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

Preprint ISSN 2198-5855

Optimal stopping with signatures

Christian Bayer[1], Paul Hager[2], Sebastian Riedel[2], John G. M. Schoenmakers[1]

submitted: November 26, 2020

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

[2] Technische Universität Berlin
Straße des 17. Juni 136
10623 Berlin
Germany
E-Mail: phager@math.tu-berlin.de
riedel@math.tu-berlin.de

No. 2790
Berlin 2020

2010 *Mathematics Subject Classification.* 60L10, 60G40.

Key words and phrases. Signatures, rough paths, optimal stopping.

Financial support from the German science foundation (DFG) via the cluster of excellence MATH+, project AA4-2, is gratefully acknowledged.

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/

Optimal stopping with signatures

Christian Bayer, Paul Hager, Sebastian Riedel, John G. M. Schoenmakers

Abstract

We propose a new method for solving optimal stopping problems (such as American option pricing in finance) under minimal assumptions on the underlying stochastic process. We consider classic and randomized stopping times represented by linear functionals of the associated rough path *signature*, and prove that maximizing over the class of *signature stopping times*, in fact, solves the original optimal stopping problem. Using the algebraic properties of the signature, we can then recast the problem as a (deterministic) optimization problem depending only on the (truncated) expected signature. The only assumption on the process is that it is a continuous (geometric) random rough path. Hence, the theory encompasses processes such as fractional Brownian motion which fail to be either semi-martingales or Markov processes.

1 Introduction

The theory of *rough paths*, see, for instance, [LCL07, FV10, FH14], provides a powerful and elegant pathwise theory of stochastic differential equations driven by general classes of stochastic processes – or, more precisely, rough paths. One of the benefits of the theory is that the resulting solution maps are continuous rather than merely measurable as in the Itō version. This property has lead to many important theoretical progresses, most notably perhaps Hairer's theory for singular non-linear SPDEs.

In addition to the theoretical advances, tools – specifically, the path *signature* – from rough path analysis play an increasingly prominent role in applications, most notably in *machine learning*, see, e.g., [AGG+18]. Intuitively, the signature $\mathbb{X}^{<\infty}$ of a path $X : [0, T] \to \mathbb{R}^d$ denotes the (infinite) collection of all iterated integrals of all components of the path against each other, i.e., of the form

$$\int_{0<t_1<\cdots<t_n<T} dX^{i_1}_{t_1} \cdots dX^{i_n}_{t_n},$$

$i_1, \ldots, i_n \in \{1, \ldots, d\}$, $n \geq 0$. To better understand the importance of the signature, let us first recall that the signature $\mathbb{X}^{<\infty}$ determines the underlying path X (up to "tree-like excursions"), which was first proved in [HL10] for paths X of bounded variation and later extended to less regular paths. This implies that, in principle, we can always work with the signature rather than the path. (A somewhat dubious proposition, as we merely replace one infinite dimensional object by another one.) However, the signature is not an arbitrary encoding of the path. Rather, Lyons' universal limit theorem suggests that the solution of a differential equation driven by a rough path X can be approximated with high accuracy by relatively few terms of the signature $\mathbb{X}^{<\infty}$. In that sense, an appropriately truncated signature can be seen as a highly efficient *compression* of X, at least in the context of dynamical systems. And, indeed, there is now ample evidence of the power of the signature as a *feature* in the sense of machine learning.

This paper is motivated by another recent application of signatures, namely the solution of stochastic optimal control problems in finance. We follow the presentation of [KLPA20], where a signature-based approach for solving optimal execution problems is developed. In a nutshell, the strategy can be summarized as follows:

DOI 10.20347/WIAS.PREPRINT.2790 Berlin 2020

1 Trading strategies for execution of a position can be understood as (continuous) functionals $\phi(X|_{[0,t]})$ of the price path, and, hence, as functionals $\theta(\mathbb{X}^{<\infty}_{0,t})$ of the signature (at least approximately).

2 Taking advantage of the algebraic structure of the signature (see Section 2 below), we may efficiently approximate continuous functionals $\theta(\mathbb{X}^{<\infty}_{0,t})$ by *linear* functionals $\langle l, \mathbb{X}^{<\infty}_{0,t}\rangle$, which further extends to the whole value function.

3 Taking advantage of the linearity, we may interchange the expectation with the linear functional, thereby reducing the optimal control problem to a problem of maximizing $l \mapsto \langle l, \mathbb{E}[\mathbb{X}^{<\infty}_{0,t}]\rangle$ over some set of dual elements l.

4 Truncate the expected signature to a finite level N.

The above strategy, in principle, only imposes very mild conditions on the underlying process X, mainly that it is continuous but rough. In particular, X does not need to be a Markov process or a semi-martingale. For this reason, we may consider the approach to be model-free.[1] Note, however, that the assumption of the existence of the expected signature $\mathbb{E}[\mathbb{X}^{<\infty}_{0,t}]$ is a rather strong assumption – in particular, ruling out many stochastic volatility models, such as the Heston model.

[KLPA20] contains extensive numerical examples, indicating the method's excellent performance in various scenarios and models, often beating benchmark methods from the financial literature. On the other hand, theoretical justification of the different approximation steps summarized above is largely missing.

In this paper, we extend the method of [KLPA20] to another important control problem in finance, namely the *optimal stopping* problem, or, in more financial terms, the pricing of *American options*. More specifically, we are concerned with the problem of computing

$$\sup_{\tau\in\mathcal{S}} \mathbb{E}\left[Y_{\tau\wedge T}\right], \tag{1.1}$$

where Y denotes a process adapted to the filtration $(\mathcal{F}_t)_{t\in[0,T]}$ generated by a rough path process $(X_t)_{t\in[0,T]}$, and $\mathcal{S}$ denotes the set of all stopping times w.r.t. the same filtration. In a financial context, Y usually denotes a reward process discounted with respect to some numéraire. At first glance the optimal stopping problem (1.1) may seem unsuitable for the signature-based approach, as typical candidate stopping times are hitting times of sets, which are generally discontinuous w.r.t. the underlying path. We solve this issue by using *randomized stopping times*, see [BTW20]. Note that extending the set $\mathcal{S}$ to also include randomized stopping times does not change the value of (1.1). In the end, we are able to prove that replacing proper stopping times by *signature stopping times* – i.e., stopping times given in terms of linear functionals of the signature $\mathbb{X}^{<\infty}$ – does not change the value of the optimal stopping problem either. More precisely, we have

Theorem 1.1. *Assume that* $\mathbb{E}[\|Y\|_\infty] < \infty$. *Then,*

$$\sup_{\tau\in\mathcal{S}} \mathbb{E}[Y_{\tau\wedge T}] = \sup_{\tau_l} \mathbb{E}[Y_{\tau_l\wedge T}],$$

where the supremum on the right-hand-side ranges over stopping times $\tau_l := \inf\{t \in [0,T] : \langle l, \widehat{\mathbb{X}}^{<\infty}_{0,t}\rangle \geq 1\}$ *connected with linear functionals* l *on the signature process* $\widehat{\mathbb{X}}^{<\infty}_{0,t}$ *(we refer to Sections 2 and 4 for precise definitions).*

[1] The full expected signature $\mathbb{E}[\mathbb{X}^{<\infty}]$ typically characterizes the distribution of the process X, see [CL16]. In that sense, we would hesitate to regard methods relying on the full (rather than truncated) expected signature as "model-free".

The theorem is presented as Theorem 4.8 below. We note that, following [KLPA20], we extend the path X by adding running time as an additional component. $\widehat{\mathbb{X}}^{<\infty}$ denotes the signature of the extended path.

In the next step we need to actually compute a maximizing signature stopping time. In this context, this most importantly implies replacing the full signature $\widehat{\mathbb{X}}^{<\infty}$ by a truncated version $\widehat{\mathbb{X}}^{\leq N}$. Using some further technical assumptions, Theorem 5.5 provides convergence of the corresponding approximations to the value of (1.1) as $N \to \infty$.

Finally, assuming that Y is a polynomial function of X – or, more generally, of $\mathbb{X}^{<\infty}$ – we can derive an approximation formula in terms of an optimization problem involving linear functionals of the expected signature $\mathbb{E}[\widehat{\mathbb{X}}^{\leq N}]$ rather than the expectation of some functional of the signature. See Corollary 5.6 for details.

Outline of the paper

Section 2 recalls basic definitions from the theory of rough paths and provides the algebraic and analytic setting of signatures. A framework for studying stopped rough paths is presented in Section 3. Finally, the optimal stopping problem based on signature stopping times is studied in Section 4. In the following Section 5 we consider the numerical approximation of the optimal signature stopping problem.

2 Preliminaries

We start by introducing the basic definitions needed for understanding signatures and their algebraic and – in the context of rough paths – analytic properties. These definitions are standard in the rough path literature, we refer to [LCL07, FH14, FV10] for a more detailed exposition.

