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

Preprint ISSN 2198-5855

Dimension reduction for path signatures

Christian Bayer[1], Martin Redmann[2]

submitted: January 14, 2025

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

[2] Martin Luther University of Halle Wittenberg
Institute of Mathematics
Theodor-Lieser-Str. 5
06120 Halle (Saale)
Germany
E-Mail: martin.redmann@mathematik.uni-halle.de

No. 3163
Berlin 2025

2020 *Mathematics Subject Classification.* 60H10, 60L10, 60L90, 65C30, 93A15.

Key words and phrases. Signature models, applications of rough analysis, stochastic differential equations, model order reduction, Petrov-Galerkin projections, financial models.

CB acknowledges support from DFG CRC/TRR 388 "Rough Analysis, Stochastic Dynamics and Related Fields", Projects B01 and B03. 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.

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/

Dimension reduction for path signatures

Christian Bayer, Martin Redmann

Abstract

This paper focuses on the mathematical framework for reducing the complexity of models using path signatures. The structure of these signatures, which can be interpreted as collections of iterated integrals along paths, is discussed and their applications in areas such as stochastic differential equations (SDEs) and financial modeling are pointed out. In particular, exploiting the rough paths view, solutions of SDEs continuously depend on the lift of the driver. Such continuous mappings can be approximated using (truncated) signatures, which are solutions of high-dimensional linear systems. In order to lower the complexity of these models, this paper presents methods for reducing the order of high-dimensional truncated signature models while retaining essential characteristics. The derivation of reduced models and the universal approximation property of (truncated) signatures are treated in detail. Numerical examples, including applications to the (rough) Bergomi model in financial markets, illustrate the proposed reduction techniques and highlight their effectiveness.

1 Introduction

The path signature (Chen [Che57]) – i.e., the sequence of all iterated integrals of the components of a path $x : [0, T] \to \mathbb{R}^d$ – plays a fundamental rule in the analysis of (rough) paths and functions on path space. On the theoretical side, the signature is the fundamental building block of Lyons' theory of *rough paths* [Lyo98], providing pathwise analysis of differential equations driven by rough paths (i.e., Hölder continuous paths of arbitrary Hölder coefficient), see [FH20; FV10] for modern accounts of the theory. Indeed, a rough path is nothing but a path together with a (properly truncated) signature.

Some of the most important properties of path signatures include:

- The signature encodes the path up to so-called tree-like excursions, and it completely determines the path, provided that we add running time as a component.
- The signature is invariant under time re-parameterization.
- The linear functionals of the signature form an algebra of functionals on path space.

In applications, signatures play a fundamental role for approximations of functionals on *(rough) path space*, comparable to polynomials on finite-dimensional spaces. Indeed, it is well-known that linear functionals on signatures are *universal approximators*, i.e., that continuous functions mapping path-space to the real numbers can be approximated by linear functionals of the signature when restricted to compact sets. Several alternative formulations also exist, we refer to [LO23] for a recent review article and [BPS23] for a version on approximation in an L^p-sense.

Of course, similar approximation results are also available in a probabilistic rather than pathwise setting, i.e., when considering stochastic processes rather than single paths. A corresponding universal approximation result in terms of signatures (of semi-martingales) can be found in [CPSF22], but we also like to mention the classical stochastic Taylor expansion in terms of iterated integrals of the Brownian motion.

DOI 10.20347/WIAS.PREPRINT.3163 Berlin 2025

The path signature has been found to be a powerful feature map in machine learning, when dealing with time series data, see T. Lyons' ICM lecture [Lyo14], for instance, [PA+18; Bue+20; Mor+20; LM22; Arr+24] for specific examples. Here, both the fact that signatures provide (efficient) encodings of paths (of arbitrary length) in a fixed vector space and the invariance under time reparameterizations are crucial.

Beyond machine learning, signatures can be used as a fundamental building block for modelling and stochastic analysis of processes with memory – or, more generally, when we only want to impose minimal assumptions on the model. An example for the use of signatures in modelling is [CGSF22], where a diffusion-type model with coefficients given as linear functionals of the signature of the underlying Brownian motion is fitted to market prices of options.

Beyond modelling, the signature can be a powerful tool for analysis and numerics of non-Markovian processes. In particular, for BSDEs (see, for instance, [BFZ24]) or stochastic optimal control ([KLPA20; BHR24]) we are often dealing with computations of conditional expectations w.r.t. the relevant sigma-algebra $\mathcal{F}_t$. In the Markovian case, this corresponds to a finite-dimensional regression problem. But in the non-Markovian case (i.e., with memory), such conditionals expectations generally depend on the whole history of the driving path up to time t. Such path-dependent conditional expectations can be efficiently computed as (linear or non-linear) functions of the signature of the underlying path, see [BPS23].

From a computational point of view, the signature poses severe challenges mainly due to its size. Indeed, a priori, the signature – the collection of *all* iterated integrals – is an infinite-dimensional object, and, hence, not directly computable. For some applications, the *signature kernel* (see [LO23]) can be used instead, which is actually computable as a solution of a PDE. Otherwise, the most natural approach is to *truncate* the signature, i.e., to restrict the computations to iterated integrals up to a given degree. Even so, the size of the (finite) sequence of iterated integrals of degree up to m increases rapidly in terms of m as well as the dimension d of the underlying space: in fact, it is easy to see that there are $1 + d + \cdots + d^m$ such terms – including one trivial term of order 0, think of an "intercept" in a linear model.

This work addresses these concerns by exploring model order reduction (MOR) techniques specifically tailored for signature-based models. MOR has a long tradition in the field of deterministic control systems [Moo81], where difficult to control and hard to observe states are neglected to receive a low-dimensional approximation of a large-scale model. This work is the foundation for establishing balancing-related MOR for controlled SDEs [BH19; BD11; BR15; BDRC17]. Complexity reduction in such probabilistic frameworks has an enormous impact, e.g., when sampling methods are applied. However, this work relies on certain stability assumptions that are not necessarily satisfied in practice. A way to overcome this issue is to use the ansatz of [GJ90] established for controlled systems of ordinary differential equations and extended to SDEs in [RJ22]. The control theory perspective on MOR does no longer work when dimension reduction is applied to financial models. Let us refer to [Red22; RBG21], where a bridge was built between the previously mentioned concepts and complexity reduction for large-scale asset price models. Detached from control concepts and stability assumptions, MOR for signatures is studied in this paper. We develop a robust mathematical framework to reduce the complexity of these models while preserving their essential features, provide a comprehensive method for deriving reduced order models of truncated signatures, with discussions on their theoretical foundations and practical implications. The efficiency of the proposed techniques is illustrated through numerical examples, including an application to the (rough) Bergomi model–a widely used model in financial markets. The results demonstrate the effectiveness of MOR in reducing the dimension without a significant loss of accuracy.

Indeed, we apply MOR to realistic signature-models of financial markets, as derived from popular equity models (the classical Bergomi model [Ber15] and the rough Bergomi model [BFG16], respectively). Starting from signature models of dimension $n = 1365$ in the Bergomi case, we find that a

reduced model of dimension $\widetilde{n} = 5$ achieves a relative error of order 10^{-2} in the implied volatility of option prices, $\widetilde{n} = 11$ is sufficient for a relative error of 10^{-4}, and $\widetilde{n} = 27$ is exact up to machine precision, see Figure 4.

In case of the rough Bergomi model, we choose an approximate signature model of dimension $n = 3280$. A reduced dimension of $\widetilde{n} = 15$ achieves a relative error of 10^{-2}, and $\widetilde{n} = 55$ achieves machine precision in the rough case, see Figure 8.

These numerical illustrations show that MOR can alleviate one of the main drawbacks of signature based models and methods, namely the fast explosion of the signature's dimension in terms of the dimension of the underlying process as well as the truncation level.

Outline We start by providing an introduction to path signatures in Section 2. The basic structures are introduced in the setting of smooth paths and later extended to signatures of continuous semimartingales. We also provide the fundamental approximation results for signatures and minimal necessary results from rough path theory. As a motivation, we summarize results on signature models in finance by [CGSF22].

Section 3 provides a derivation of MOR for signatures. Note that (truncated) signatures of continuous semi-martingales solve a linear system of SDEs with specific structure, to which general MOR for SDEs is adapted. The reduced signature equation is introduced in Section 3.3.

Finally, we provide numerical examples showcasing the success of MOR for signature models in finance in Section 4.

2 Signatures

2.1 Signatures of smooth paths

Consider a smooth (i.e., $\mathcal{C}^1$) path[1] $x : [0,T] \to V$ taking values in a Banach space V – for our purposes, we assume $d := \dim V < \infty$. Given a multi-index $I = (i_1, \dots, i_n) \in \{1, \dots, d\}^n$, $n \geq 1$, we denote the iterated integrals

$$\mathbb{x}^I_{s,t} := \int_{s<t_1<\cdots<t_n<t} \dot{x}^{i_1}_{t_1} \mathrm{d}t_1 \cdots \dot{x}^{i_n}_{t_n} \mathrm{d}t_n =: \int_{s<t_1<\cdots<t_n<t} \mathrm{d}x^{i_1}_{t_1} \cdots \mathrm{d}x^{i_n}_{t_n} \in \mathbb{R}, \tag{2.1}$$

where $0 \leq s \leq t \leq T$. Here, x^i_t denotes the ith coordinate of the path x evaluated at time t for a fixed basis $e_1, \dots, e_d$ of V. We prefer to use the notation "$\mathrm{d}x_t$" rather than "$\dot{x}_t \mathrm{d}t$" because the former corresponds to the standard notation in the non-smooth case to be considered later – think of paths of Brownian motion.

Note that the collection of all iterated integrals $\mathbb{x}^I_{s,t}$ for $I \in \{1, \dots, d\}^n$ with fixed $n \geq 1$ takes values in $V^{\otimes n}$ recursively defined by $V^{\otimes 0} := \mathbb{R}$, and $V^{\otimes(n+1)} := V^{\otimes n} \otimes V$. Specifically, we write

$$\left(\mathbb{x}^I_{s,t}\right)_{I \in \{1,\dots,d\}^n} =: \int_{s<t_1<\cdots<t_n<t} \mathrm{d}x_{t_1} \otimes \cdots \otimes \mathrm{d}x_{t_n} \in V^{\otimes n}.$$

The *signature* is the collection of all such iterated integrals, formally

$$\mathbb{x}^{<\infty}_{s,t} := 1 + \sum_{n=1}^{\infty} \int_{s<t_1<\cdots<t_n<t} \mathrm{d}x_{t_1} \otimes \cdots \otimes \mathrm{d}x_{t_n} \in T((V)), \tag{2.2}$$

[1] In fact, all results of this subsection remain valid for continuous bounded variation paths.

with the *extended tensor algebra* $T((V))$ being defined as

$$T((V)) \coloneqq \prod_{n=0}^{\infty} V^{\otimes n}. \tag{2.3}$$

Here, the initial term "1" is considered as the (sole) entry in $V^{\otimes 0} \simeq \mathbb{R}$. Note that elements of the extended tensor algebra $T((V))$ have infinitely many terms, and are, hence, comparable to formal power series – in d non-commuting variables $e_1, \dots, e_d$. For numerical purposes, signatures need to be truncated at a finite degree. Hence, we also define

$$\mathbb{x}_{s,t}^{\leq m} \coloneqq 1 + \sum_{n=1}^{m} \int_{s<t_1<\cdots<t_n<t} \mathrm{d}x_{t_1} \otimes \cdots \otimes \mathrm{d}x_{t_n} \in T^m(V) \coloneqq \bigoplus_{n=0}^{m} V^{\otimes n}. \tag{2.4}$$

Following our analogy from above, an element of the truncated tensor algebra $T^m(V)$ can be compared with a polynomial of degree m – note, however, that the tensor product is not commutative.

