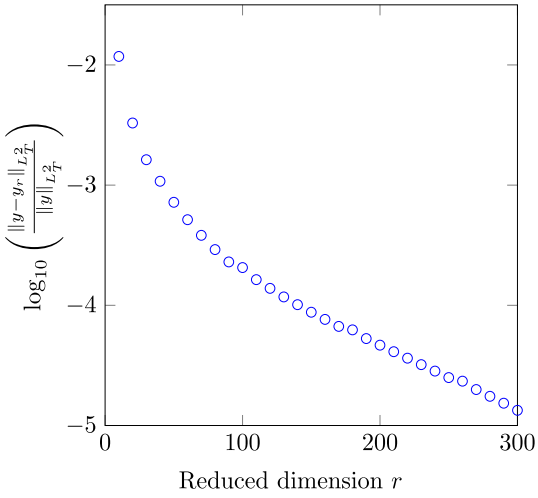

Fig. 8 Logarithmic relative L^2_T-error for the dimension reduction of (4.10) with $n = 1000$ and different reduced dimensions $r = 10, 20, 30, \ldots, 300$

A Resolvent positive operators

This section covers the essential information on resolvent positive operators that are required in this paper. We refer to [8] for a more detailed and more general discussion. In particular, we are interested in such operators on $(H^n, \langle\cdot,\cdot\rangle_F)$ which shall be the Hilbert space of symmetric $n \times n$ matrices and $\langle\cdot,\cdot\rangle_F$ denotes the Frobenius inner product. Further suppose that H^n_+ is the associated subset of symmetric positive semidefinite matrices. We begin with the definition of positive and resolvent positive operators on H^n.

Definition A.1 A linear operator $\mathcal{L} : H^n \to H^n$ is called positive if $\mathcal{L}(H^n_+) \subset H^n_+$. It is resolvent positive if there is an $\alpha_0 \in \mathbb{R}$ such that for all $\alpha > \alpha_0$ the operator $(\alpha I - \mathcal{L})^{-1}$ is positive.

The Lyapunov operator defined in (1.4) is resolvent positive observing that it is a composition of a resolvent positive operator $X \mapsto AX + XA^\top$ and a positive part $X \mapsto \sum_{i,j=1}^{d} N_i X N_j^\top k_{ij}$, see [8]. Below, we state an equivalent characterization of resolvent positive operators and refer once more to [8, Section 3.2.2] for a more general framework.

Theorem A.2 *A linear operator $\mathcal{L} : H^n \to H^n$ is resolvent positive if and only if $\langle V_1, V_2 \rangle_F = 0$ implies $\langle \mathcal{L} V_1, V_2 \rangle_F \geq 0$ for $V_1, V_2 \in H^n_+$.*

Acknowledgements MR is supported by the DFG via the individual grant "Low-order approximations for large-scale problems arising in the context of high-dimensional PDEs and spatially discretized SPDEs"–project number 499366908.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest We have no Conflict of interest to disclose.

References

1. Antoulas, A.C.: Approximation of large-scale dynamical systems. Adv. in Design and Control. SIAM (2005)
2. Bayer, C., Belomestny, D., Redmann, M., Riedel, S., Schoenmakers, J.: Solving linear parabolic rough partial differential equations. J. Math. Anal. Appl. **490**(1), 124236 (2020)
3. Benner, P., Cohen, A., Ohlberger, M., Willcox, K. (eds): Model reduction and approximation, volume 15 of Computational Science & Engineering. Society for Industrial and Applied Mathematics (SIAM), Theory and algorithms, Philadelphia, (2017)
4. Benner, P., Redmann, M.: Model reduction for stochastic systems. Stoch. PDE: Anal. Comp. **3**(3), 291–338 (2015)
5. Bhatia, R.: Matrix analysis, vol. 169. Springer, Berlin (1997)
6. Caruana, M., Friz, P.: Partial differential equations driven by rough paths. J. Differ. Equ. **247**(1), 140–173 (2009)
7. Caruana, M., Friz, P.K., Oberhauser, H.: A (rough) pathwise approach to a class of non-linear stochastic partial differential equations. Ann. Inst. H. Poincaré C Anal Non Linéaire, 28(1):27–46, (2011)
8. Damm, T.: Rational matrix equations in stochastic control. Lecture Notes in Control and Information Sciences, vol. 297. Springer, Berlin (2004)
9. Deya, A.: Numerical schemes for rough parabolic equations. Appl. Math. Optim. **65**(2), 253–292 (2012)
10. Diehl, J.: *Topics in stochastic differential equations and rough path theory*. PhD thesis, Technical University of Berlin, (2012)
11. Diehl, J., Oberhauser, H., Riedel, S.: A Lévy area between Brownian motion and rough paths with applications to robust nonlinear filtering and rough partial differential equations. Stoch. Process. Appl. **125**(1), 161–181 (2015)