2.1 The tensor algebra

Let V be a finite-dimensional $\mathbb{R}$-vector space with basis $\{e_1, \dots, e_d\}$. The dual space is denoted by V^* with dual basis $\{e_1^*, \dots, e_d^*\}$. We define the *tensor algebra* and the *extended tensor algebra* by setting

$$T(V) := \bigoplus_{n=0}^{\infty} V^{\otimes n} \quad \text{and} \quad T((V)) := \prod_{n=0}^{\infty} V^{\otimes n}$$

where $V^{\otimes n}$ denotes the n-th tensor power of V with the convention $V^{\otimes 0} := \mathbb{R}$, $V^{\otimes 1} := V$. Note that there is a natural pairing between $T((V))$ and $T(V^*)$ which we denote by

$$\langle \cdot, \cdot \rangle : T(V^*) \times T((V)) \to \mathbb{R}.$$

We define sum and product of two elements $\mathbf{a} = (a_n)_{n=0}^{\infty}, \mathbf{b} = (b_n)_{n=0}^{\infty} \in T((V))$ by setting

$$\mathbf{a} + \mathbf{b} := (a_n + b_n)_{n=0}^{\infty},$$
$$\mathbf{a} \otimes \mathbf{b} := (\sum_{i+j=n} a_i \otimes b_j)_{n=0}^{\infty}.$$

For $\lambda \in \mathbb{R}$, we define $\lambda \mathbf{a} := (\lambda a_n)_{n=0}^{\infty}$. We also let $\mathbf{0} := (0, 0, \dots)$ and $\mathbf{1} := (1, 0, 0, \dots)$. Note that

$$\mathbf{1} \otimes \mathbf{a} = \mathbf{a} \otimes \mathbf{1} = \mathbf{a}$$

for every $\mathbf{a} \in T((V))$. The *truncated tensor algebra* is defined by

$$T^N(V) := \bigoplus_{n=0}^{N} V^{\otimes n}.$$

We define maps $\pi_n \colon T((V)) \to V^{\otimes n}$ and $\pi_{\leq N} \colon T((V)) \to T^N(V)$ by $\pi_n(\mathbf{a}) = a_n$ and $\pi_{\leq N}(\mathbf{a}) = (a_0, \ldots, a_N)$ where $\mathbf{a} = (a_n)_{n=0}^{\infty}$. We will sometimes abuse notation and write $\mathbf{0}$ and $\mathbf{1}$ for the elements $\pi_{\leq N}(\mathbf{0})$ and $\pi_{\leq N}(\mathbf{1})$ in the truncated tensor algebra.

Next, we consider norms on $T((V))$ and $T(V^*)$. On V, we choose the l^∞-norm, i.e. for $v = \lambda_1 e_1 + \ldots + \lambda_d e_d$, we set $|v| := \max_i |\lambda_i|$. For elements in V^*, we use the l^1-norm, i.e. $|v^*| := |\lambda_1| + \ldots + |\lambda_d|$ for $v^* = \lambda_1 e_1^* + \ldots + \lambda_d e_d^*$. On the tensor powers of V resp. V^*, we use the corresponding norms, too. Note that the norms on the tensor products $V^{\otimes n}$ are *admissible*, meaning that if $v = a_1 \otimes \ldots \otimes a_k$ and $\sigma v := a_{\sigma(1)} \otimes \ldots \otimes a_{\sigma(k)}$ for a permutation σ, $|\sigma v| = |v|$, and $|v \otimes w| \leq |v||w|$. We set

$$|\mathbf{a}| := \sup_{i \in \mathbb{N}_0} |a_i| \in [0, \infty] \quad \text{for } \mathbf{a} = (a_i)_{i=0}^{\infty} \in T((V))$$

and

$$|\mathbf{b}| := \sum_{i=0}^{\infty} |b_i| \in [0, \infty) \quad \text{for } \mathbf{b} = (b_i)_{i=0}^{\infty} \in T(V^*).$$

Note that we always have

$$|\langle \mathbf{b}, \mathbf{a} \rangle| \leq |\mathbf{b}||\pi_{\leq N}(\mathbf{a})| \leq |\mathbf{b}||\mathbf{a}|$$

where $N = \max\{i \in \mathbb{N}_0 \,:\, b_i \neq 0\}$.

2.2 Shuffles

In the following, calculations will mainly be performed in the space $T(V^*)$. In order to simplify notations, we will replace expressions like $e_{i_1}^* \otimes \cdots \otimes e_{i_n}^*$ by the much simpler form $i_1 \cdots i_n$. More precisely, let $\mathcal{W}(\mathcal{A}_d)$ be the linear span of words composed by the letters in the dictionary $\mathcal{A}_d = \{1, \ldots, d\}$. The empty word is denoted by $\varnothing \in \mathcal{W}(\mathcal{A}_d)$. We can naturally define the sum $l_1 + l_2$ and the scalar product λl for elements $l, l_1, l_2 \in \mathcal{W}(\mathcal{A}_d)$ and $\lambda \in \mathbb{R}$. If $w = i_1 \cdots i_n$ and $v = j_1 \cdots j_m$ are two words, the *concatenation* is defined by

$$wv := i_1 \cdots i_n j_1 \cdots j_m.$$

This operation is extended bi-linearly to elements in $\mathcal{W}(\mathcal{A}_d)$. The basis elements $\{e_{i_1}^* \otimes \cdots \otimes e_{i_n}^* : i_1, \ldots, i_n \in \{1, \ldots, d\}\}$ in $(V^*)^{\otimes n}$ can be identified with words via the map

$$e_{i_1}^* \otimes \cdots \otimes e_{i_n}^* \mapsto i_1 \cdots i_n$$

which induces an isomorphism $T(V^*) \cong \mathcal{W}(\mathcal{A}_d)$. We can also think of $\mathcal{W}(\mathcal{A}_d)$ as the space of non-commutative polynomials where the unknown are given by the letters $\{1, \ldots, d\}$. For a word $w = i_1 \cdots i_n$, set $\deg(w) := n$ and $\deg(\varnothing) := 0$. If $l = \lambda_1 w_1 + \ldots + \lambda_n w_n \in \mathcal{W}(\mathcal{A}_d)$ with $\lambda_1, \ldots, \lambda_n \in \mathbb{R} \setminus \{0\}$ and $w_1, \ldots, w_n$ words, we define

$$\deg(l) := \max_{i=1\ldots,n} \deg(w_i).$$

Apart from concatenation, there is a second important product defined on $\mathcal{W}(\mathcal{A}_d)$ which is called *shuffle product*: For a word w, we set

$$w \sqcup\!\sqcup \varnothing := \varnothing \sqcup\!\sqcup w := w.$$

If wi and vj are words and $i, j \in \mathcal{A}_d$ are letters, we recursively define $wi \sqcup\!\sqcup vj \in \mathcal{W}(\mathcal{A}_d)$ by

$$wi \sqcup\!\sqcup vj := (w \sqcup\!\sqcup vj)i + (wi \sqcup\!\sqcup v)j.$$

This operation is extended bi-linearly to a product $\sqcup\!\sqcup\colon \mathcal{W}(\mathcal{A}_d) \times \mathcal{W}(\mathcal{A}_d) \to \mathcal{W}(\mathcal{A}_d)$. Note that $\sqcup\!\sqcup$ is associative, commutative and distributive over $+$. If $P \in \mathbb{R}[x]$ is a commutative polynomial with unknown variable x, i.e. $P(x) = \lambda_0 + \lambda_1 x + \ldots \lambda_n x^n$, we define $P^{\sqcup\!\sqcup}\colon \mathcal{W}(\mathcal{A}_d) \to \mathcal{W}(\mathcal{A}_d)$ by setting

$$P^{\sqcup\!\sqcup}(l) := \lambda_0 \varnothing + \lambda_1 l + \lambda_2 (l \sqcup\!\sqcup l) + \ldots + \lambda_n l^{\sqcup\!\sqcup n}, \tag{2.1}$$

where $l^{\sqcup\!\sqcup k}$ denotes k-th shuffle product of $l \in \mathcal{W}(\mathcal{A}_d)$ with itself.

We define

$$G(V) := \{\mathbf{a} \in T((V)) \setminus \{\mathbf{0}\} \;:\; \langle l_1 \sqcup\!\sqcup l_2, \mathbf{a}\rangle = \langle l_1, \mathbf{a}\rangle\langle l_2, \mathbf{a}\rangle \text{ for every } l_1, l_2 \in T(V^*)\} \tag{2.2}$$

and call it the set of *group-like elements*. Note that $\pi_0(\mathbf{g}) = 1$ for every $\mathbf{g} \in G(V)$. One can show that $(G(V), \otimes)$ is a group with identity $\mathbf{1}$ and inverse given by

$$\mathbf{g}^{-1} = \sum_{n \geq 0} (\mathbf{1} - \mathbf{g})^{\otimes n}.$$

We also set $G^N(V) := \pi_{\leq N}(G(V))$ which is a free nilpotent group of order N with respect to the "truncated multiplication" $\mathbf{a} \otimes_{G^N(V)} \mathbf{b} := \pi_N(\mathbf{a} \otimes \mathbf{b})$, for $\mathbf{a}, \mathbf{b} \in G^N(V)$. However, we will not distinguish between the multiplication symbols on $G^N(V)$ and $G(V)$ and use $\otimes$ in both cases.

Remark 2.1. *The relation $\langle l_1 \sqcup\!\sqcup l_2, \mathbf{a}\rangle = \langle l_1, \mathbf{a}\rangle\langle l_2, \mathbf{a}\rangle$ for $\mathbf{a} \in G(V)$ implies that*

$$P(\langle l, \mathbf{a}\rangle) = \langle P^{\sqcup\!\sqcup}(l), \mathbf{a}\rangle \tag{2.3}$$

for any polynomial P. This is really the justification for introducing the shuffle product, as it provides an explicit linearization of polynomials in the signature.