Both $T((V))$ and $T^m(V)$ are algebras under the tensor product $\otimes$ – the product of formal power series in non-commuting variables. We refer to [Bay+23a] for more information. Using this product, we see that the signature (formally) solves a controlled ODE, namely $\dot{\mathbb{x}}_{0,t}^{<\infty} = \mathbb{x}_{0,t}^{<\infty} \otimes \dot{x}_t$ or (in a notation easier to adapt for non-smooth paths)

$$\mathrm{d}\mathbb{x}_{0,t}^{<\infty} = \mathbb{x}_{0,t}^{<\infty} \otimes \mathrm{d}x_t = \sum_{i=1}^{d} \mathbb{x}_{0,t}^{<\infty} \otimes e_i \mathrm{d}x_t^i, \quad \mathbb{x}_{0,0}^{<\infty} = 1 \in T((V)), \tag{2.5}$$

where e_i, $i = 1, \dots d$, denote the standard basis vectors of V as lifted to elements of $T((V))$. The truncated signature satisfies the same ODE, but interpreted on the truncated tensor algebra $T^m(V)$, i.e., with a nilpotent tensor product, which will – abusing notation – still be denoted $\otimes$. That is, the truncated signature satisfies

$$\mathrm{d}\mathbb{x}_{0,t}^{\leq m} = \mathbb{x}_{0,t}^{\leq m} \otimes \mathrm{d}x_t = \sum_{i=1}^{d} \mathbb{x}_{0,t}^{\leq m} \otimes e_i \mathrm{d}x_t^i, \quad \mathbb{x}_{0,0}^{\leq m} = 1 \in T^m(V). \tag{2.6}$$

Remark 2.1. We do not endow $T((V))$ with a topology, and, hence, the ODE (2.5) is only defined in a formal way. Several topologies for $T((V))$ have been considered in the rough path literature, including natural Hilbert or Banach sub-spaces of $T((V))$, see [CO22] for more details. In this paper, we will only ever use the full signature defined on $T((V))$ for motivation, the mathematical analysis will take place on $T^m(V)$ for finite truncation m. As a finite dimensional vector space, $T^m(V)$ will be equipped with the standard Euclidean metric.

The signature is invariant under re-parameterization of the path: I.e., if we consider a (smooth, increasing) map, say, $\gamma : [R, S] \to [0, T]$, and a path $\bar{x} : [R, S] \to V$, $u \mapsto x_{\gamma(u)}$, then $\mathbb{x}_{\gamma(s),\gamma(t)}^{<\infty} = \bar{\mathbb{x}}_{s,t}^{<\infty}$, $R \leq s \leq t \leq S$. Up to re-parameterization and so-called tree-like excursions, the signature $\mathbb{x}_{0,T}^{<\infty}$ uniquely determines the path x, see [HL10]. Invariance under parameterization as well as the possibility of tree-like excursions can be avoided by adding the component t to the path. We will often prefer to work with time-extended paths – i.e., paths with a component $x_t^1 \equiv t$.

We will also need to consider the dual algebra $\mathcal{W}_d \simeq T(V^*) \coloneqq \bigoplus_{n=0}^{\infty} (V^*)^{\otimes n}$ of linear functionals on $T((V))$. Here, $\mathcal{W}_d$ denotes the linear span of all words $\mathbf{w} = i_1 \cdots i_k$, $k \geq 0$, in the alphabet $\{1, \dots, d\}$, which is an algebra under the concatenation product on words, extended with the distributive property. Consider a generic linear functional $\ell \in \mathcal{W}_d$ applied to a generic element $\mathfrak{a} \in T((V))$. We can represent ℓ as a linear combination $\ell = \sum_{i=1}^{k} \gamma_i \mathbf{w}_i$ of words $\mathbf{w}_j = i_1^j \cdots i_{K_j}^j$ of length K_j, $j = 1, \dots, k$, for some $k \geq 1$, and

$$\mathfrak{a} = \sum_{n=0}^{\infty} \sum_{I=(i_1,\dots,i_n) \in \{1,\dots,d\}^n} \alpha_I e_{i_1} \otimes \cdots \otimes e_{i_n} \in T((V)),$$

owing to the fact that words $i_1^j \cdots i_{K_j}^j$ form a basis of $\mathcal{W}_d$ and tensor products $e_{i_1} \otimes \cdots \otimes e_{i_n}$ form a basis of any truncated tensor algebra $T^m(V)$.[2] We set

$$\langle \ell \, , \mathfrak{a} \rangle \coloneqq \sum_{j=1}^{k} \gamma_j \, \langle \mathbf{w}_j \, , \mathfrak{a} \rangle \coloneqq \sum_{j=1}^{k} \gamma_j \alpha_{(i_1^j, \ldots, i_{K_j}^j)} \in \mathbb{R}. \tag{2.7}$$

2.2 Signatures for Hölder continuous paths

Let $x \colon [0,T] \to V$ be an α-Hölder continuous path, $\alpha \in (0,1]$. This means that the α-Hölder seminorm

$$\sup_{s<t} \frac{\|x_t - x_s\|}{|t-s|^\alpha}$$

is finite. Then, we write $x \in C^\alpha$. We briefly sketch the concepts of rough paths and (truncated) signatures for $x \in C^\alpha$ – see [FH20] for more details. Formally, a rough path $\mathbf{x}$ is a two-parameter function from the simplex $\{\, 0 \le s \le t \le T \,\}$ taking values in the truncated tensor algebra $T^m(V)$.

Remark 2.2. The different "levels" of a rough path increment $\mathbf{x}_{s,t}$, $s \le t$, have different interpretations. The V-valued component is the increment $x_{s,t} = x_t - x_s$ of the underlying C^α path x itself. However, Hölder paths are not regular enough to allow us to solve controlled differential equations of the form

$$\mathrm{d}y_t = U(y_t)\mathrm{d}x_t.$$

However, a formal Taylor expansion shows that higher order Euler approximations of the differential equation in terms of the iterated integral of x of order up to $m = \lfloor 1/\alpha \rfloor$ *would converge*. Of course, the catch is that – once again – C^α paths are not regular enough for the iterated integrals to make sense in a classical way when $\alpha \le 1/2$. On the flip side, if we enhance the path increments $x_{s,t}$ with higher order terms $\mathbb{x}_{s,t}^I$ "behaving like iterated integrals" (see [FH20] for details), then we can solve the corresponding controlled differential equations for smooth enough vector fields in a pathwise, deterministic way.

It is well-known that such *rough path lifts* of C^α paths x are *always* possible, but *not* unique when $\alpha \le 1/2$ (think about Itô versus Stratonovich solutions of stochastic differential equations – corresponding to two different rough path lifts of Brownian motion).

We introduce a metric for two parameter functions taking values in $T^m(V)$ by

$$\varrho_\alpha(\mathbf{x}, \tilde{\mathbf{x}}) \coloneqq \sum_{n=1}^{m} \sup_{s<t} \frac{\|\mathbb{x}_{s,t}^{(n)} - \tilde{\mathbb{x}}_{s,t}^{(n)}\|}{|t-s|^{n\alpha}},$$

where $\mathbf{x}_{s,t} \coloneqq 1 + \sum_{n=1}^m \mathbb{x}_{s,t}^{(n)} \in T^m(V)$ and $\tilde{\mathbf{x}}$ is defined accordingly – with $\mathbb{x}_{s,t}^{(1)} = x_{s,t} \in V$ and $\mathbb{x}_{s,t}^{(n)} \in V^{\otimes n}$, $n = 2, \ldots, m$. Now, we can define the desired concepts.

Definition 2.3. Let $x \in C^\alpha$ and choose $N \coloneqq \lfloor 1/\alpha \rfloor$. A two-parameter function

$$\mathbf{x}_{s,t} = 1 + \sum_{n=1}^{N} \mathbb{x}_{s,t}^{(n)} \in T^N(V)$$

with $\mathbb{x}_{s,t}^{(1)} = x_{s,t}$ is called a *(geometric α-Hölder) rough path associated to* x if there exists a sequence of smooth paths x^ϵ with truncated signature $(\mathbb{x}^\epsilon)_{s,t}^{\le N}$, such that

$$\varrho_\alpha(\mathbf{x}, (\mathbb{x}^\epsilon)^{\le N}) \to 0, \quad \text{as } \epsilon \to 0.$$

[2] The infinite sum defining $\mathfrak{a}$ needs to be understood as a formal sum.

We denote the set of all geometric α-Hölder rough paths by $\mathscr{C}_g^\alpha([0,T];V)$. We also introduce $\widehat{\mathscr{C}}_g^\alpha([0,T];V) := \left\{ \mathbf{x} \in \mathscr{C}_g^\alpha([0,T];V) \;\middle|\; \forall t \in [0,T] : \, x^1_{0,t} = t \right\}$, where $x_{s,t}$ denotes the level-1 component of $\mathbf{x}_{s,t}$.

For general m, the (truncated) signature $\mathbb{x}^{\leq m}$ and $\mathbb{x}^{<\infty}$ of a geometric α-Hölder rough path $\mathbf{x}$ can be defined as the limit of $(\mathbb{x}^\epsilon)^{\leq m}$ and $(\mathbb{x}^\epsilon)^{<\infty}$, respectively, as $\epsilon \to 0$.

Definition 2.4. Suppose that W is another (finite-dimensional) Banach space, $\mathbf{x} \in \mathscr{C}_g^\alpha([0,T];V)$ and $U : W \to L(V,W)$. An α-Hölder path $y : [0,T] \to W$ is called a *solution of the rough differential equation*

$$\mathrm{d}y(t) = U(y(t))\,\mathrm{d}\mathbf{x}_t, \quad y(0) = y_0 \in W, \tag{2.8}$$

if $y(0) = y_0$ and for a sequence (x^ϵ) of smooth paths with $\varrho_\alpha(\mathbf{x}, (\mathbb{x}^\epsilon)^{\leq N}) \to 0$ (as $\epsilon \to 0$), the solutions y^ϵ of

$$\mathrm{d}y^\epsilon(t) = U(y^\epsilon(t))\,\mathrm{d}x^\epsilon_t, \quad y^\epsilon(0) = y_0,$$

exist and converge in the α-Hölder metric to y.

By construction of a solution of a rough differential equation, the truncated signature solves the equation

$$\mathrm{d}\mathbb{x}^{\leq m}_{0,t} = \mathbb{x}^{\leq m}_{0,t} \otimes \mathrm{d}\mathbf{x}_t = \sum_{i=1}^{d} \mathbb{x}^{\leq m}_{0,t} \otimes e_i \mathrm{d}\mathbf{x}^i_t, \quad \mathbb{x}^{\leq m}_{0,0} = 1 \in T^m(V). \tag{2.9}$$

Next, we formulate a result on existence and uniqueness of a solution of (2.8) as well as on properties of the solution map.