12. Friz, P.K., Gassiat, P., Lions, P.-L., Souganidis, P.E.: Eikonal equations and pathwise solutions to fully non-linear SPDEs. Stoch. Partial Differ. Equ. Anal. Comput. **5**(2), 256–277 (2017)
13. Friz, P.K., Hairer, M.: A course on rough paths With an introduction to regularity structures, 2nd edn. Springer, Cham (2020)
14. Friz, P.K., Victoir, N.B.: Multidimensional stochastic processes as rough paths, Theory and applications. Cambridge Studies in Advanced Mathematics, vol. 120. Cambridge University Press, Cambridge (2010)
15. Gubinelli, M., Lejay, A., Tindel, S.: Young integrals and SPDEs. Potential Anal. **25**(4), 307–326 (2006)
16. Gubinelli, M., Tindel, S.: Rough evolution equations. Ann. Probab. **38**(1), 1–75 (2010)
17. Hairer, M.: Rough stochastic PDEs. Comm. Pure Appl. Math. **64**(11), 1547–1585 (2011)
18. Hairer, M.: Solving the KPZ equation. Ann. Math. **178**(2), 559–664 (2013)
19. Hairer, M.: A theory of regularity structures. Invent. Math. **198**(2), 269–504 (2014)
20. Hairer, M., Matetski, K.: Discretisations of rough stochastic PDEs. Ann. Probab. **46**(3), 1651–1709 (2018)
21. Hairer, M., Weber, H.: Rough Burgers-like equations with multiplicative noise. Probab. Theory Related Fields **155**(1–2), 71–126 (2013)
22. Hong, J., Huang, C., Wang, X.: Symplectic Runge-Kutta methods for Hamiltonian systems driven by Gaussian rough paths. Appl. Numer. Math. **129**, 120–136 (2018)
23. Khasminskii, R.Z.: Stochastic stability of differential equations. Monographs and Textbooks on Mechanics of Solids and Fluids. Mechanics: Analysis, 7. Alphen aan den Rijn, The Netherlands; Rockville, Maryland, USA: Sijthoff & Noordhoff. (1980)
24. Levin, D., Lyons, T., Ni, H. Learning from the past, predicting the statistics for the future, learning an evolving system. arXiv:1309.0260v6 (2016)
25. Lyons, T.J.: Differential equations driven by rough signals. Rev. Mat. Iberoamericana **14**(2), 215–310 (1998)
26. Lyons, T.J., Caruana, M., Lévy, T.: *Differential equations driven by rough paths*, volume 1908 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007. Lectures from the 34th Summer School on Probability Theory held in Saint-Flour, July 6–24, With an introduction concerning the Summer School by Jean Picard (2004)
27. Mao, X.: Stochastic differential equations and applications, 2nd edn. Woodhead Publishing, Sawston (2007)
28. Redmann, M.: Type II singular perturbation approximation for linear systems with Lévy noise. SIAM J. Control Optim. **56**(3), 2120–2158 (2018)
29. Redmann, M.: Bilinear Systems-A New Link to $\mathcal{H}_2$-norms, Relations to Stochastic Systems, and Further Properties. SIAM J. Control Optim. **59**(4), 2477 (2021)
30. Redmann, M., Bayer C., Goyal, P. Low-Dimensional Approximations of High-Dimensional Asset Price Models. *SIAM Journal on Financial Mathematics*, 12(1):1–28 (2021)
31. Redmann, M., Riedel, S.: Runge-Kutta methods for rough differential equations. *J. Stoch. Anal.* **3**(4) (2022)
32. Riedel, S., Scheutzow, M.: Rough differential equations with unbounded drift term. J. Differ. Equ. **262**(1), 283–312 (2017)