2.3 Rough paths and their signatures

Now that the algebraic setting for signatures is developed (for the purposes of this paper), we can finally consider the analytic properties of (rough) paths. More concretely, given a path $X : [0, T] \to V$ (of sufficient regularity), we will associate to it a function $\mathbb{X}$ taking values in the truncated tensor algebra, which is the fundamental building block of rough path theory. Set $\Delta_T := \{(s, t) \in [0, T]^2 : 0 \leq s \leq t \leq T\}$. For a map $\mathbb{X}\colon \Delta_T \to T^N(V)$, we define its p-*variation*

$$\|\mathbb{X}\|_{p-\mathrm{var};[s,t]} := \max_{k=1,\ldots,N} \sup_{\mathcal{D} \subset [s,t]} \left(\sum_{t_i \in \mathcal{D}} |\pi_k(\mathbb{X}_{t_i, t_{i+1}})|^{\frac{p}{k}} \right)^{\frac{k}{p}}$$

where the supremum ranges over all partitions $\mathcal{D}$ of $[s, t]$. We will use the notation $\|\mathbb{X}\|_{p-\mathrm{var}} := \|\mathbb{X}\|_{p-\mathrm{var};[0,T]}$. For $\mathbb{X}, \mathbb{Y}\colon \Delta_T \to T^N(V)$, we define the p-*variation distance*

$$d_{p-\mathrm{var};[s,t]}(\mathbb{X}, \mathbb{Y}) := \|\mathbb{X} - \mathbb{Y}\|_{p-\mathrm{var};[s,t]}$$

and set $d_{p-\mathrm{var}}(\mathbb{X},\mathbb{Y}) := d_{p-\mathrm{var};[0,T]}(\mathbb{X},\mathbb{Y})$. A *weakly geometric p-rough path* $\mathbb{X}$ is a continuous path $\mathbb{X}\colon [0,T] \to G^{\lfloor p \rfloor}(V)$ with $\mathbb{X}_0 = \mathbf{1}$ and $\|\mathbb{X}\|_{p-\mathrm{var}} < \infty$ where we set $\mathbb{X}_{s,t} := \mathbb{X}_s^{-1} \otimes \mathbb{X}_t$ for $s \leq t$. Note that $\mathbb{X}_t = \mathbb{X}_{0,t}$. We denote the space of weakly geometric p-rough paths by $\mathcal{W}\Omega_T^p$ and equip it with the distance $d_{p-\mathrm{var}}$. If $X\colon [0,T] \to V$ is a continuous path of bounded variation, we define its *signature* $\mathbb{X}^{<\infty}\colon [0,T] \to T((V))$ by

$$\pi_k(\mathbb{X}_t^{<\infty}) := \int_{0<t_1<\ldots<t_k<t} dX_{t_1} \otimes \cdots \otimes dX_{t_k}.$$

The *truncated signature* $\mathbb{X}^{\leq N}\colon [0,T] \to T^N(V)$ is defined by $\mathbb{X}^{\leq N} := \pi_{\leq N}(\mathbb{X}^{<\infty})$. It can be checked that $\mathbb{X}^{<\infty}$ takes values in $G(V)$ and we set $\mathbb{X}_{s,t}^{<\infty} := (\mathbb{X}_s^{<\infty})^{-1} \otimes \mathbb{X}_t^{<\infty}$ so that

$$\pi_k(\mathbb{X}_{s,t}^{<\infty}) = \int_{s<t_1<\ldots<t_k<t} dX_{t_1} \otimes \cdots \otimes dX_{t_k}.$$

One can also show that $\mathbb{X}^{\leq N}$ is an element in $\mathcal{W}\Omega_T^p$ for every $p \geq 1$ with $N = \lfloor p \rfloor$.

A *geometric p-rough path* $\mathbb{X}$ is a weakly geometric rough path $\mathbb{X} \in \mathcal{W}\Omega_T^p$ for which there exists a sequence of piecewise smooth paths (X_n) such that $d_{p-\mathrm{var}}(\mathbb{X}, \mathbb{X}_n^{\leq \lfloor p \rfloor}) \to 0$ as $n \to \infty$. The space of geometric rough paths is denoted by Ω_T^p. It can be shown that the inclusion $\Omega_T^p \subset \mathcal{W}\Omega_T^p$ is strict and that Ω_T^p is a Polish space. From Lyons' Extension theorem [LCL07, Theorem 3.7], every geometric rough path $\mathbb{X} \in \Omega_T^p$ has a unique lift $\mathbb{X}^{<\infty}$ which is a path in $G(V)$, satisfying $\|\pi_{\leq N}(\mathbb{X}^{<\infty})\|_{p-\mathrm{var}} < \infty$ for every $N \geq 1$ and $\pi_{\leq \lfloor p \rfloor}(\mathbb{X}^{<\infty}) = \mathbb{X}$. We call $\mathbb{X}^{<\infty}$ the *signature of the rough path* $\mathbb{X}$ and $\mathbb{X}^{\leq N} := \pi_{\leq N}(\mathbb{X}^{<\infty})$ its *truncated signature*.

Similarly, for $V = \mathbb{R}^{1+d}$, we define the space $\widehat{\Omega}_T^p$ as the closure of rough path lifts $\widehat{\mathbb{X}}^{\leq \lfloor p \rfloor}$ in the p-variation distance where $\widehat{X}_t = (t, X_t) \in \mathbb{R}^{1+d}$ and X is piecewise smooth. It follows that $\widehat{\Omega}_T^p$ is Polish.

Remark 2.2. *Following the notation introduced above, the letter 1 corresponds to the running time component t of the path $\widehat{X}$, whereas the components of X correspond to the letters $2, \ldots, d+1$, respectively.*

Example 2.3 (Brownian motion as a rough path)**.** *Let X be a d-dimensional Brownian motion. In this case a natural lift to a geometric rough path $\mathbb{X} \in \Omega_T^p$ with $p \in (2,3)$ is given by*

$$\mathbb{X}_{s,t} = \left(1, X_{s,t}, \int_s^t X_{s,u} \otimes \circ dX_u\right), \quad 0 \leq s \leq t \leq T.$$

where $X_{s,t} = X_t - X_s$ and for all $i, j \in \mathcal{A}_d$ the tensor valued Stratonovich integral is given by

$$\left\langle ij, \int_s^t X_{s,u} \otimes \circ dX_u \right\rangle = \int_s^t X_{s,u}^i \circ dX_u^j = \int_s^t X_{s,u}^i dX_u^j + \frac{1}{2}[X^i, X^j]_{s,t}.$$

Indeed, to see that $\mathbb{X} \in \mathcal{W}\Omega_T^p$, on may readily check that $\mathbb{X}_{s,t} \in G^2(V)$ is an immediate consequence of the product rule and the rough path regularity of $\mathbb{X}$ follows from a generalized Kolmogorov criterion (see [FH14, Theorem 3.1]). As it is well known that the integral with respect to the piecewise linear approximation of Brownian motion converges to the Stratonovich integral, we also see that $\mathbb{X} \in \Omega_T^p$. The signature $\mathbb{X}^{<\infty}$ of the enhanced Brownian motion $\mathbb{X}$ is then given by the iterated integrals of all orders, i.e. for any word $w = i_1 \cdots i_k \in \mathcal{W}(\mathcal{A}_d)$ we have

$$\langle i_1 \cdots i_k, \mathbb{X}_{0,t}^{<\infty} \rangle = \int_{0<t_1<\ldots<t_k<t} \circ dX_{t_1}^{i_1} \cdots \circ dX_{t_k}^{i_k}.$$

An explicit form of the expected signature is also known due to Fawcett [Faw02]

$$\mathbb{E}(\mathbb{X}_{0,t}^{<\infty}) = \exp\left(\frac{1}{2}t\sum_{i=1}^{d} e_i \otimes e_i\right) = \sum_{n=0}^{\infty}\frac{1}{n!}\frac{t^n}{2^n}\left(\sum_{i=1}^{d} e_i\otimes e_i\right)^{\otimes n}.$$

Note that this construction of a geometric rough path works in principle for all continuous semimartingales and we refer to [FV10, Section 14] for more detail.

Example 2.4 (Fractional Brownian motion)**.** *Let X be one-dimensional fractional Brownian motion with Hurst parameter $H \in (0,1)$, i.e. X is a zero mean Gaussian process with covariance function*

$$\mathbb{E}(X_sX_t) = \frac{1}{2}\left(|s|^{2H} + |t|^{2H} - |t-s|^{2H}\right), \quad 0 \le s \le t.$$

Recall that the sample paths of X are $(H-\varepsilon)$-Hölder continuous for any $\varepsilon > 0$. In case $H = 1/2$ then X is just a standard Brownian motion, in case $H \neq 1/2$ then X is not a Markov process and not a semimartingale. However, since X is one-dimensional ($V = \mathbb{R}$) there is a trivial lift to a geometric rough path $\mathbb{X} \in \Omega_T^p$ for any $p \in (1/H, 1+1/H)$ given by

$$\mathbb{X}_{s,t} = \left(1, X_{s,t}, \frac{1}{2}(X_{s,t})^{\otimes 2}, ..., \frac{1}{[p]!}(X_{s,t})^{\otimes [p]}\right) \equiv \exp_{[p]}(X_{s,t}) \in G^{[p]}(V), \quad 0 \le s \le t \le T.$$

As we will see in the next section, we are particulary interested in the process $\widehat{X}$ defined by $\widehat{X}_t = (t, X_t) \in \mathbb{R}^2$. Since the first component of $\widehat{X}$ is of locally bounded variation, there is a lift to a geometric rough path $\widehat{\mathbb{X}} \in \widehat{\Omega}_T^p$ (see [FV10, Theorem 9.26]). Intuitively speaking we can make use of the abundant regularity of the first component $\widehat{X}$ in order to define iterated integrals by imposing the integration by parts rule. More precisely in case $p > 2$ we have

$$\langle 12, \widehat{\mathbb{X}}_{s,t}\rangle = \langle 2, \widehat{\mathbb{X}}_{s,t}\rangle\langle 1, \widehat{\mathbb{X}}_{s,t}\rangle - \langle 21, \widehat{\mathbb{X}}_{s,t}\rangle = X_{s,t}(t-s) - \int_s^t X_{s,u}du,$$

and the right hand side is clearly well defined. Using the shuffle identity, this reasoning can be carried on to express all components of the signature $\widehat{\mathbb{X}}^{<\infty}$ in terms of increments of X, finite-variation integrals and products thereof.