Theorem 2.5. *Given $\mathbf{x} \in \mathscr{C}^\alpha([0,T],V)$ for $\alpha \in \left(0, \frac{1}{2}\right]$ and $U \in \mathcal{C}_b^{N+1}(W, L(V,W))$ or linear, where $N := \lfloor 1/\alpha \rfloor$. Then, there is a unique solution $y \in C^\alpha([0,T],W)$ for (2.8). Moreover, the solution map $\mathbf{x} \mapsto g(\mathbf{x}) = y$ of (2.8) is locally Lipschitz continuous.*

Proof. We refer to [FH20] for the case $\alpha > 1/3$. The more general (p-variation) framework can be found in [FV10]. □

Remark 2.6. As evidenced by the difference between Itô and Stratonovich solutions to SDEs, it is, in general, not possible to find solutions to controlled differential equations driven by rough signals x such that the solution map ($x \mapsto y$) is continuous in a pathwise sense. The rough path approach factorizes this map into two parts:

1 the lift of x to a rough path $\mathbf{x}$ (discontinuous);

2 the solution map $\mathbf{x} \mapsto y$ (locally Lipschitz continuous by Theorem 2.5).

This fact is sometimes called the "rough path principle".

2.3 Signatures of semi-martingales

If, instead of a single deterministic path x, we are given a continuous semi-martingale $X : \Omega \times [0,T] \to V$, then the statements above remain true, mutatis mutandis, provided that iterated integrals are defined as *Stratonovich integrals* rather than Itô integrals.

We define the (truncated) signatures $\mathbb{X}_{s,t}^{<\infty}$, $\mathbb{X}_{s,t}^{\leq m}$ as in (2.2) and (2.4), but with the iterated integrals (2.1) replaced by

$$\mathbb{X}_{s,t}^{I} \coloneqq \int_{s<t_1<\cdots<t_n<t} \circ \mathrm{d}X_{t_1}^{i_1} \otimes \cdots \otimes \circ \mathrm{d}X_{t_n}^{i_n}. \tag{2.10}$$

Indeed, we can see that the thus defined rough path lift $\mathbf{X}$ a.s. takes values in $\mathscr{C}_g^\alpha([0,T];V)$ for any $1/3 < \alpha < 1/2$ – we again refer to [FH20] for details. Moreover, the solution of a stochastic differential equation, say

$$\mathrm{d}Y_t = U(Y_t) \circ \mathrm{d}X_t, \quad Y_0 = y_0 \in W,$$

coincides a.s. with the (pathwise) solution of the rough differential equation

$$\mathrm{d}Y_t = U(Y_t)\mathrm{d}\mathbf{X}_t, \quad Y_0 = y_0 \in W,$$

where $U : W \to L(V,W)$ is sufficiently smooth.

Remark 2.7. From a modeling point, it may be advantageous to include the quadratic variation of X in the construction of the signature, i.e., to consider the signature of $(t, X_t, \langle X \rangle_t)$. While it is included in our framework, we will not require such an extension.

Remark 2.8. The setting can also incorporate semimartingales with jumps, see, for instance, [CPSF22] for a precise statement and proof of universality of signatures in that case.

2.4 Universal approximation theorem

We will now formulate *universality* of signatures, a well known result in the literature (see, for instance, [KLPA20; BPS23]). Recall that $\widehat{\mathscr{C}}_g^\alpha([0,T];V)$ denotes the set of geometric α-Hölder rough paths $\mathbf{x}$ such that the first component x^1 of the underlying path x is equal to running time, see Definition 2.3. As noted in the smooth case in Section 2.1, the signature $\mathbb{x}_{0,T}^{<\infty}$ of a time-extended rough path $\mathbf{x} \in \widehat{\mathscr{C}}_g^\alpha([0,T];V)$ characterizes $\mathbf{x}$ – and, hence, x up to the initial value x_0. This, together with the fact that linear functionals of the signature form an algebra – also already mentioned for smooth paths in Section 2.1, but equally true for rough paths – allow us to apply the Stone-Weierstrass theorem.

Theorem 2.9. *For any compact subset $\mathcal{K} \subset \widehat{\mathscr{C}}_g^\alpha([0,T];V)$, any continuous function $f : \widehat{\mathscr{C}}_g^\alpha([0,T];V) \to \mathbb{R}$, and any $\epsilon > 0$, we can find a linear functional $\ell \in \mathcal{W}_d$ such that*

$$\sup_{\mathbf{x}\in\mathcal{K}} \left| f(\mathbf{x}) - \left\langle \ell, \mathbb{x}_{0,T}^{<\infty} \right\rangle \right| < \epsilon.$$

Proof of Theorem 2.9. For completeness, we give a short proof of this well-known result, a direct consequence of the Stone-Weierstrass theorem. Hence, we need to prove that

$$\mathcal{A} \coloneqq \left\{ \mathbf{x} \mapsto \left\langle \ell, \mathbb{x}_{0,T}^{<\infty} \right\rangle \;\middle|\; \ell \in \mathcal{W}_d,\ \mathbf{x} \in \widehat{\mathscr{C}}_g^\alpha([0,T];V) \right\}$$

is a subalgebra of $C\left(\widehat{\mathscr{C}}_g^\alpha([0,T];V);\mathbb{R}\right)$ which is point-separating and contains a non-zero constant function. First note that for any $\ell \in \mathcal{W}_d$, the map $\mathbf{x} \mapsto \left\langle \ell, \mathbb{x}_{0,T}^{<\infty} \right\rangle$ is continuous, see, for instance, [FV10]. $\mathcal{A}$ is point-separating since the signature $\mathbb{x}^{<\infty}$ uniquely determines the rough path $\mathbf{x}$ by the discussion above. The constant function $\mathbf{x} \mapsto 1$ is obviously contained in $\mathcal{A}$ by simply choosing $\ell = \emptyset$, where $\emptyset$ denotes the empty word.

This leaves us to prove that $\mathcal{A}$ is an algebra. In addition to the concatenation product, $\mathcal{W}_d$ is also equipped with a commutative shuffle product $\sqcup\!\sqcup : \mathcal{W}_d \times \mathcal{W}_d \to \mathcal{W}_d$, see, e.g., [Bay+23a], and the

following *shuffle identity* holds for signatures: For $\ell_1, \ell_2 \in \mathcal{W}_d$ and a geometric rough path $\mathbf{x}$, we have

$$\left\langle \ell_1 \,, \mathbb{x}_{s,t}^{<\infty} \right\rangle \left\langle \ell_2 \,, \mathbb{x}_{s,t}^{<\infty} \right\rangle = \left\langle \ell_1 \sqcup\!\sqcup \ell_2 \,, \mathbb{x}_{s,t}^{<\infty} \right\rangle . \tag{2.11}$$

Hence, $\mathcal{A}$ is an algebra and the proof is complete. □

Example 2.10. Let g be the solution map of (2.8), i.e., $y = g(\mathbf{x})$, mapping a rough path $\mathbf{x}$ to a path y. Then, we know by Theorem 2.5 that this map is continuous. For that reason, a potential continuous functional of interest in Theorem 2.9 can be the ith component y_i of y evaluated at T (e.g., think of $W = \mathbb{R}^d$) meaning that $f(\mathbf{x}) = y_i(T)$.

As noted in Section 2.3, we can also lift semi-martingales X to the associated rough path $\mathbf{X}$ taking values in $\widehat{\mathscr{C}}_g^\alpha([0,T];V)$– assuming $X_t^1 \equiv t$ –, and Theorem 2.9 applies:

Corollary 2.11. *Given a continuous functional $F : \widehat{\mathscr{C}}_g^\alpha([0,T];V) \to \mathbb{R}$ and a compact subset $\mathcal{K} \subset \widehat{\mathscr{C}}_g^\alpha([0,T];V)$. Then, for every $\epsilon > 0$, there is $\ell \in \mathcal{W}_d$, such that*

$$\sup_{\omega \in \mathbf{X}^{-1}(\mathcal{K})} \left| F(\mathbf{X}(\omega)) - \left\langle \ell \,, \mathbb{X}_{0,T}^{<\infty}(\omega) \right\rangle \right| ,$$

interpreting the rough path lift as a random variable $\mathbf{X} : \Omega \to \widehat{\mathscr{C}}_g^\alpha([0,T];V)$.

We also refer to [CPSF22, Corollary 3.8] for a more general version applicable to jump-semimartingales.

2.5 Dynamic approximations with signatures

Theorem 2.9 shows that a function $f(x|_{[0,T]})$ of a path[3] defined on $[0,T]$ can be approximated by linear functionals of the path's signature $\left\langle \ell \,, \mathbb{x}_{0,T}^{<\infty} \right\rangle$ on the interval $[0,T]$. Suppose that we are instead given a path-valued functional, i.e., $y(t) = f(t, x|_{[0,t]})$ – think of a stochastic process y adapted to the filtration generated by x. Under the required regularity conditions, Theorem 2.9 immediately implies that we can find approximations

$$y(t) = f(t, x|_{[0,t]}) \approx \left\langle \ell(t) \,, \mathbb{x}_{0,t}^{<\infty} \right\rangle$$

for suitable linear functionals $\ell(t)$. Can we find *uniform approximations* in the sense that

$$y(t) = f(t, x|_{[0,t]}) \approx \left\langle \ell \,, \mathbb{x}_{0,t}^{<\infty} \right\rangle$$

with a linear functional ℓ independent of t? On a computational side, this seems plausible, since $\mathbb{x}_{0,t}^{<\infty}$ contains all monomials in t when $x_t^1 \equiv t$.

To formalise the problem, we introduce the space of *stopped rough paths*, going back to [Dup19], see also, [KLPA20; Bay+23a], for a more didactic presentation we also refer to [BHR24].

Indeed, consider $\mathcal{C}_T := \bigcup_{t\in[0,T]} C([0,t];V)$ understood as a disjoint union. Note that an element $x|_{[0,t]} \in \mathcal{C}_T$ is the restriction of a continuous path defined on the interval $[0,T]$ to a sub-interval $[0,t]$, $t \le T$. A metric is defined on $\mathcal{C}_T$ setting

$$d_{\mathcal{C}_T}(x|_{[0,t]}, y|_{[0,s]}) := \sup_{u\in[0,T]} |x_{t\wedge u} - y_{s\wedge u}| + |t-s| ,$$

[3] For simplicity, we will consider a smooth path x for the time being, but the discussion is equally valid for the rough case.

and it turns out that $\mathcal{C}_T$ equipped with this metric is a Polish space.

We will use this framework to study dynamic functional approximations, i.e., we want so approximate time dependent functions of a path of the form $f(t, x|_{[0,t]})$, $0 \le t \le T$. Note that we can easily understand such a function f as a function $F : \mathcal{C}_T \to \mathbb{R}$. The natural question is now whether such functions of restrictions of paths can also be approximated by linear functionals of the signature.

Of course, the space $\mathcal{C}_T$ is too large to allow for such approximations. As before, we will restrict ourselves to a proper rough path version thereof. Let, for fixed α, denote

$$\Lambda_T^\alpha := \bigcup_{t\in[0,T]} \widehat{\mathscr{C}}_g^\alpha([0,t];V). \tag{2.12}$$

(We restrict ourselves to the time-extended case from the beginning because any such universal approximation result will anyway require us to add time as a component.) Using a similar definition (see [BPS23]), we can define a rough path metric d on Λ_T^α, under which it is a Polish space. In this framework, repeating the arguments used in the proof of Theorem 2.9 gives (see [BHR24] for more details)

Corollary 2.12. *Let $\mathcal{K} \subset \Lambda_T^\alpha$ be compact and $F \in C(\Lambda_T^\alpha; \mathbb{R})$, then for every $\epsilon > 0$ there is an $\ell \in \mathcal{W}_d$ such that*

$$\sup_{\mathbf{x}|_{[0,t]}\in\mathcal{K}} \left| F(\mathbf{x}|_{[0,t]}) - \left\langle \ell, \mathbb{x}_{0,t}^{\le\infty} \right\rangle \right| < \epsilon.$$

The analogue result for lifts $\mathbf{X}$ of semimartingales also holds a.s.

Remark 2.13. The above universal approximation theorems – as typical in the area – are formulated in terms of uniform convergence on compacts. We admit that this concept is problematic, especially in infinite dimensions, as compact sets are often very small. Note that when we consider stochastic processes, e.g., the lift $\mathbf{X}|_{[0,t]}$ of a semimartingale taking values in Λ_T^α (a Polish space), *tightness* implies that for every $\delta > 0$ we can find a compact subset $[K] \subset \Lambda_T^\alpha$ such that $P(\mathbf{X} \notin \mathcal{K}) < \delta$, which allows us to replace "ϵ-close on compact sets" by "ϵ-close with probability $1 - \delta$".

More general global universal approximation results can be formulated in terms of *weighted spaces* (see [CST23]) or in terms of L^p-norms by using so-called "robust signatures" (see [CO22] for the definition of robust signatures and [BPS23; SA23] for the universality in L^p.)

2.6 Example: Signature models in finance

As motivated by the universality of the signature, we consider the problem of approximating a fixed linear functional ℓ of the signature $\mathbb{x}_{0,T}^{\le m}$ of a time-augmented smooth path, or, alternatively, a fixed linear functional ℓ of the signature $\mathbb{X}_{0,T}^{\le m}$ of a (time-extended) continuous semimartingale, by a linear functional of an alternative path $\widetilde{\mathbb{x}}$ or $\widetilde{\mathbb{X}}$ taking values in a space $\mathbb{R}^{\widetilde{n}}$ with $\widetilde{n} \ll \dim T^m(V)$.

As guiding example, we consider *signature models* for financial markets as introduced in [CGSF22]. Consider a d-dimensional underlying base process X, which is assumed to be a (time-extended) continuous semimartingale. Given a linear functional $\ell \in \mathcal{W}_d$, we then consider an asset price process given by

$$S_t = S_t^{(\ell)} := \left\langle \ell, \mathbb{X}_{0,t}^{\le m} \right\rangle, \tag{2.13}$$

where the truncation degree m is chosen high enough that all words in ℓ of degree larger than m have zero coefficient. Note that conditions on ℓ can be formulated such that the resulting price process satisfies fundamental requirements of mathematical finance such as no-arbitrage.[4] Universality of

[4] As such conditions are usually easier to formulate w.r.t. Itô rather than Stratonovich form, those conditions may look simpler in terms of the associated Itô signature.

signatures implies that signature models form a very flexible class of models, capable of approximating many desirable properties of asset price models. In addition, even though the price process (2.13) is not a Markov process, efficient numerical routines for option pricing exist in some cases.[5] As explained in [CGSF22], the model parameter ℓ can be efficiently estimated either from time series of asset prices, or from option price data. However, the dimension of the "state space" $T^m(V)$ of (2.13) is generally very large. Indeed, the dimension (in this sense) of the numerical examples in [CGSF22] are generally of order 100, but more complex markets might easily lead to much higher dimensional approximations.

In this paper, we assume that we are given a calibrated signature model $S = S^{(\ell)}$ for a fixed linear functional ℓ of degree m. The purpose is to derive reduced order models, i.e., a linear SDE with solution $\widetilde{\mathbb{X}}$ in $\mathbb{R}^{\widetilde{n}}$ and a linear functional $\widetilde{\ell} : \mathbb{R}^{\widetilde{n}} \to \mathbb{R}$ such that

$$\left\langle \widetilde{\ell}, \widetilde{\mathbb{X}} \right\rangle \approx S^{(\ell)} \quad \text{and} \quad \widetilde{n} \ll \dim T^m(V). \tag{2.14}$$

Note that the requirement of $\widetilde{\mathbb{X}}$ to solve a linear SDE is natural, given that already $\mathbb{X}^{\leq m}$ solves a linear SDE, see (2.6).

2.7 Signature differential equation

In the following, we will concentrate on the semi-martingale setting, but note that many discussions are equally valid in the (deterministic or stochastic) rough path framework (or even the bounded variation setting). As seen in (2.6) above, the truncated signature of the semi-martingale X satisfies

$$\mathrm{d}\mathbb{X}_{0,t}^{\leq m} = \mathbb{X}_{0,t}^{\leq m} \otimes \circ \mathrm{d}X_s = \sum_{i=1}^{d} \mathbb{X}_{0,t}^{\leq m} \otimes e_i \circ \mathrm{d}X_t^i, \quad \mathbb{X}_{0,0}^{\leq m} = 1 \in T^m(V), \tag{2.15}$$

where the tensor product $\otimes$ is truncated to $T^m(V)$. Since (2.15) is a linear SDE (driven by a semi-martingale), it fits within the context of stochastic model order reduction, see, for instance, [Red22; RBG21]. For concreteness, we will rewrite the SDE in terms of coordinates. We will use the canonical coordinates

$$\underbrace{1}_{\in V^{\otimes 0} \subset T^m(V)}, \underbrace{e_1, \dots, e_d}_{\in V \subset T^m(V)}, \underbrace{e_1 \otimes e_1, \dots e_d \otimes e_d}_{\in V^{\otimes 2} \subset T^m(V)}, \dots, \underbrace{e_1^{\otimes m}, \dots, e_d^{\otimes m}}_{\in V^{\otimes m} \subset T^m(V)},$$

which we order by length first and lexicographically within elements of the same length. The linear vector fields $V_i : T^m(V) \to T^m(V)$, $\mathfrak{a} \mapsto \mathfrak{a} \otimes e_i$ (in the sense of the truncated tensor product), can then be represented by matrices $N_i \in \mathbb{R}^{n \times n}$, where

$$n = n(d, m) = \sum_{i=0}^{m} d^i = \frac{d^{m+1} - 1}{d - 1} = \dim T^m(V).$$

The matrices $N_1, \dots, N_d$ can be given in closed form, see Box 1.

We end up with the SDE

$$\mathrm{d}\mathbb{X}_{0,t}^{\leq m} = \sum_{i=1}^{d} N_i \mathbb{X}_{0,t}^{\leq m} \circ \mathrm{d}X_t^i, \quad \mathbb{X}_{0,0}^{\leq m} = (1, 0, \dots, 0)^\top \in \mathbb{R}^n, \tag{2.16}$$

where we use the same symbol $\mathbb{X}_{0,t}^{\leq m} \in \mathbb{R}^n$ for the version of the signature in coordinates.

[5] Essentially, when the payoff function can be efficiently approximated by a linear functional of the signature of S.

Box 1 Matrices representing the vector fields driving the truncated signature.

Initialize matrices $N_1 = \cdots = N_d = 0$
for $i = 1, \dots, d$ **do**
  for $k = 1, 2, \dots, (d^m - 1)/(d - 1)$ **do**
    $N_i[1 + (k-1)d + i, k] = 1$
  end for
end for

Remark 2.14. Note that the matrices $N_1, \dots, N_d$ are *nilpotent* with order $m + 1$, i.e., for any $i_1, \dots, i_m, i_{m+1} \in \{1, \dots, d\}$ and $1 \le k \le m$ we have

$$N_{i_1} \cdot N_{i_2} \cdots N_{i_k} \neq 0, \text{ but } N_{i_1} \cdot N_{i_2} \cdots N_{i_m} \cdot N_{i_{m+1}} = 0.$$

This is clear from the construction: applying the vector field V_i to an element $\mathfrak{a} \in V^{\otimes n} \subset T^m(V)$ yields an element $V_i(\mathfrak{a}) \in V^{\otimes(n+1)} \subset T^m(V)$, for any $0 \le n < m$. Hence, iteratively applying the vector fields $m + 1$ times always gives the zero element, as we are "pushed out" of the truncated tensor algebra.

3 Dimension reduction

As the first step in the analysis, we rewrite (2.16) as an Itô stochastic differential equation. In order to simplify the dimension reduction, we will impose assumptions assuring that the Itô version is still linear.

Assumption 3.1. X_t is a time-extended Brownian motion, i.e., $X_t^1 = t$ and $B_t \coloneqq (B_t^2, \dots, B_t^d)^\top \coloneqq (X_t^2, \dots, X_t^d)^\top$ is a $(d-1)$-dimensional Brownian motion with associated matrix $K = (k_{ij})_{i,j=2,\dots,d}$ that determines the covariance by $\mathbb{E}[B_t B_t^\top] = Kt$.

Remark 3.2. Alternatively, we can also consider a general semi-martingale, provided that the quadratic variation processes are also part of X. The driving process X is a semi-martingale such that all quadratic co-variations $[X^i, X^j]_t$ are already (constant multiples) of coordinates of X. More specifically, $X_t = (\bar{X}_t, \widehat{X}_t)$, where $\bar{X}$ is a (proper) semimartingale and $\widehat{X}$ is a bounded variation process such that for any i, j there is a constant k_{ij} and an index c_{ij} such that $[\bar{X}^i, \bar{X}^j] \equiv k_{ij} \widehat{X}^{c_{ij}}$. Note that Assumption 3.1 is a special case of this semi-martingale setting. For simplicity, we will concentrate on the former case in what follows and leave the situation of this remark for future work.

We aim to rewrite (2.16) in the Itô sense. The associated Itô-Stratonovic correction term is $0.5 \sum_{i,j=2}^d N_i N_j k_{ij}$, where k_{ij} is the ij-th entry of the matrix K determining the covariance of the Brownian motion. The drift coefficient is $A = N_1 + 0.5 \sum_{i,j=2}^d N_i N_j k_{ij}$ leading to

$$\mathrm{d}\mathcal{X}_t = A\mathcal{X}_t \mathrm{d}t + \sum_{i=2}^d N_i \mathcal{X}_t \mathrm{d}B_t^i, \quad \mathcal{X}_0 = z, \quad \mathcal{Y}_t = L\mathcal{X}_t, \quad t \in [0, T], \tag{3.1}$$

where $z \in \mathbb{R}^n$ is a generic initial state and the rows of $L \in \mathbb{R}^{p \times n}$ represent linear functionals of interest. Let us recall that we are particularly interested in $z = \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix}^\top$ giving us $\mathcal{X}_t = \mathbb{X}_{0,t}^{\le m}$.