3 The space of stopped rough paths

We will now consider rough paths $\mathbb{Z}$ defined on some intervals $[0,s] \subset [0,T]$. In order to naturally model the notion of *adaptedness* to a filtration, we will consider functionals of the restriction of a rough path $\mathbb{Z}$ to a subinterval of its domain. Hence, the analysis of the corresponding control problem requires us to define a distance of rough paths with different domains. Following [KLPA20], we will use a distance motivated by Dupire's functional Itō calculus, see [Dup19, CF10]. This means, when we compare a path Z^1 defined on $[0,s]$ and another path Z^2 defined on $[0,t]$ with $s < t$, we will extend Z^1 to $[0,t]$ by $Z_u^1 \coloneqq Z_s^1$ for $s \le u \le t$. We will, in principle, use the same construction for rough paths, but recall that we are considering paths $u \mapsto (u, X_u)$ in our framework, and extending the time component of such a path in a constant way does not make much sense. Instead, we will apply Dupire's extension to the X-component, but use the linear extension (i.e., the exact one) for the time component.

More precisely, let $\mathbb{Z}|_{[0,s]} \in \widehat{\Omega}_s^p$ and $s \le t$. By definition, there exists a sequence $Z_u^n = (u, X_u^n)$ where $X^n\colon [0,s] \to \mathbb{R}^d$ is a piecewise smooth path such that $d_{p-\mathrm{var};[0,s]}(\mathbb{Z}|_{[0,s]}, \mathbb{Z}^{n;\le\lfloor p\rfloor}) \to 0$ as

$n \to \infty$. Set $\widetilde{X}^n_u := X^n_{u\wedge s}$ for $u \in [0,t]$ and $\widetilde{Z}^n_u := (u, \widetilde{X}^n_u)$. One can check that $\widetilde{\mathbb{Z}}^{n;\leq \lfloor p \rfloor}$ is a Cauchy sequence in $\widehat{\Omega}^p_t$, and we denote the limit by $\widetilde{\mathbb{Z}}|_{[0,t]}$. One can also check that the definition of $\widetilde{\mathbb{Z}}|_{[0,t]}$ does not depend on the choice of the sequence X^n. By construction we have that $\widetilde{\mathbb{Z}}|_{[0,s]} = \mathbb{Z}|_{[0,s]}$, which motivates the following definition.

Definition 3.1. *For $T > 0$, we set $\Lambda_T := \bigcup_{t\in[0,T]} \widehat{\Omega}^p_t$ and call it the* space of stopped rough paths*. We equip it with the metric*

$$d(\mathbb{X}|_{[0,t]}, \mathbb{Y}|_{[0,s]}) := d_{p-\mathrm{var};[0,t]}(\mathbb{X}|_{[0,t]}, \widetilde{\mathbb{Y}}|_{[0,t]}) + |t-s|$$

where we assume $s \leq t$ and $\widetilde{\mathbb{Y}}|_{[0,t]}$ is the stopped rough path constructed as explained above.

Let us mention that Λ_T is Polish. For this and related simple technical facts about the topology of Λ_T, we refer to the appendix. Later on we will use that $\mathbb{1}_{\{\tau(\omega)\leq t\}}$ can be represented as a measurable map of the restricted signature.

Lemma 3.2. *Let $\widehat{\mathbb{X}}$ be a stochastic process in $\widehat{\Omega}^p_T$ and set $\mathcal{F}_t := \sigma(\widehat{\mathbb{X}}_{0,s} : 0 \leq s \leq t) = \sigma(\widehat{\mathbb{X}}|_{[0,t]})$. Let τ be a stopping time with respect to $(\mathcal{F}_t)$. Then there is a Borel measurable map $\theta\colon \Lambda_T \to \{0,1\}$ such that*

$$\theta(\widehat{\mathbb{X}}(\omega)|_{[0,t]}) = \mathbb{1}_{\{\tau(\omega)\leq t\}}$$

for every $\omega \in \Omega$.

Proof. For every $t \in [0,T]$, $\{\tau \leq t\}$ is $\sigma(\widehat{\mathbb{X}}|_{[0,t]})$-measurable, hence there is a set $A_t \in \mathcal{B}(\widehat{\Omega}^p_t)$ such that $(\widehat{\mathbb{X}}|_{[0,t]})^{-1}(A_t) = \{\tau \leq t\}$. It follows that

$$\mathbb{1}_{\{\tau(\omega)\leq t\}} = \mathbb{1}_{A_t}(\widehat{\mathbb{X}}(\omega)|_{[0,t]}) \tag{3.1}$$

for every $\omega \in \Omega$. Define $\phi\colon \Lambda_T \to [0,T] \times \widehat{\Omega}^p_T$ as $\phi(\mathbb{X}|_{[0,t]}) = (t, \widetilde{\mathbb{X}}|_{[0,T]})$ where $\widetilde{\mathbb{X}}|_{[0,T]}$ denotes the stopped process defined in Definition 3.1 . Note that ϕ is continuous, thus measurable. Define $f\colon [0,T] \times \widehat{\Omega}^p_T \to \mathbb{R}$ as $f(t,\widehat{\mathbb{X}}) := \mathbb{1}_{A_t}(\widehat{\mathbb{X}}|_{[0,t]})$. For fixed t, $\widehat{\mathbb{X}} \to \widehat{\mathbb{X}}|_{[0,t]}$ is continuous and $\widehat{\mathbb{X}}|_{[0,t]} \mapsto \mathbb{1}_{A_t}(\widehat{\mathbb{X}}|_{[0,t]})$ is measurable, therefore $\widehat{\mathbb{X}} \mapsto f(t,\widehat{\mathbb{X}})$ is measurable. For $n \in \mathbb{N}$, define $I^n_k := [k/2^nT, (k+1)/2^nT)$ for $k = 0,\dots,2^n-2$, $I^n_{2^n-1} := [(2^n-1)/2^nT, T]$ and $t^n_k := k/2^nT$. Set

$$f_n(t,\widehat{\mathbb{X}}) := \sum_{k=0}^{2^n-1} f(t^n_k, \widehat{\mathbb{X}})\mathbb{1}_{I^n_k}(t)$$

which is measurable for every $n \in \mathbb{N}$. Set

$$\widetilde{f}(t,\widehat{\mathbb{X}}) := \limsup_{m\to\infty}\limsup_{n\to\infty} f_n(t+1/m, \widehat{\mathbb{X}}) \quad \text{and} \quad \theta(\widehat{\mathbb{X}}|_{[0,t]}) := (\widetilde{f}\circ\phi)(\widehat{\mathbb{X}}|_{[0,t]}).$$

The map θ is thus measurable and satisfies

$$\theta(\widehat{\mathbb{X}}(\omega)|_{[0,t]}) = \limsup_{m\to\infty}\limsup_{n\to\infty}\sum_{k=0}^{2^n-1} \mathbb{1}_{\{\tau(\omega)\leq t^n_k\}}\mathbb{1}_{I^n_k}(t+1/m) = \mathbb{1}_{\{\tau(\omega)\leq t\}}. \qquad \square$$

Definition 3.3. *We set $\mathcal{T} := C(\Lambda_T, \mathbb{R})$ and call it the space of* continuous stopping policies*. The space of* signature stopping policies *$\mathcal{T}_{\mathrm{sig}} \subset \mathcal{T}$ is defined as*

$$\mathcal{T}_{\mathrm{sig}} = \left\{\theta \in \mathcal{T} : \exists l \in T((\mathbb{R}^{1+d})^*) \text{ such that } \theta(\widehat{\mathbb{X}}|_{[0,t]}) = \langle l, \widehat{\mathbb{X}}^{<\infty}_{0,t}\rangle \;\forall \widehat{\mathbb{X}}|_{[0,t]} \in \Lambda_T\right\}.$$

Note that every $l \in T((\mathbb{R}^{1+d})^*)$ defines a $\theta \in \mathcal{T}$ by setting $\theta(\widehat{\mathbb{X}}|_{[0,t]}) = \langle l, \widehat{\mathbb{X}}^{<\infty}_{0,t} \rangle$. The important result about $\mathcal{T}_{\mathrm{sig}}$ is the following:

Lemma 3.4. *Let $\mathbb{P}$ be a probability measure on $(\widehat{\Omega}^p_T, \mathcal{B}(\widehat{\Omega}^p_T))$. Then, for every $\varepsilon > 0$, there is a compact set $\mathcal{K} \subset \widehat{\Omega}^p_T$ such that*

1 $\mathbb{P}(\mathcal{K}) > 1 - \varepsilon$,

2 *$\mathcal{T}_{\mathrm{sig}}$, restricted of $\mathcal{K}$, is dense in $\mathcal{T}$. More precisely, for every $\theta \in \mathcal{T}$ there is a sequence $\theta_n \in \mathcal{T}_{\mathrm{sig}}$ such that*

$$\sup_{\widehat{\mathbb{X}} \in \mathcal{K};\ t \in [0,T]} |\theta_n(\widehat{\mathbb{X}}|_{[0,t]}) - \theta(\widehat{\mathbb{X}}|_{[0,t]})| \to 0$$

as $n \to \infty$.