3.1 Identifying less relevant signature information

We introduce the fundamental solution to the signature equation (3.1) as the process $\Phi(t,t_0)$ satisfying

$$\mathrm{d}\Phi(t,t_0) = A\Phi(t,t_0)\mathrm{d}t + \sum_{i=2}^{d} N_i\Phi(t,t_0)\mathrm{d}B_t^i, \quad \Phi(t_0,t_0) = I \in \mathbb{R}^{n\times n}, \tag{3.2}$$

with $t_0 \in [0,T)$. Therefore, we can write $\mathcal{X}_t = \Phi(t,t_0)\mathcal{X}_{t_0}$ for all $t \in [t_0,T]$. Based on that, we set

$$P := \int_0^T \mathbb{E}\left[\mathcal{X}_u\mathcal{X}_u^\top\right]\mathrm{d}u = \int_0^T \mathbb{E}\left[\Phi(u,0)zz^\top\Phi(u,0)^\top\right]\mathrm{d}u, \tag{3.3}$$

$$Q := \int_0^T \mathbb{E}\left[\Phi(u,0)^\top L^\top L\Phi(u,0)\right]\mathrm{d}u. \tag{3.4}$$

We can choose an orthonormal basis (ONB) consisting of eigenvectors of these time-averaged covariances to represent the truncated signature. Let $(p_k)_{k=1,\dots,n}$ be an ONB for $\mathbb{R}^n$ of eigenvectors of P and $(q_k)_{k=1,\dots,n}$ the one associated to Q. This yields the following

$$\begin{aligned} \mathcal{X}_t &= \sum_{k=1}^{n} \langle \mathcal{X}_t, p_k\rangle_{\mathbb{R}^n}\, p_k, \quad t \in [0,T], \\ \mathcal{Y}_t &= L\Phi(t,t_0)\mathcal{X}_{t_0} = \sum_{k=1}^{n} \langle \mathcal{X}_{t_0}, q_k\rangle_{\mathbb{R}^n}\, L\Phi(t,t_0)q_k, \quad t \in [t_0,T]. \end{aligned} \tag{3.5}$$

The next proposition tells us which directions p_k and q_k are of minor relevance.

Proposition 3.3. *Let $(\lambda_k)_{k=1,\dots,n}$ and $(\mu_k)_{k=1,\dots,n}$ be the eigenvalues of the ONB $(p_k)_{k=1,\dots,n}$ and $(q_k)_{k=1,\dots,n}$, respectively. Then, we obtain for the k-th summands of* (3.5) *that*

$$\begin{aligned} \mathbb{E}\int_0^T \langle \mathcal{X}_u, p_k\rangle_{\mathbb{R}^n}^2 \mathrm{d}u &= \lambda_k, \\ \mathbb{E}\int_{t_0}^T \langle \mathcal{X}_{t_0}, q_k\rangle_{\mathbb{R}^n}^2 \left\|L\Phi(u,t_0)q_k\right\|_{\mathbb{R}^p}^2 \mathrm{d}u &\le \mu_k \mathbb{E}\langle \mathcal{X}_{t_0}, q_k\rangle_{\mathbb{R}^n}^2. \end{aligned}$$

Proof. The first identity is trivial as $\mathbb{E}\int_0^T \langle \mathcal{X}_u, p_k\rangle_{\mathbb{R}^n}^2 \mathrm{d}u = p_k^\top \int_0^T \mathbb{E}[\mathcal{X}_u\mathcal{X}_u^\top]\mathrm{d}u\, p_k = p_k^\top P p_k = \lambda_k$ exploiting that p_k has norm 1. We further note that $\mathcal{X}_{t_0}$ is $\mathcal{F}_{t_0}$-measurable and that $u \mapsto \Phi(u,t_0)$ only depends on increments of B after t_0. Therefore, they are independent and hence $\mathbb{E}[\langle \mathcal{X}_{t_0}, q_k\rangle_{\mathbb{R}^n}^2 \left\|L\Phi(u,t_0)q_k\right\|_{\mathbb{R}^p}^2] = \mathbb{E}\langle \mathcal{X}_{t_0}, q_k\rangle_{\mathbb{R}^n}^2 \mathbb{E}\left\|L\Phi(u,t_0)q_k\right\|_{\mathbb{R}^p}^2$. Now, we obtain by substitution that

$$\begin{aligned} \mathbb{E}\int_{t_0}^T \left\|L\Phi(u,t_0)q_k\right\|_{\mathbb{R}^p}^2 \mathrm{d}u &= \mathbb{E}\int_{t_0}^T \left\|L\Phi(u-t_0,0)q_k\right\|_{\mathbb{R}^p}^2 \mathrm{d}u \\ &= \int_0^{T-t_0} \mathbb{E}\left\|L\Phi(u,0)q_k\right\|_{\mathbb{R}^p}^2 \mathrm{d}u \le q_k Q q_k = \mu_k. \end{aligned}$$

Above, we exploited that $\mathbb{E}\big[\Phi(u,t_0)q_kq_k^\top\Phi(u,t_0)^\top\big] = \mathbb{E}\big[\Phi(u-t_0,0)q_kq_k^\top\Phi(u-t_0,0)^\top\big]$ as both expressions satisfy (3.6) with $M = q_kq_k^\top$. This concludes the proof. □

Proposition 3.3 tell us that p_k has a low contribution to $\mathcal{X}_t$ if the associated eigenvalue λ_k is small. Moreover, neglecting q_k in $\mathcal{X}_{t_0}$ has a minor impact on the quantity of interest after t_0 if μ_k is small unless $\mathcal{X}_{t_0}$ is very large in the direction q_k. Consequently, we know that we can remove directions p_k and q_k that correspond to small eigenvalues. In order to be able to do this simultaneously, we construct a coordinate transform below that ensures that $p_k = q_k = \mathrm{e}_k$. Here, e_k is k-th canonical basis vector of $\mathbb{R}^n$. Therefore, unimportant direction can be identified with components of a transformed signature. Before stressing this aspect further, a strategy for the computation of P and Q is provided in the following section.

3.2 Computation of P and Q

The later dimension reduction procedure relies on having P and Q available. For that reason, let us briefly discuss how these matrices can be computed in practice.

Let us introduce the Lyapunov operator $\mathcal{L}(Z) = AZ + ZA^\top + \sum_{i,j=2}^d N_i Z N_j^\top k_{ij}$, where Z is a matrix of suitable dimension. This notion is used to recall the following well-known result.

Proposition 3.4. *Given $M \in \mathbb{R}^{n\times n}$, the function $t \mapsto \mathbb{E}[\Phi(t,t_0)M\Phi(t,t_0)^\top]$ solves*

$$\frac{\mathrm{d}}{\mathrm{d}t} Z_t = \mathcal{L}(Z_t), \quad Z_{t_0} = M, \quad t \geq t_0. \tag{3.6}$$

Proof. We define $\Phi_t := \Phi(t,t_0)$ and make use of Itô's product rule yielding

$$\begin{aligned}
\mathrm{d}(\Phi_t M \Phi_t^\top) &= \mathrm{d}(\Phi_t) M \Phi_t^\top + \Phi_t M \mathrm{d}(\Phi_t^\top) + \mathrm{d}(\Phi_t) M \mathrm{d}(\Phi_t^\top) \\
&= \Big(A\Phi_t \mathrm{d}t + \sum_{i=2}^d N_i \Phi_t \mathrm{d}B_t^i\Big) M \Phi_t^\top + \Phi_t M \Big(\Phi_t^\top A^\top \mathrm{d}t + \sum_{i=2}^d \Phi_t^\top N_i^\top \mathrm{d}B_t^i\Big) \\
&\quad + \sum_{i,j=2}^d N_i \Phi_t M \Phi_t^\top N_j^\top k_{ij} \mathrm{d}t.
\end{aligned}$$

Taking the expectation, we obtain $\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\Phi_t M \Phi_t^\top] = \mathcal{L}\left(\mathbb{E}[\Phi_t M \Phi_t^\top]\right)$. □

Let $\mathrm{vec}(\cdot)$ be the vectorization of a matrix. Using its relation to the Kronecker product $\tilde{\otimes}$ between two matrices, (3.6) becomes

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{vec}(Z_t) = \mathcal{K}\,\mathrm{vec}(Z_t), \quad \mathrm{vec}(Z_{t_0}) = \mathrm{vec}(M), \quad t \geq t_0, \tag{3.7}$$

where the matrix representation of the Lyapunov operator is

$$\mathcal{K} = A \,\tilde{\otimes}\, I + I \,\tilde{\otimes}\, A + \sum_{i,j=2}^d N_i \,\tilde{\otimes}\, N_j k_{ij}.$$

As P defined in (3.3) is the integral of the solution of (3.6) with $t_0 = 0$ and $M = zz^\top$, we therefore have that

$$\mathrm{vec}(P) = \int_0^T \mathrm{vec}(Z_t)\mathrm{d}t = \int_0^T \mathrm{e}^{\mathcal{K}t}\,\mathrm{vec}(zz^\top)\mathrm{d}t$$

inserting the solution representation for (3.7). According to Remark 2.14, the matrices $N_1, \ldots, N_d$ are nilpotent with order $m+1$ leading to $\mathcal{K}^j = 0$ for $j \geq 2m+1$ and hence $\mathrm{e}^{\mathcal{K}t} = \sum_{j=0}^{2m} \frac{t^j}{j!}\mathcal{K}^j$. Exploiting this fact yields

$$\mathrm{vec}(P) = \sum_{j=0}^{2m} \frac{T^{j+1}}{(j+1)!}\mathcal{K}^j\,\mathrm{vec}(zz^\top). \tag{3.8}$$

Although $\mathcal{K}$ potentially is a huge matrix, P can be computed from (3.8) since $\mathcal{K}$ is extremely sparse making the matrix-vector multiplication cheap. In order to compute P for very large n, it is beneficial to devectorize (3.8) leading to the explicit representation

$$P = \sum_{j=0}^{2m} \frac{T^{j+1}}{(j+1)!}\mathcal{L}^j(zz^\top). \tag{3.9}$$

The computation of Q can be conducted using similar arguments. Vectorizing (3.4) yields

$$\begin{aligned}\operatorname{vec}(Q) &= \int_0^T \mathbb{E}\Big[\operatorname{vec}\Big(\Phi(u,0)^\top L^\top L\Phi(u,0)\Big)\Big]\mathrm{d}u\\ &= \int_0^T \mathbb{E}\Big[\Phi(u,0)^\top \tilde{\otimes}\, \Phi(u,0)^\top\Big]\operatorname{vec}\Big(L^\top L\Big)\mathrm{d}u.\end{aligned}$$

Based on Proposition 3.4 and (3.7), we know that

$$\mathrm{e}^{\mathcal{K}(t-t_0)}\operatorname{vec}(M) = \operatorname{vec}\Big(\mathbb{E}\Big[\Phi(t,t_0)M\Phi(t,t_0)^\top\Big]\Big) = \mathbb{E}\Big[\Phi(t,t_0)\,\tilde{\otimes}\,\Phi(t,t_0)\Big]\operatorname{vec}(M).$$

Since this identity is true for all matrices M, we find that $\mathbb{E}\Big[\Phi(t,t_0)\,\tilde{\otimes}\,\Phi(t,t_0)\Big] = \mathrm{e}^{\mathcal{K}(t-t_0)}$ and consequently

$$\operatorname{vec}(Q) = \int_0^T \mathrm{e}^{\mathcal{K}^\top u}\operatorname{vec}\Big(L^\top L\Big)\mathrm{d}u = \sum_{j=0}^{2m}\frac{T^{j+1}}{(j+1)!}(\mathcal{K}^\top)^j\operatorname{vec}(L^\top L) \tag{3.10}$$

using once more that $\mathcal{K}^j = 0$ for $j \geq 2m+1$. Relation (3.10) now is the basis for the computation of Q. Equivalently, we can write

$$Q = \sum_{j=0}^{2m}\frac{T^{j+1}}{(j+1)!}(\mathcal{L}^*)^j(L^\top L), \tag{3.11}$$

where the Lyapunov operator's adjoint w.r.t. the Frobenius inner product is

$$\mathcal{L}^*(Z) = A^\top Z + ZA + \sum_{i,j=2}^{d} N_i^\top Z N_j k_{ij}.$$

3.3 Reduced order signature approximation

We conduct a state space transformation via a nonsingular matrix $\mathcal{T} \in \mathbb{R}^{n\times n}$. To do so, we define $\hat{\mathcal{X}}_t = \mathcal{T}\mathcal{X}_t$ and and associated coefficients $(\hat{A}, \hat{N}_i, \hat{L}) = (\mathcal{T}A\mathcal{T}^{-1}, \mathcal{T}N_i\mathcal{T}^{-1}, L\mathcal{T}^{-1})$. The purpose of this transformation is that an equivalent system (same quantity of interest) is supposed to be obtained, in which the redundant information can be removed easily by truncation of state components. In particular, $\mathcal{T}$ is chosen in a way that it simultaneously diagonalizes P and Q. Hence, the corresponding eigenvectors (p_k) and (q_k) are the canonical basis (e_k) of $\mathbb{R}^n$ at the same time. Exploiting Proposition 3.3, we can then identify unimportant components $\langle\hat{\mathcal{X}}., \mathrm{e}_k\rangle_{\mathbb{R}^n}$ of $\hat{\mathcal{X}}$ and truncate those. Let us refer to the discussion below this proposition once more and provide further details below.