Proof. [KLPA20, Lemma B.3]. □

4 Signature stopping rules

In the next step, we convert the stopping policies defined above into actual *stopping times*. Of course, for a given stopping time, there is no reason why it should be representable by a continuous stopping policy, or even a signature stopping policy. Indeed, relevant stopping times – such as hitting times of even nice sets – are often discontinuous functions of the underlying path. We will see, however, that stopping times, in particular the optimal stopping times for our problem, can be approximated by stopping times induced by signatures stopping policies, in the sense that the corresponding value functions converge.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. In this section, $\widehat{\mathbb{X}}$ denotes a stochastic process in $\widehat{\Omega}^p_T$ and $Y\colon [0,T] \times \Omega \to \mathbb{R}$ is a real-valued continuous stochastic process adapted to the filtration $(\mathcal{F}_t)$, $\mathcal{F}_t = \sigma(\widehat{\mathbb{X}}_{0,s} \ :\ 0 \le s \le t)$. We are trying to solve the optimal stopping problem for Y, i.e., in a financial context Y corresponds to a cash-flow process. For simplicity, we assume that $X_0 = 0$.

In this section, we will encounter various flavors of stopping times. In particular, as already indicated above, proper stopping times are often difficult to approximate. We will, instead, consider *randomized stopping times*, which relax proper stopping times and lead to much more regular approximation problems. We note that similar techniques have been used in [BTW20] in the context of numerical methods for American option pricing.

Definition 4.1.

1 *By $\mathcal{S}$, we denote the space of all $(\mathcal{F}_t)$-stopping times.*

2 *Let Z be a non-negative random variable independent of $\widehat{\mathbb{X}}$ and such that $\mathbb{P}(Z = 0) = 0$. For a continuous stopping policy $\theta \in \mathcal{T}$, we define the randomized stopping time*

$$\tau^r_\theta := \inf \left\{ t \ge 0 \ :\ \int_0^{t \wedge T} \theta(\widehat{\mathbb{X}}|_{[0,s]})^2 \, ds \ge Z \right\} \tag{4.1}$$

where $\inf \emptyset = +\infty$.

3 *Let $l \in T((\mathbb{R}^{1+d})^*)$ and Z as in (ii). Then we define the* randomized signature stopping time

$$\tau^r_l := \inf \left\{ t \ge 0 \ :\ \int_0^{t \wedge T} \langle l, \widehat{\mathbb{X}}^{<\infty}_{0,s} \rangle^2 \, ds \ge Z \right\}. \tag{4.2}$$

DOI 10.20347/WIAS.PREPRINT.2790

Next we prove that stopping times can be approximated by randomized stopping times based on continuous stopping policies.

Proposition 4.2. *For every stopping time $\tau \in \mathcal{S}$, there exists a sequence $\theta_n \in \mathcal{T}$ such that the randomized stopping times $\tau^r_{\theta_n}$ satisfy $\tau^r_{\theta_n} \to \tau$ almost surely as $n \to \infty$. In particular, if $\mathbb{E}[\|Y\|_\infty] < \infty$, then*

$$\sup_{\theta \in \mathcal{T}} \mathbb{E}[Y_{\tau^r_\theta \wedge T}] = \sup_{\tau \in \mathcal{S}} \mathbb{E}[Y_{\tau \wedge T}].$$

Proof. Let τ be a stopping time. From Lemma 3.2, we know that there is a measurable map $\theta \colon \Lambda_T \to \{0, 1\}$ such that

$$\theta(\widehat{\mathbb{X}}|_{[0,t]}) = \mathbb{1}_{\{\tau \le t\}}.$$

Using [Wiś94, Theorem 1], we can find a sequence of continuous functions $\widetilde{\theta}_n \in \mathcal{T}$ such that $\widetilde{\theta}_n(\widehat{\mathbb{X}}|_{[0,t]}) \to \mathbb{1}_{\{\tau \le t\}}$ almost surely w.r.t. $\lambda|_{[0,T]} \otimes \mathbb{P}$ where $\lambda|_{[0,T]}$ denotes the Lebesgue measure on $[0, T]$. W.l.o.g, we may assume that $0 \le \widetilde{\theta}_n \le 1$. Set $\theta_n := (2\widetilde{\theta}_n)^n$. Then

$$\lim_{n \to \infty} \theta_n(\widehat{\mathbb{X}}|_{[0,t]}) \to \begin{cases} +\infty & \text{if } t \ge \tau \\ 0 & \text{if } t < \tau. \end{cases}$$

It follows that $\tau^r_{\theta_n} \to \tau$ almost surely as $n \to \infty$. Using the dominated convergence theorem, this implies that

$$\sup_{\theta \in \mathcal{T}} \mathbb{E}[Y_{\tau^r_\theta \wedge T}] \ge \sup_{\tau \in \mathcal{S}} \mathbb{E}[Y_{\tau \wedge T}].$$

To show the converse inequality, take $\theta \in \mathcal{T}$. From independence,

$$\mathbb{E}[Y_{\tau^r_\theta \wedge T} \,|\, \widehat{\mathbb{X}}] = \int_0^\infty Y_{\tau_z \wedge T}\, \mathbb{P}_Z(dz)$$

where

$$\tau_z := \inf \left\{ t \ge 0 \,:\, \int_0^{t \wedge T} \theta(\widehat{\mathbb{X}}|_{[0,s]})^2 \, ds \ge z \right\}.$$

Note that this is a stopping time for every $z \ge 0$. Taking expectation, it follows that

$$\mathbb{E}[Y_{\tau^r_\theta \wedge T}] = \int_0^\infty \mathbb{E}[Y_{\tau_z \wedge T}]\, \mathbb{P}_Z(dz) \le \sup_{\tau \in \mathcal{S}} \mathbb{E}[Y_{\tau \wedge T}]$$

which implies the claim. □

Note that we cannot generally assume that $\theta_n \to \theta$ implies $\tau^r_{\theta_n} \to \tau^r_\theta$, as is shown by the following counter-example. So even randomized stopping times are not continuous w.r.t. the underlying stopping policies.

Example 4.3. *Consider $\vartheta, \vartheta_n \colon [0, 3] \to [0, \infty)$ defined by*

$$\vartheta(t) = \begin{cases} 1 - t & \text{if } t \in [0, 1] \\ 0 & \text{if } t \in [1, 2] \\ t - 2 & \text{if } t \in [2, 3] \end{cases} \quad \text{and} \quad \vartheta_n(t) = \begin{cases} (1 - \frac{1}{n})(1 - t) & \text{if } t \in [0, 1] \\ 0 & \text{if } t \in [1, 2] \\ t - 2 & \text{if } t \in [2, 3]. \end{cases}$$

DOI 10.20347/WIAS.PREPRINT.2790

Although $\vartheta_n \to \vartheta$ as $n \to \infty$, we have

$$\inf\left\{t \geq 0 \,:\, \int_0^{t\wedge 3} \vartheta(s)\,ds \geq \frac{1}{2}\right\} = 1 \quad \textit{and} \quad \inf\left\{t \geq 0 \,:\, \int_0^{t\wedge 3} \vartheta_n(s)\,ds \geq \frac{1}{2}\right\} > 2$$

for all $n \geq 1$.

As mentioned above, randomized stopping times *regularize* the optimal stopping problem. Indeed, given a randomized stopping time τ_θ^r defined in terms of an independent random variable Z as in Definition 4.1, if we integrate the stopped process $Y_{\tau_\theta^r \wedge T}$ w.r.t. Z, we obtain a smooth function of θ – which is clearly not true without regularization, see Remark 4.5 below.

Lemma 4.4. *Let S be an $(\mathcal{F}_t)$-stopping time and let F_Z denote the cumulative distribution function of Z. Then*

$$\mathbb{E}[Y_{\tau_\theta^r \wedge S} \,|\, \widehat{\mathbb{X}}] = \int_0^S Y_t\, d\widetilde{F}(t) + Y_S(1 - \widetilde{F}(S)) = \int_0^S (1 - \widetilde{F}(t))\, dY_t + Y_0$$

where the second integral is implicitly defined by integration by parts and

$$\widetilde{F}(t) := F_Z\left(\int_0^t \theta(\widehat{\mathbb{X}}|_{[0,s]})^2\, ds\right).$$

In particular, if Z has a density ϱ,

$$\mathbb{E}[Y_{\tau_\theta^r \wedge S}] = \mathbb{E}\left[\int_0^S Y_t \theta(\widehat{\mathbb{X}}|_{[0,t]})^2 \varrho\left(\int_0^t \theta(\widehat{\mathbb{X}}|_{[0,s]})^2\, ds\right)\, dt + Y_S(1 - \widetilde{F}(S))\right].$$

Proof. Recall that $\tau_\theta^r \in [0, T] \cup \{\infty\}$. For $t \in [0, \infty)$, we have

$$\mathbb{P}(\tau_\theta^r \leq t \,|\, \widehat{\mathbb{X}}) = \mathbb{P}\left(\int_0^{t\wedge T} \theta(\widehat{\mathbb{X}}|_{[0,s]})^2\, ds \geq Z \,|\, \widehat{\mathbb{X}}\right) = F_Z\left(\int_0^{t\wedge T} \theta(\widehat{\mathbb{X}}|_{[0,s]})^2\, ds\right) = \widetilde{F}(t)$$

and

$$\mathbb{P}(\tau_\theta^r = \infty \,|\, \widehat{\mathbb{X}}) = \mathbb{P}\left(\int_0^T \theta(\widehat{\mathbb{X}}|_{[0,s]})^2\, ds < Z \,|\, \widehat{\mathbb{X}}\right) = 1 - \widetilde{F}(T).$$

It follows that for $f\colon [0, \infty] \to \mathbb{R}$ integrable,

$$\mathbb{E}[f(\tau_\theta^r) \,|\, \widehat{\mathbb{X}}] = \int_0^T f(t)\, d\widetilde{F}(t) + f(\infty)(1 - \widetilde{F}(T))$$

and therefore

$$\mathbb{E}[Y_{\tau_\theta^r \wedge S} \,|\, \widehat{\mathbb{X}}] = \int_0^T Y_{t\wedge S}\, d\widetilde{F}(t) + Y_S(1 - \widetilde{F}(T)) = \int_0^S Y_t\, d\widetilde{F}(t) + Y_S(1 - \widetilde{F}(S)).$$

□

Remark 4.5. *In the deterministic case $Z = z > 0$ almost surely, we have*

$$\widetilde{F}(t) = \mathbb{1}_{[z,\infty)}\left(\int_0^t \theta(\widehat{\mathbb{X}}|_{[0,s]})^2\, ds\right),$$

thus

$$\mathbb{E}[Y_{\tau_\theta^r \wedge T}] = \mathbb{E}\left[\int_0^T \mathbb{1}_{[0,z)}\left(\int_0^t \theta(\widehat{\mathbb{X}}|_{[0,s]})^2\, ds\right)\, dY_t\right] + \mathbb{E}[Y_0].$$

DOI 10.20347/WIAS.PREPRINT.2790

Using the regularization by randomization, we can now prove that randomized stopping times induced by continuous stopping policies can, in fact, be approximated by randomized signature stopping times, in the sense of convergence of the optimal stopping problems.