First, we can conclude that the modified signature $\hat{\mathcal{X}}$ fulfills

$$\mathrm{d}\hat{\mathcal{X}}_t = \hat{A}\hat{\mathcal{X}}_t\mathrm{d}t + \sum_{i=2}^{d}\hat{N}_i\hat{\mathcal{X}}_t\mathrm{d}B_t^i, \quad \hat{\mathcal{X}}_0 = \mathcal{T}z, \quad \mathcal{Y}_t = \hat{L}\hat{\mathcal{X}}_t, \quad t\in[0,T], \tag{3.12}$$

meaning that $\mathcal{Y}$ is invariant under this transformation. Now, we see directly from (3.2) that the fundamental solution of (3.12) is $\hat{\Phi}(t,t_0) = \mathcal{T}\Phi(t,t_0)\mathcal{T}^{-1}$. Consequently, the time-averaged covariances of (3.12) are

$$\begin{aligned}\hat{P} &:= \int_0^T \mathbb{E}\left[\hat{\Phi}(u,0)\mathcal{T}z(\mathcal{T}z)^\top\hat{\Phi}(u,0)^\top\right]\mathrm{d}u = \mathcal{T}P\mathcal{T}^\top,\\ \hat{Q} &:= \int_0^T \mathbb{E}\left[\hat{\Phi}(u,0)^\top\hat{L}^\top\hat{L}\hat{\Phi}(u,0)\right]\mathrm{d}u = \mathcal{T}^{-\top}Q\mathcal{T}^{-1}.\end{aligned} \tag{3.13}$$

The following proposition states the particular transformation required for a simultaneous diagonalization of the time-averaged covariances.

Proposition 3.5. *Given that P and Q are positive definite, we obtain $\hat{P} = \hat{Q} = \Sigma = \operatorname{diag}(\sigma_1, \dots, \sigma_n)$ using the balancing transformation*

$$\mathcal{T} = \Sigma^{\frac{1}{2}} U^\top L_P^{-1}, \tag{3.14}$$

with the factorization $P = L_P L_P^\top$ and the spectral decomposition $L_P^\top Q L_P = U \Sigma^2 U^\top$, where U is orthogonal. Moreover, σ_i^2 are the eigenvalues of PQ.

Proof. Inserting (3.14) in (3.13) yields $\hat{P} = \Sigma^{\frac{1}{2}} U^\top L_P^{-1} P L_P^{-\top} U \Sigma^{\frac{1}{2}} = \Sigma$ and $\hat{Q} = \Sigma^{-\frac{1}{2}} U^\top L_P^\top Q L_P U \Sigma^{-\frac{1}{2}} = \Sigma$. By definition, σ_i^2 are the eigenvalues of $L_P^\top Q L_P$. Exploiting that $L_P^\top Q L_P$ has the same spectrum like $L_P L_P^\top Q = PQ$ concludes the proof. □

Given that we have used the balancing transformation (3.14) in (3.12), we can identify less relevant directions with components of $\hat{\mathcal{X}}$ that are associated to small eigenvalues σ_k^2 of PQ. This is a consequence of Proposition 3.3, where the transformation ensures that $p_k = q_k$ is the k-th column of the $n \times n$ identity matrix and $\lambda_k = \mu_k = \sigma_k$. For that reason, the spectrum of PQ delivers a good truncation criterion and therefore the intuition on how to fix the reduced dimension $\widetilde{n} \ll n$. In order to find the reduced equation, we partition the solution of (3.12) with $\mathcal{T}$ like in (3.14) as follows $\hat{\mathcal{X}}_t = \begin{bmatrix} \hat{\mathcal{X}}_t^1 \\ \hat{\mathcal{X}}_t^2 \end{bmatrix}$. $\hat{\mathcal{X}}_t^1$ taking values in $\mathbb{R}^{\widetilde{n}}$ corresponds to the large values $\sigma_1, \dots, \sigma_{\widetilde{n}}$ and $\hat{\mathcal{X}}_t^2$ to the small values $\sigma_{\widetilde{n}+1}, \dots, \sigma_n$ in the sense that $\sigma_{\widetilde{n}+1} \ll \sigma_{\widetilde{n}}$. The respective partition of the balanced coefficients

$$\hat{A} = \begin{bmatrix} \widetilde{A} & \star \\ \star & \star \end{bmatrix}, \quad \mathcal{T} z = \begin{bmatrix} \widetilde{z} \\ \star \end{bmatrix}, \quad \hat{N}_i = \begin{bmatrix} \widetilde{N}_i & \star \\ \star & \star \end{bmatrix}, \quad \hat{L} = \begin{bmatrix} \widetilde{L} & \star \end{bmatrix} \tag{3.15}$$

with $\widetilde{A}, \widetilde{N}_i \in \mathbb{R}^{\widetilde{n} \times \widetilde{n}}$, $\widetilde{z} \in \mathbb{R}^{\widetilde{n}}$ and $\widetilde{L} \in \mathbb{R}^{p \times \widetilde{n}}$ leads to the reduced system

$$\mathrm{d}\widetilde{\mathcal{X}}_t = \widetilde{A} \widetilde{\mathcal{X}}_t \mathrm{d}t + \sum_{i=2}^{d} \widetilde{N}_i \widetilde{\mathcal{X}}_t \mathrm{d}B_t^i, \quad \widetilde{\mathcal{X}}_0 = \widetilde{z}, \quad \widetilde{\mathcal{Y}}_t = \widetilde{L} \widetilde{\mathcal{X}}_t, \quad t \in [0, T]. \tag{3.16}$$

In detail, the reduced model (3.16) is obtained by removing the equation for $\hat{\mathcal{X}}_t^2$ in (3.12) and by setting $\hat{\mathcal{X}}_t^2 = 0$ in the dynamics of $\hat{\mathcal{X}}_t^1$. This is motivated by th minor relevance of $\hat{\mathcal{X}}_t^2$ leading to a reduced output $\widetilde{\mathcal{Y}} \approx \mathcal{Y}$. In particular, choosing $p = 1$ and $z = \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix}^\top$, the reduced variable $\widetilde{\mathcal{X}}$ is a candidate for $\widetilde{\mathbb{X}}$ in (2.14).

4 Numerical examples

4.1 Bergomi model

In the first example, we use a standard stochastic volatility model, namely the *Bergomi model* [Ber15], as our starting point. The Bergomi model is highly praised for its flexibility, and the ability to accurately fit equity markets, even with relatively few parameters. More concretely, the n-factor Bergomi model is given by

$$\mathrm{d}S_t = \sqrt{\xi_t^t} S_t \mathrm{d}Z_t, \tag{4.1a}$$

$$\mathrm{d}_t \xi_t^T = \frac{\omega}{\sqrt{\sum_{i,j=1}^n w_i w_j \rho_{ij}}} \xi_t^T \sum_{i=1}^n w_i \mathrm{e}^{-k_i (T-t)} \mathrm{d}W_t^i, \tag{4.1b}$$

with initial values S_0 for the stock price and an initial forward variance curve $\xi_0^\cdot$. Here, $Z, W^1, \dots, W^n$ are correlated standard Brownian motions, where ρ denotes the correlation matrix of $(W^1, \dots, W^n)$ and – following [Guy22] – we denote the correlation between the Brownian motion Z driving the stock price and the Brownian motion W^i by ρ_{Si}.

Remark 4.1. As indicated above, the Bergomi model is a *forward variance model*, i.e., it models the whole forward variance curve $\xi_t^T := \mathbb{E}[v_T \mid \mathcal{F}_t]$, $T \geq t \geq 0$, not just the instantaneous variance $v_t = \xi_t^t$. This yields great flexibility for calibration, noting that the initial forward variance curve $\xi_0^\cdot$ – an essential parameter of the model – can be read out from market data (variance swaps or vanilla option prices using the *log-strip formula*), and, hence, does not need to be calibrated in principle. The specific dynamics of the forward variance in (4.1b) corresponds to taking the exponential of a weighted sum of Ornstein-Uhlenbeck processes.

In the first numerical example, we take a two-factor Bergomi model, following the parameterization of [Guy22, Section 3], which is a variant of the parameter Set I in [Ber15, p. 229] with constant initial forward variance curve. Specifically, we choose the parameters presented in Table 1. We here use the notation $\theta_1 = w_1$, and the convention $w_2 = 1 - \theta_1$.

ω	k_1	k_2	θ_1	ρ_{12}	ρ_{S1}	ρ_{S2}	S_0	$\xi_0^\cdot$
3	2.63	0.42	0.69	0.7	-0.9	-0.9	1	0.04

Table 1: Parameters for the Bergomi model used in our numerical example following [Guy22, Section 3].

In our first numerical example, we take the Bergomi model as ground-truth, and proceed to

1 train a signature model in the sense of [CGSF22] with different truncation levels m, i.e., a model for the asset price of the form $S_t \approx S_t^{(\ell)} = \left\langle \ell, \mathbb{X}_{0,t}^{\leq m} \right\rangle$, see Section 2.6;

2 vectorize the truncated signature $\mathbb{X}_{0,t}^{\leq m}$ and formulate Itô-SDE (3.1) with respective initial state, so that $S_t^{(\ell)} = \mathcal{Y}_t = L\mathcal{X}_t$;

3 approximate the truncated signature by the method described in Section 3.3; obtain a reduced order signature approximation $S_t^{(\ell)} \approx \widetilde{\mathcal{Y}}_t = \widetilde{L}\widetilde{\mathcal{X}}_t$ from (3.16).