Proposition 4.6. *Assume that Z has a continuous density ϱ and that $\mathbb{E}[\|Y\|_\infty] < \infty$. Then*

$$\sup_{\theta\in\mathcal{T}} \mathbb{E}[Y_{\tau^r_\theta \wedge T}] = \sup_{\theta\in\mathcal{T}_{\mathrm{sig}}} \mathbb{E}[Y_{\tau^r_\theta \wedge T}].$$

It follows that

$$\sup_{\theta\in\mathcal{T}} \mathbb{E}[Y_{\tau^r_\theta \wedge T}] = \sup_{l\in T((\mathbb{R}^{1+d})^*)} \mathbb{E}[Y_{\tau^r_l \wedge T}].$$

Proof. It is enough to show that $\sup_{\theta\in\mathcal{T}} \mathbb{E}[Y_{\tau^r_\theta \wedge T}] \le \sup_{\theta\in\mathcal{T}_{\mathrm{sig}}} \mathbb{E}[Y_{\tau^r_\theta \wedge T}]$. Let $\theta \in \mathcal{T}$. From Lemma 3.4, we know that for every $\varepsilon > 0$, there is a compact set $\mathcal{K} \subset \widehat{\Omega}^p_T$ such that for $A := \{\widehat{\mathbb{X}} \in \mathcal{K}\}$, we have $\mathbb{P}(A) \ge 1 - \varepsilon$, and a sequence $\theta_n \in \mathcal{T}_{\mathrm{sig}}$ such that

$$\lim_{n\to\infty} \sup_{\mathbb{X}\in\mathcal{K};t\in[0,T]} |\theta_n(\mathbb{X}|_{[0,t]}) - \theta(\mathbb{X}|_{[0,t]})| = 0. \tag{4.3}$$

Let $\mathcal{K}$ be such a compact set, the precise choice will be made later. Set

$$\widetilde{F}_n(t) = F_Z\left(\int_0^t \theta_n(\widehat{\mathbb{X}}|_{[0,s]})^2\, ds\right) \quad \text{and} \quad \widetilde{F}(t) = F_Z\left(\int_0^t \theta(\widehat{\mathbb{X}}|_{[0,s]})^2\, ds\right).$$

Then,

$$\begin{aligned} &|\mathbb{E}[Y_T(1-\widetilde{F}_n(T))\,;\,A] - \mathbb{E}[Y_T(1-\widetilde{F}(T))\,;\,A]| \\ &\le \mathbb{E}[|Y_T||\widetilde{F}_n(T) - \widetilde{F}(T)|\,;\,A] \\ &\le \mathbb{E}[|Y_T|] \sup_{\mathbb{X}\in\mathcal{K}} \left| F_Z\left(\int_0^T \theta_n(\mathbb{X}|_{[0,s]})^2\, ds\right) - F_Z\left(\int_0^T \theta(\mathbb{X}|_{[0,s]})^2\, ds\right)\right|. \end{aligned}$$

Since F_Z is continuous and uniformly continuous on compact sets,

$$\sup_{\mathbb{X}\in\mathcal{K}} \left| F_Z\left(\int_0^T \theta_n(\mathbb{X}|_{[0,s]})^2\, ds\right) - F_Z\left(\int_0^T \theta(\mathbb{X}|_{[0,s]})^2\, ds\right)\right| \to 0$$

as $n \to \infty$. Indeed: we first show that

$$\sup_{\mathbb{X}\in\mathcal{K};t\in[0,T]} |\theta_n(\mathbb{X}|_{[0,t]})^2 - \theta(\mathbb{X}|_{[0,t]})^2| \to 0 \tag{4.4}$$

as $n \to \infty$. Since $\sup_{n\ge1} \sup_{\mathbb{X}\in\mathcal{K};t\in[0,T]} |\theta_n(\mathbb{X}|_{[0,t]})| < \infty$, the functions θ and θ_n take their values in a compact set, hence (4.4) follows from (4.3). Property (4.3) also implies that

$$\sup_{\mathbb{X}\in\mathcal{K}} \left| \int_0^T \theta_n(\mathbb{X}|_{[0,s]})^2\, ds - \int_0^T \theta(\mathbb{X}|_{[0,s]})^2\, ds \right| \to 0$$

as $n \to \infty$. Using continuity of F_Z and uniform continuity on compact sets implies the claim. It follows that

$$\lim_{n\to\infty} |\mathbb{E}[Y_T(1-\widetilde{F}_n(T))\,;\,A] - \mathbb{E}[Y_T(1-\widetilde{F}(T))\,;\,A]| = 0.$$

Since $|\widetilde{F}_n(T) - \widetilde{F}(T)| \leq 2$,

$$|\mathbb{E}[Y_T(1-\widetilde{F}_n(T))\,;\,A^c] - \mathbb{E}[Y_T(1-\widetilde{F}(T))\,;\,A^c]| \leq 2\mathbb{E}[|Y_T|\,;\,A^c]$$

and this quantity can me made arbitrarily small by the choice of $\mathcal{K}$.

With the same arguments, we can show that

$$\left|\mathbb{E}\left[\int_0^T Y_t\left(\theta_n(\widehat{\mathbb{X}}|_{[0,t]})^2\varrho\left(\int_0^t \theta_n(\widehat{\mathbb{X}}|_{[0,s]})^2\,ds\right) - \theta(\widehat{\mathbb{X}}|_{[0,t]})^2\varrho\left(\int_0^t \theta(\widehat{\mathbb{X}}|_{[0,s]})^2\,ds\right)\right)\,dt\right]\right| \to 0$$

as $n \to \infty$ which implies the claim. □

Finally, we note that we do not need randomization for the approximation by stopping times based on signature stopping policies to work.

Definition 4.7. *We define a* signature stopping time *as the hitting time of a closed half-plane, i.e. for $l \in T((\mathbb{R}^{1+d})^*)$, we set*

$$\tau_l := \inf\left\{t \in [0,T]\,:\,\langle l, \widehat{\mathbb{X}}^{<\infty}_{0,t}\rangle \geq 1\right\}.$$

Theorem 4.8. *Given $\mathbb{E}[\|Y\|_\infty] < \infty$, we have*

$$\sup_{l\in T((\mathbb{R}^{1+d})^*)} \mathbb{E}[Y_{\tau_l\wedge T}] = \sup_{\tau\in\mathcal{S}} \mathbb{E}[Y_{\tau\wedge T}].$$

Proof. Using Proposition 4.2 and 4.6, it suffices to show that

$$\sup_{l\in T((\mathbb{R}^{1+d})^*)} \mathbb{E}[Y_{\tau_l^r\wedge T}] \leq \sup_{l\in T((\mathbb{R}^{1+d})^*)} \mathbb{E}[Y_{\tau_l\wedge T}].$$

Choose $l \in T((\mathbb{R}^{1+d})^*)$. Then

$$\mathbb{E}[Y_{\tau_l^r\wedge T}\,|\,\widehat{\mathbb{X}}] = \int_0^\infty Y_{\tau_z\wedge T}\,\mathbb{P}_Z(dz)$$

where

$$\tau_z := \inf\left\{t \geq 0\,:\,\int_0^{t\wedge T}\langle l, \widehat{\mathbb{X}}^{<\infty}_{0,s}\rangle^2\,ds \geq z\right\} = \inf\left\{t \in [0,T]\,:\,\langle (l \sqcup\!\sqcup\, l)\mathbf{1}/z, \widehat{\mathbb{X}}^{<\infty}_{0,t}\rangle \geq 1\right\}$$

which is a signature stopping time for every $z > 0$. Taking expectation, we obtain

$$\mathbb{E}[Y_{\tau_l^r\wedge T}] = \int_0^\infty \mathbb{E}[Y_{\tau_z\wedge T}]\,\mathbb{P}_Z(dz) \leq \sup_{\ell\in T((\mathbb{R}^d)^*)} \mathbb{E}[Y_{\tau_\ell\wedge T}]$$

as claimed. □

Remark 4.9. *In the case of X being a standard Markov process in $\mathbb{R}^d$ and $Y_t = G(t, X_t)$ for a continuous function G, it is known that*

$$\sup_{\tau\in\mathcal{S}} \mathbb{E}[Y_{\tau\wedge T}] = \sup_{\tau\in\mathfrak{D}} \mathbb{E}[Y_{\tau\wedge T}]$$

where $\mathfrak{D}$ denotes the set of all hitting times of closed sets in $\mathbb{R}^{1+d}$ of the process $t \mapsto (t, X_t)$ [Shi08, Corollary 3 on p. 129]. Our Theorem can be seen as an extension of this classical result to non-Markovian processes.

5 Approximation of the stopping problem

In this section, we will study randomized signature stopping times for $Z \sim \mathrm{Exp}(1)$. In the former section, we have seen that

$$\sup_{\tau\in\mathcal{S}} \mathbb{E}[Y_{\tau\wedge T}] = \sup_{l\in T((\mathbb{R}^{1+d})^*)} \mathbb{E}[Y_{\tau_l^r\wedge T}] = \sup_{l\in T((\mathbb{R}^{1+d})^*)} \mathbb{E}\left[\int_0^T \exp\left(-\int_0^t \langle l, \widehat{\mathbb{X}}^{<\infty}_{0,s}\rangle^2\, ds\right) dY_t\right] + \mathbb{E}[Y_0]. \tag{5.1}$$

Recall that for group-like elements $\mathbf{a} \in G(V)$ as defined in (2.2), polynomials of linear functionals in $\mathbf{a}$ can be expressed in terms of shuffle products of the linear functionals themselves, see (2.3). Consider the main term on the right-hand side of (5.1): we have

- The innermost term $\langle l, \widehat{\mathbb{X}}^{<\infty}_{0,s}\rangle^2$ is a polynomial of a linear functional of the signature. It can, therefore, be expressed as a linear functional of the signature, more precisely, $\langle l, \widehat{\mathbb{X}}^{<\infty}_{0,s}\rangle^2 = \langle l \sqcup\!\sqcup\, l, \widehat{\mathbb{X}}^{<\infty}_{0,s}\rangle$.
- Given an element of the signature, integrating against a component of the underlying path produces another element of the signature. Concretely,

  $$\int_0^t \langle l, \widehat{\mathbb{X}}^{<\infty}_{0,s}\rangle^2 ds = \langle (l \sqcup\!\sqcup\, l)\mathbf{1}, \widehat{\mathbb{X}}^{<\infty}_{0,t}\rangle,$$

  recalling that the time-component of our driving path $\widehat{X}_t = (t, X_t)$ was associated with the letter $\mathbf{1}$, see Remark 2.2.
- Next we need to apply the exponential function to $\langle (l \sqcup\!\sqcup\, l)\mathbf{1}, \widehat{\mathbb{X}}^{<\infty}_{0,t}\rangle$. Unfortunately, the exponential function is not a polynomial, so we cannot directly apply the shuffle product. However, as we shall see below, there is a corresponding *exponential shuffle*, which comes with certain restrictions. Nonetheless, we shall see that we will still obtain a linear functional of the signature for our purposes.
- Finally, we integrate against Y and take the expectation. **If** Y can itself be represented as a linear functional of the signature, integrating another linear functional of the signature against Y will result in yet another linear functional of the signature. In this case, we can finally interchange the expectation, and the right-hand side of (5.1) can be represented as a $\sup$ over a linear functional of the *expected signature* $\mathbb{E}[\widehat{\mathbb{X}}^{<\infty}_{0,T}]$ of $\widehat{X}$.

In the remainder of this section, we will follow through with this program.

We start with a definition of an exponential function based on the shuffle product.

Definition 5.1. *Let V be finite-dimensional $\mathbb{R}$-vector space. For $l \in T(V^*)$ with $l = a_0\varnothing + \widetilde{l}$ and $\langle \widetilde{l}, \mathbf{1}\rangle = 0$ we define the* exponential shuffle

$$\exp^{\sqcup\!\sqcup}(l) := \exp(a_0)\exp^{\sqcup\!\sqcup}(\widetilde{l}), \quad \textit{where} \quad \exp^{\sqcup\!\sqcup}(\widetilde{l}) := \sum_{r=0}^{\infty} \frac{1}{r!}\widetilde{l}^{\sqcup\!\sqcup r}. \tag{5.2}$$

Since obviously $\pi_{\le N}(\widetilde{l}^{\sqcup\!\sqcup r}) = 0$ for $r > N$, the infinite sum is well defined as an element in the extended tensor algebra $T((V^*))$. One may straightforwardly check that $\exp^{\sqcup\!\sqcup}(\widetilde{l}_1+\widetilde{l}_2) = \exp^{\sqcup\!\sqcup}(\widetilde{l}_1)\exp^{\sqcup\!\sqcup}(\widetilde{l}_2)$ for $\widetilde{l}_1, \widetilde{l}_2 \in T(V^*)$ such that $\langle \widetilde{l}_1, \mathbf{1}\rangle = \langle \widetilde{l}_2, \mathbf{1}\rangle = 0$. Thus, in particular, one has

$$\exp^{\sqcup\!\sqcup}(l_1 + l_2) = \exp^{\sqcup\!\sqcup}(l_1)\exp^{\sqcup\!\sqcup}(l_2) \quad \text{for all } l_1, l_2 \in T(V^*). \tag{5.3}$$

We can now prove that the exponential shuffle linearizes the exponential function for group-like elements. In this context, keep in mind that $\exp^{\sqcup\!\sqcup}(l) \notin T(V^*)$ and, hence, it is not a well-defined linear functional on $T((V))$. It is, however, trivially well-defined as a linear functional on $T(V)$, and, hence, can be applied to any projection $\pi_{\leq N}(\mathbf{v})$, $\mathbf{v} \in T((V))$. In addition, as the lemma shows, we can apply $\exp^{\sqcup\!\sqcup}(l)$ to group-like elements.

Lemma 5.2. *Let $l \in T(V^*)$ and $\mathbf{g} \in G(V)$. One then has*

$$|\exp(\langle l, \mathbf{g}\rangle) - \langle \exp^{\sqcup\!\sqcup}(l), \pi_{\leq N}(\mathbf{g})\rangle| \leq 4\exp(\langle l, 1\rangle)\frac{\left(|l||\pi_{\leq \deg(l)}(\mathbf{g})|\right)^{\lfloor N/\deg(l)\rfloor+1}}{(\lfloor N/\deg(l)\rfloor+1)!}$$

for $N > 2\deg(l)|l||\pi_{\leq \deg(l)}(\mathbf{g})|$.

Proof. Let us write $l = a_0\varnothing + \widetilde{l}$ with $\langle \widetilde{l}, \mathbf{1}\rangle = 0$, where for mutually different words $w_1, ..., w_n$,

$$\widetilde{l} = \lambda_1 w_1 + ... + \lambda_n w_n, \text{ and set } M := \deg(l) = \max_{1\leq i\leq n}\deg(w_i), \quad m := \min_{1\leq i\leq n}\deg(w_i) \geq 1.$$

We then have

$$\pi_{\leq N}\left(\exp^{\sqcup\!\sqcup}(\widetilde{l})\right) = \sum_{r=0}^{\lfloor N/M\rfloor}\frac{\widetilde{l}^{\sqcup\!\sqcup r}}{r!} + \pi_N\left(\sum_{r=\lfloor N/\deg(l)\rfloor+1}^{\lfloor N/m\rfloor}\frac{\widetilde{l}^{\sqcup\!\sqcup r}}{r!}\right) \in T^N(V^*) \quad \text{for any} \quad N \geq 1.$$

Hence,

$$\begin{aligned}\langle \pi_{\leq N}\left(\exp^{\sqcup\!\sqcup}(l)\right), \mathbf{g}\rangle &= \exp(a_0)\sum_{r=0}^{\lfloor N/M\rfloor}\frac{\langle \widetilde{l}, \mathbf{g}\rangle^r}{r!} + \exp(a_0)\sum_{r=\lfloor N/M\rfloor+1}^{\lfloor N/m\rfloor}\frac{1}{r!}\langle \pi_N\left(\widetilde{l}^{\sqcup\!\sqcup r}\right), \mathbf{g}\rangle \\ &=: \exp(a_0)\sum_{r=0}^{\lfloor N/M\rfloor}\frac{\langle \widetilde{l}, \mathbf{g}\rangle^r}{r!} + R_N^{(1)},\end{aligned}$$

and since $\mathbf{g} \in G(V)$ it holds that

$$\begin{aligned}\langle \pi_{\leq N}\left(\widetilde{l}^{\sqcup\!\sqcup r}\right), \mathbf{g}\rangle &= \sum_{\substack{i_1,...,i_r=1 \\ \deg(w_{i_1})+...+\deg(w_{i_r})\leq N}}^{n} \lambda_{i_1}...\lambda_{i_r}\langle w_{i_1} \sqcup\!\sqcup ... \sqcup\!\sqcup w_{i_r}, \mathbf{g}\rangle \\ &= \sum_{\substack{i_1,...,i_r=1 \\ \deg(w_{i_1})+...+\deg(w_{i_r})\leq N}}^{n} \langle \lambda_{i_1}w_{i_1}, \mathbf{g}\rangle ... \langle \lambda_{i_r}w_{i_r}, \mathbf{g}\rangle.\end{aligned}$$