In the specific example of the two-dimensional Bergomi model, we choose the underlying state process $X_t = (t, Z_t, W_t^1, W_t^2)$, i.e., we consider the truncated signature $\mathbb{X}_{0,t}^{\leq m}$ at level m of a d-dimensional process with $d = 4$ and $m = 5$. This results in an Itô-SDE (3.1) with state dimension $n = 1365$ with quantity of interest $\mathcal{Y}_t$. We apply the dimension reduction scheme of Section 3 to (3.1) in order to obtain a reduced system (3.16) of order $\widetilde{n} \ll n$ with quantity of interest $\widetilde{\mathcal{Y}}_t$. The values $\sigma_k := \sqrt{\operatorname{eig}_k(PQ)}$ introduced in Proposition 3.5 provide an algebraic criterion for a good choice of $\widetilde{n}$. According to Section 3.3, the smaller σ_k, the less important the associated state component in the balanced system (3.12) is. Figure 1 shows σ_k for $k \in \{1, \dots, 50\}$. We observe a strong decay in k and notice that $\sigma_{28} < 10^{-8}$ which is below the machine precision. Consequently, (3.1) has a high reduction potential and allows for an exact approximation in case of choosing $\widetilde{n} = 27$. The corresponding reduction errors $\sqrt{\mathbb{E}\int_0^T \left\|\mathcal{Y}_t - \widetilde{\mathcal{Y}}_t\right\|_{\mathbb{R}^p}^2 dt}$ can be found in Figure 2. We have a true approximation error for $\widetilde{n} < 27$. If $\widetilde{n} \geq 27$, the error is numerically zero. In addition to the L^2-error, the quality of the reduced signature system is tested in a finance context. First, we determine the fair price $\mathbb{E}\left[\max\{S_T^{(\ell)} - K_T, 0\}\right]$ of a European option, where the values of the strike price $K_T = (0.8 + j \cdot 0.02)^{\sqrt{T}}$ $(j = 0, 1, \dots, 20)$ are chosen depending on the maturity T. The computed prices are then treated as prices from a Black-Scholes model and the associated volatilities are

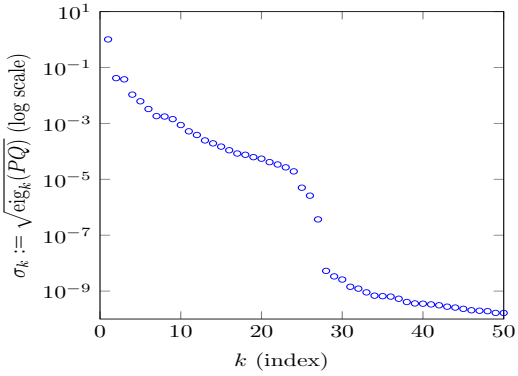

Figure 1: Square root of first 50 out of $n = 1365$ eigenvalues of PQ for signature model associated to (4.1).

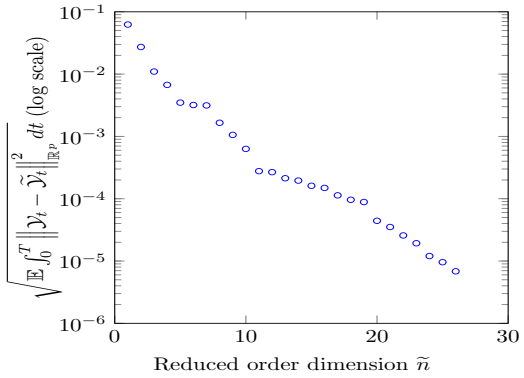

Figure 2: L^2-error between output of the signature model of (4.1) ($n = 1365$) and reduced system output for $\widetilde{n} = 1, \ldots, 26$.

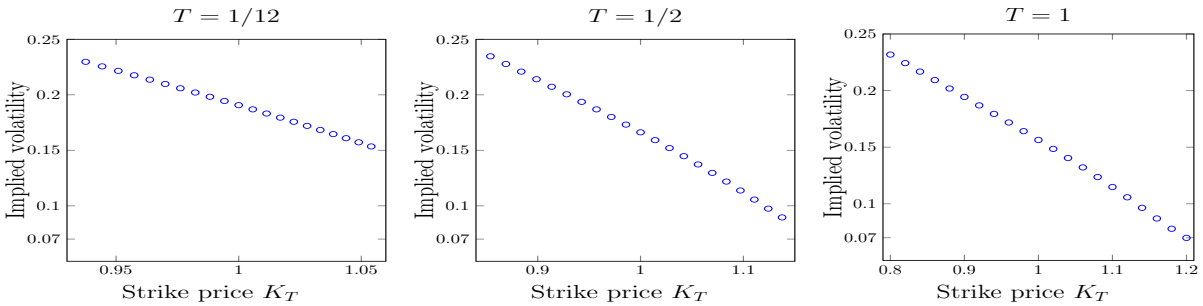

Figure 3: Implied volatilities of (4.1) for $T = 1/12, 1/2, 1$ and strike prices $K_T = (0.8 + j \cdot 0.02)^{\sqrt{T}}$ with $j = 0, 1, \ldots, 20$.

derived (while the interest rate is zero). These quantities are called implied volatilities (IV). The IV are depicted in Figure 3 for $T = 1/12, 1/2, 1$. We repeat this procedure in the reduced setting and obtain approximating IV. The relative error between the IV in the full and the reduced dynamics can be found in Figure 4 with reduced dimensions $\widetilde{n} = 5, 11, 27$ and terminal times $T = 1/12, 1/2, 1$. We can see that the error is around one percent for $T = 1/12, 1/2$ if we set $\widetilde{n} = 5$. However, the same reduced dimension shows errors of up to six percent for $T = 1$, since relatively small IV come with a higher relative approximation error. Therefore, it can be reasonable to enlarge the reduced dimension to $\widetilde{n} = 11$. This yields relative deviations in the IV of around 10^{-4} only and is hence a very good approximation regardless of the maturity T. As expected, we obtain IV errors that can be totally neglected when fixing $\widetilde{n} = 27$, see Figure 4 once more. This is a very significant reduction in comparison to the original dimension $n = 1365$.

4.2 Rough Bergomi model

Our second example is the *rough Bergomi model* introduced in [BFG16], a work-horse model within the class of *rough volatility models*, see [Bay+23b] for an exposition. Intuitively, the rough Bergomi model is a variant of the Bergomi model, where the exponential memory kernel in the variance process is replaced by a fractional kernel, in essence leading to the variance process being an exponential of a fractional Brownian motion. While seemingly innocuous, this change destroys the Markovian

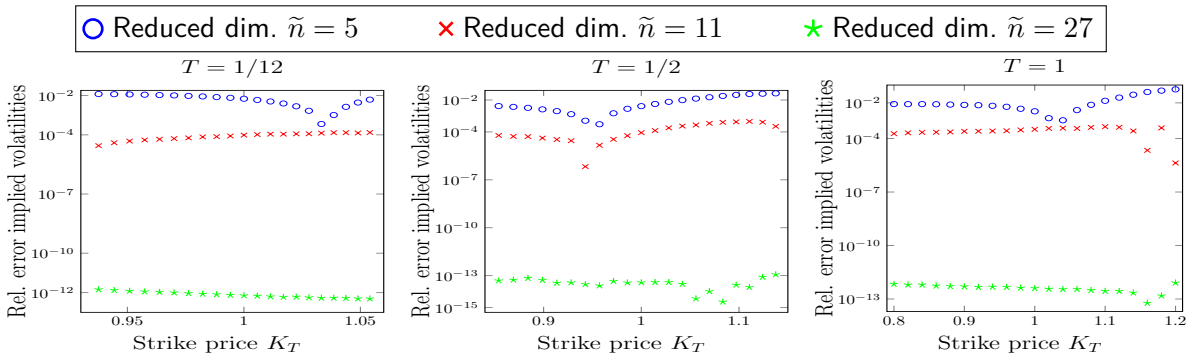

Figure 4: Relative error between volatilities of signature model associated to (4.1) and reduced systems with $\widetilde{n} = 5, 11, 27$, strike prices $K_T = (0.8 + j \cdot 0.02)^{\sqrt{T}}$ and $T = 1/12, 1/2, 1$.

structure of the resulting model. More precisely, we consider the model

$$\mathrm{d}S_t = \sqrt{v_t} S_t \mathrm{d}Z_t, \tag{4.2a}$$

$$v_t = \xi_0(t) \exp\left(\eta \widehat{W}_t^H - \frac{1}{2}\eta^2 t^{2H}\right), \tag{4.2b}$$

where $\widehat{W}_t^H := \sqrt{2H} \int_0^t (t-s)^{H-1/2} \mathrm{d}W_s$ denotes a *Riemann–Liouville fractional Brownian motion*, and W and Z are standard Brownian motions with correlation ρ. Note that the rough Bergomi model is – like the Bergomi model – a forward variance model, (depending on the initial forward variance curve ξ_0).

H	η	ρ	S_0	$\xi_0(\cdot)$
0.3	2.3	-0.9	1	0.04

Table 2: Parameters for the rough Bergomi model used in our numerical example.

The term "rough" in the *rough Bergomi model* reflects the empirical observation that the Hurst index H should be chosen less than $\frac{1}{2}$ – leading to a power law explosion of the ATM implied volatility skew. In fact, empirical studies show that H is often chosen very close to 0, e.g., $H = 0.07$ reported in [BFG16] based on a calibration on SPX option prices as of February 4, 2010.

We again fit a signature model with underlying state process $X_t = (t, Z_t, W_t)$ (i.e. $d = 3$) to a rough Bergomi model, taking model parameters reported in Table 2. The parameters are realistic, with the possible exception of our choice of $H = 0.3$. While clearly within the range of Hurst parameters observed in the large scale study [BLP22], it is comparatively large for rough volatility models. In our experience, fitting a signature model to a rough volatility model with very small H is, however, hard, especially while keeping the truncation degree manageable. Therefore, we choose $H = 0.3$ as a compromise of a clearly fractional, non-Markovian model which is still easily fitable by a signature model. Note that we again use a constant initial forward variance curve. In particular, we choose $m = 7$ as the level of the truncated signature $\mathbb{X}_{0,t}^{\leq m}$ leading to an equation in (3.1) with state dimension $n = 3280$. The reduction technique of Section 3 provides a reduced system (3.16) with state dimension $\widetilde{n} \ll n$ with the aim of having an accurate approximation of the quantity of interest, i.e., $\widetilde{\mathcal{Y}}_t \approx \mathcal{Y}_t$. The algebraic values $\sigma_k := \sqrt{\mathrm{eig}_k(PQ)}$ in Figure 5 tell us about the significance of state variables and hence the right reduced dimension $\widetilde{n}$. We observe that σ_k is numerically zero for $k > 55$. Therefore, a reduced system of order $\widetilde{n} = 55$ is an exact model. However, choosing $\widetilde{n}$ much smaller than this can go along with a little approximation error as well. We illustrate the L^2-performance for all case with a true error in Figure 6. We, e.g., observe that the L^2-error between $\widetilde{\mathcal{Y}}$ with $\widetilde{n} > 10$ and $\mathcal{Y}$ (linear functional of the signature) is below 0.01.

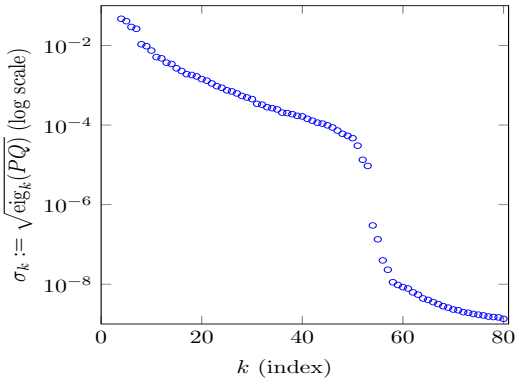

Figure 5: Square root of first 80 out of $n = 3280$ eigenvalues of PQ for signature model associated to (4.2).

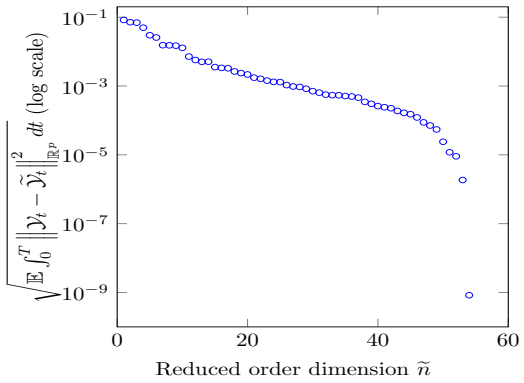

Figure 6: L^2-error between output of the signature model of (4.2) ($n = 3280$) and reduced system output for $\widetilde{n} = 1, \ldots, 54$.

Moreover, we are interested in the quality of the signature approximation (3.16) when IV are aimed to be reproduced. First of all, let us note the true IV of the signature model in Figure 7 for $T = 1/12, 1/2, 1$. We approximated these values by the IV of reduced system for $\widetilde{n} = 55$ and, as expected,

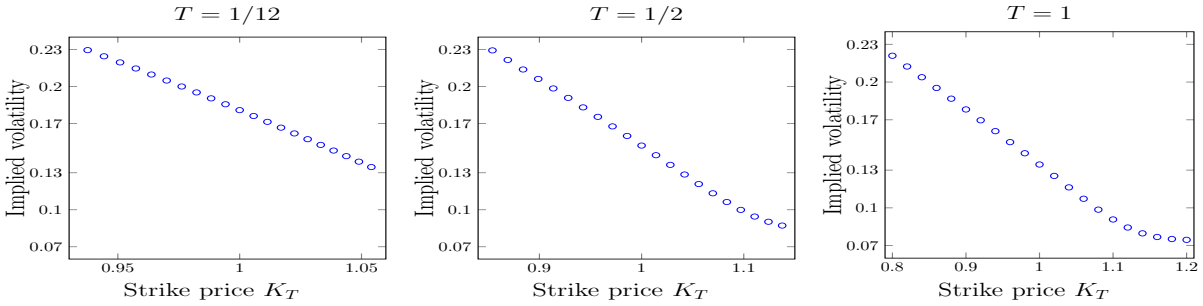

Figure 7: Implied volatilities of (4.2) for $T = 1/12, 1/2, 1$ and strike prices $K_T = (0.8 + j \cdot 0.02)^{\sqrt{T}}$ with $j = 0, 1, \ldots, 20$.

we obtain an error that can be fully neglected, see Figure 8. We demonstrate the case of $\widetilde{n} = 15$ in the same figure to illustrate that even a reduction to such a small dimension (in comparison to $n = 3280$) a relative error of less that 0.01 can be guaranteed that is often around 10^{-3}.

References

[Arr+24] P. Arrubarrena, M. Lemercier, B. Nikolic, T. Lyons, and T. Cass. "Novelty Detection on Radio Astronomy Data using Signatures". In: *arXiv preprint arXiv:2402.14892* (2024).

[Bay+23a] C. Bayer, P. Hager, S. Riedel, and J. Schoenmakers. "Optimal stopping with signatures". In: *The Annals of Applied Probability* 33.1 (2023), pp. 238–273.

[Bay+23b] C. Bayer, P. K. Friz, M. Fukasawa, J. Gatheral, A. Jacquier, and M. Rosenbaum. *Rough volatility*. SIAM, 2023.

[BD11] P. Benner and T. Damm. "Lyapunov equations, energy functionals, and model order reduction of bilinear and stochastic systems." English. In: *SIAM J. Control Optim.* 49.2 (2011), pp. 686–711. DOI: 10.1137/09075041X.

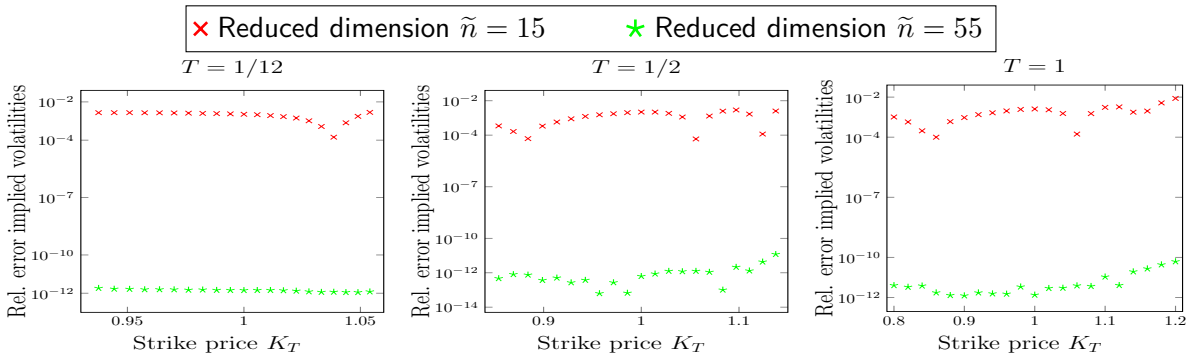

Figure 8: Relative error between volatilities of signature model associated to (4.2) and reduced systems with $\widetilde{n} = 15, 55$, strike prices $K_T = (0.8 + j \cdot 0.02)^{\sqrt{T}}$ and $T = 1/12, 1/2, 1$.

[BDRC17] P. Benner, T. Damm, and Y. R. Rodriguez Cruz. "Dual pairs of generalized Lyapunov inequalities and balanced truncation of stochastic linear systems". In: *IEEE Trans. Autom. Contr.* 62.2 (2017), pp. 782–791.

[Ber15] L. Bergomi. *Stochastic volatility modeling*. CRC press, 2015.

[BFG16] C. Bayer, P. K. Friz, and J. Gatheral. "Pricing under rough volatility". In: *Quantitative Finance* 16.6 (2016), pp. 887–904.

[BFZ24] E. Bayraktar, Q. Feng, and Z. Zhang. "Deep Signature Algorithm for Multidimensional Path-Dependent Options". In: *SIAM Journal on Financial Mathematics* 15.1 (2024), pp. 194–214. DOI: 10.1137/23M1571563.

[BH19] S. Becker and C. Hartmann. "Infinite-dimensional bilinear and stochastic balanced truncation with error bounds". In: *Math. Control Signals, Syst.* 31 (2019), pp. 1–37.

[BHR24] C. Bayer, P. Hager, and S. Riedel. "Optimal stopping for non-Markovian asset price processes". In: *XXXX*. Springer, 2024+.

[BLP22] M. Bennedsen, A. Lunde, and M. S. Pakkanen. "Decoupling the short-and long-term behavior of stochastic volatility". In: *Journal of Financial Econometrics* 20.5 (2022), pp. 961–1006.

[BPS23] C. Bayer, L. Pelizzari, and J. Schoenmakers. "Primal and dual optimal stopping with signatures". In: *arXiv preprint:2312.03444* (2023).

[BR15] P. Benner and M. Redmann. "Model Reduction for Stochastic Systems." English. In: *Stoch PDE: Anal Comp* 3.3 (2015), pp. 291–338. ISSN: 2194-0401; 2194-041X. DOI: 10.1016/j.laa.2014.12.005.

[Bue+20] H. Buehler, B. Horvath, T. Lyons, I. Perez Arribas, and B. Wood. "Generating financial markets with signatures". In: *Available at SSRN 3657366* (2020).

[CGSF22] C. Cuchiero, G. Gazzani, and S. Svaluto-Ferro. "Signature-based models: theory and calibration". In: *arXiv preprint:2207.13136* (2022).

[Che57] K.-T. Chen. "Integration of paths, geometric invariants and a generalized Baker-Hausdorff formula". In: *Annals of Mathematics* (1957), pp. 163–178.

[CO22] I. Chevyrev and H. Oberhauser. "Signature moments to characterize laws of stochastic processes". In: *Journal of Machine Learning Research* 23.176 (2022), pp. 1–42.

[CPSF22] C. Cuchiero, F. Primavera, and S. Svaluto-Ferro. "Universal approximation theorems for continuous functions of càdlàg paths and Lévy-type signature models". In: *arXiv preprint:2208.02293* (2022).

[CST23] C. Cuchiero, P. Schmocker, and J. Teichmann. "Global universal approximation of functional input maps on weighted spaces". In: *arXiv preprint:2306.03303* (2023).

[Dup19] B. Dupire. "Functional Itô calculus". In: *Quantitative Finance* 19.5 (2019), pp. 721–729.

[FH20] P. K. Friz and M. Hairer. *A course on rough paths*. Springer, 2020.

[FV10] P. K. Friz and N. B. Victoir. *Multidimensional stochastic processes as rough paths: theory and applications*. Vol. 120. Cambridge University Press, 2010.

[GJ90] W. Gawronski and J. Juang. "Model reduction in limited time and frequency intervals". In: *Int. J. Syst. Sci.* 21.2 (1990), pp. 349–376. DOI: 10.1080/00207729008910366.

[Guy22] J. Guyon. "The VIX future in Bergomi models: Fast approximation formulas and joint calibration with S&P 500 skew". In: *SIAM Journal on Financial Mathematics* 13.4 (2022), pp. 1418–1485.

[HL10] B. Hambly and T. Lyons. "Uniqueness for the signature of a path of bounded variation and the reduced path group". In: *Annals of Mathematics* (2010), pp. 109–167.

[KLPA20] J. Kalsi, T. Lyons, and I. Perez Arribas. "Optimal Execution with Rough Path Signatures". In: *SIAM Journal on Financial Mathematics* 11.2 (2020), pp. 470–493. DOI: 10.1137/19M1259778.

[LM22] T. Lyons and A. D. McLeod. "Signature methods in machine learning". In: *arXiv preprint arXiv:2206.14674* (2022).

[LO23] D. Lee and H. Oberhauser. "The Signature Kernel". In: *arXiv preprint arXiv:2305.04625* (2023).

[Lyo14] T. Lyons. "Rough paths, signatures and the modelling of functions on streams". In: *arXiv preprint arXiv:1405.4537* (2014).

[Lyo98] T. Lyons. "Differential equations driven by rough signals". In: *Revista Matemática Iberoamericana* 14.2 (1998), pp. 215–310.

[Moo81] B. Moore. "Principal component analysis in linear systems: Controllability, observability, and model reduction". In: *IEEE Trans. Autom. Control* 26.1 (1981), pp. 17–32.

[Mor+20] J. Morrill, A. Fermanian, P. Kidger, and T. Lyons. "A generalised signature method for multivariate time series feature extraction". In: *arXiv preprint arXiv:2006.00873* (2020).

[PA+18] I. Perez Arribas, G. M. Goodwin, J. R. Geddes, T. Lyons, and K. E. A. Saunders. "A signature-based machine learning model for distinguishing bipolar disorder and borderline personality disorder". In: *Translational psychiatry* 8.1 (2018), p. 274.

[RBG21] M. Redmann, C. Bayer, and P. Goyal. "Low-dimensional approximations of high-dimensional asset price models". In: *SIAM J. Financial Math.* 12.1 (2021), pp. 1–28.

[Red22] M. Redmann. "Solving high-dimensional optimal stopping problems using optimization based model order reduction". In: *Applied Mathematical Finance* 29.2 (2022), pp. 110–140.

[RJ22] M. Redmann and N. Jamshidi. "Gramian-based model reduction for unstable stochastic systems". In: *Mathematics of Control, Signals, and Systems* 34 (2022), pp. 855–881.

[SA23] A. Schell and R. Alaifari. "Nonparametric Regression of Stochastic Processes via Signatures". https://www.sam.math.ethz.ch/sam_reports/reports_final/reports2023/2023-45_rev1.pdf. 2023.