Hence we have

$$\begin{aligned}\left|\langle \pi_{\leq N}\left(\widetilde{l}^{\sqcup\!\sqcup r}\right), \mathbf{g}\rangle\right| &\leq \sum_{i_1,...,i_r=1}^{n}|\langle \lambda_{i_1}w_{i_1}, \mathbf{g}\rangle| ... |\langle \lambda_{i_r}w_{i_r}, \mathbf{g}\rangle| = \left(\sum_{i=1}^{n}|\langle \lambda_i w_i, \mathbf{g}\rangle|\right)^r \\ &\leq |\widetilde{l}|^r\left|\pi_{\leq \deg(l)}(\mathbf{g})\right|^r, \quad \text{and so}\end{aligned}$$

$$\left|R_N^{(1)}\right| \leq \exp(a_0)\sum_{r=\lfloor N/M\rfloor+1}^{\infty}\frac{|\widetilde{l}|^r\left|\pi_{\leq \deg(l)}(\mathbf{g})\right|^r}{r!} \leq 2\exp(a_0)\frac{\left(|\widetilde{l}||\pi_{\leq \deg(l)}(\mathbf{g})|\right)^{\lfloor N/M\rfloor+1}}{(\lfloor N/M\rfloor+1)!}$$

for $N > N_{l,\mathbf{g}} := 2M|\widetilde{l}|\pi_{\leq \deg(l)}(\mathbf{g})|$. One further has (note that $\mathbf{g}_0 = \mathbf{1}$), due to a similar estimation,

$$\exp(\langle l, \mathbf{g}\rangle) = \exp(a_0)\exp(\langle \widetilde{l}, \mathbf{g}\rangle) = \exp(a_0) \sum_{r=0}^{\lfloor N/M \rfloor} \frac{\langle \widetilde{l}, \mathbf{g}\rangle^r}{r!} + R_N^{(2)} \quad \text{with}$$

$$\left|R_N^{(2)}\right| \leq 2\exp(a_0) \frac{\left|\langle \widetilde{l}, \mathbf{g}\rangle\right|^{\lfloor N/M \rfloor + 1}}{(\lfloor N/M \rfloor + 1)!} \leq 2\exp(a_0) \frac{\left(\left|\widetilde{l}\right| |\pi_{\leq \deg(l)}(\mathbf{g})|\right)^{\lfloor N/M \rfloor + 1}}{(\lfloor N/M \rfloor + 1)!}$$

for $N > N_{l,\mathbf{g}}$. Finally, by noting that $\langle \pi_{\leq N}(\exp^{ш}(l)), \mathbf{g}\rangle = \langle \exp^{ш}(l), \pi_{\leq N}(\mathbf{g})\rangle$, and then taking all together we obtain the stated result. □

Remark 5.3. *The equation* $\langle \exp^{ш}(l), \mathbf{g}\rangle = \exp(\langle l, \mathbf{g}\rangle)$ *is confusing at first glance, because* $\mathbf{g} \mapsto \langle \exp^{ш}(l), \mathbf{g}\rangle$ seems *linear, whereas* $\mathbf{g} \mapsto \exp(\langle l, \mathbf{g}\rangle)$ *clearly is not. Note, however, that* $\exp^{ш}(l) \in T((V^*))$ *and, hence, does not define a linear map on* $T((V))$*. Indeed, the group* $G(V)$ *is not closed under linear combination, and, hence, Lemma 5.2 does simply not apply to a linear combination of elements* $\mathbf{g}_1, \mathbf{g}_2 \in G(V)$.

The exponential shuffle satisfies a differential equation, which we shall use later. Note that terms of the form $\langle \exp^{ш}(l\mathbf{1}), \widehat{\mathbb{X}}_{0,t}^{\leq N}\rangle$ are (classically) differentiable in t.

Lemma 5.4. *For every polynomial* $l = \lambda_1 w_1 + \ldots + \lambda_n w_n \in T((\mathbb{R}^{1+d})^*)$,

$$\frac{d}{dt}\langle \exp^{ш}(l\mathbf{1}), \widehat{\mathbb{X}}_{0,t}^{\leq N}\rangle = \sum_{i=1}^{n} \langle \lambda_i w_i, \widehat{\mathbb{X}}_{0,t}^{<\infty}\rangle \langle \exp^{ш}(l\mathbf{1}), \widehat{\mathbb{X}}_{0,t}^{\leq N - \deg(w_i) - 1}\rangle.$$

Proof. Note that

$$\frac{d}{dt}\langle w\mathbf{1}, \widehat{\mathbb{X}}_{0,t}^{<\infty}\rangle = \frac{d}{dt}\int_0^t \langle w, \widehat{\mathbb{X}}_{0,s}^{<\infty}\rangle \, ds = \langle w, \widehat{\mathbb{X}}_{0,t}^{<\infty}\rangle$$

for every word w. Hence, for $l = \lambda_1 w_1 + \ldots + \lambda_n w_n$, one always has $\langle l\mathbf{1}, \mathbf{1}\rangle = 0$ and so by (5.2),

$$\begin{aligned}
&\frac{d}{dt}\langle \exp^{ш}(l\mathbf{1}), \widehat{\mathbb{X}}_{0,t}^{\leq N}\rangle \\
&= \frac{d}{dt} \sum_{0 \leq k_1 \deg(w_1\mathbf{1}) + \ldots + k_n \deg(w_n\mathbf{1}) \leq N} \frac{\langle \lambda_1 w_1\mathbf{1}, \widehat{\mathbb{X}}_{0,t}^{<\infty}\rangle^{k_1}}{k_1!} \cdots \frac{\langle \lambda_n w_n\mathbf{1}, \widehat{\mathbb{X}}_{0,t}^{<\infty}\rangle^{k_n}}{k_n!} \\
&= \sum_{0 \leq k_1 \deg(w_1\mathbf{1}) + \ldots + k_n \deg(w_n\mathbf{1}) \leq N} \langle \lambda_1 w_1, \widehat{\mathbb{X}}_{0,t}^{<\infty}\rangle \frac{\langle \lambda_1 w_1\mathbf{1}, \widehat{\mathbb{X}}_{0,t}^{<\infty}\rangle^{k_1 - 1}}{(k_1 - 1)!} \frac{\langle \lambda_2 w_2\mathbf{1}, \widehat{\mathbb{X}}_{0,t}^{<\infty}\rangle^{k_2}}{k_2!} \cdots \frac{\langle \lambda_n w_n\mathbf{1}, \widehat{\mathbb{X}}_{0,t}^{<\infty}\rangle^{k_n}}{k_n!} \\
&\quad + \ldots + \langle \lambda_n w_n, \widehat{\mathbb{X}}_{0,t}^{<\infty}\rangle \frac{\langle \lambda_1 w_1\mathbf{1}, \widehat{\mathbb{X}}_{0,t}^{<\infty}\rangle^{k_1}}{k_1!} \cdots \frac{\langle \lambda_{n-1} w_{n-1}\mathbf{1}, \widehat{\mathbb{X}}_{0,t}^{<\infty}\rangle^{k_{n-1}}}{k_{n-1}!} \frac{\langle \lambda_n w_n\mathbf{1}, \widehat{\mathbb{X}}_{0,t}^{<\infty}\rangle^{k_n - 1}}{(k_n - 1)!}
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{0 \leq k_1 \deg(w_1\mathbf{1}) + \ldots + k_n \deg(w_n\mathbf{1}) \leq N} \frac{\langle \lambda_1 w_1\mathbf{1}, \widehat{\mathbb{X}}_{0,t}^{<\infty}\rangle^{k_1 - 1}}{(k_1 - 1)!} \frac{\langle \lambda_2 w_2\mathbf{1}, \widehat{\mathbb{X}}_{0,t}^{<\infty}\rangle^{k_2}}{k_2!} \cdots \frac{\langle \lambda_n w_n\mathbf{1}, \widehat{\mathbb{X}}_{0,t}^{<\infty}\rangle^{k_n}}{k_n!} \\
&= \sum_{0 \leq (k_1 + 1) \deg(w_1\mathbf{1}) + \ldots + k_n \deg(w_n\mathbf{1}) \leq N} \frac{\langle \lambda_1 w_1\mathbf{1}, \widehat{\mathbb{X}}_{0,t}^{<\infty}\rangle^{k_1}}{k_1!} \frac{\langle \lambda_2 w_2\mathbf{1}, \widehat{\mathbb{X}}_{0,t}^{<\infty}\rangle^{k_2}}{k_2!} \cdots \frac{\langle \lambda_n w_n\mathbf{1}, \widehat{\mathbb{X}}_{0,t}^{<\infty}\rangle^{k_n}}{k_n!} \\
&= \langle \exp^{ш}(l\mathbf{1}), \widehat{\mathbb{X}}_{0,t}^{\leq N - \deg(w_1\mathbf{1})}\rangle. \square
\end{aligned}$$

We are now ready to formulate the main result of this section. Consider the optimization problem (5.1), which we modify by expressing the exponential by the exponential shuffle. Then we obtain convergence to the value of the optimal stopping problem. The proof requires us to localize w.r.t. the rough path metric. Other than that, the below formulation is now essentially implementable: In particular, the result is formulated in terms of truncated signatures, which is necessary also from a numerical point of view.

Theorem 5.5. *For given $\kappa > 0$, we define the stopping time*

$$S = S_\kappa = \inf\{t \geq 0 \,:\, \|\widehat{\mathbb{X}}\|_{p-var;[0,t]} \geq \kappa\} \wedge T.$$

Assume $Z \sim \mathrm{Exp}(1)$ and $\mathbb{E}\left[\|Y\|_\infty\right] < \infty$. Then

$$\sup_{l\in T((\mathbb{R}^{1+d})^*)} \mathbb{E}[Y_{\tau_l^r \wedge T}] = \lim_{\kappa\to\infty}\lim_{K\to\infty}\lim_{N\to\infty}\sup_{|l|+\deg(l)\leq K} \mathbb{E}\left[\int_0^{S_\kappa} \langle \exp^{\sqcup\!\sqcup}(-(l \sqcup\!\sqcup l)1), \widehat{\mathbb{X}}^{\leq N}_{0,t}\rangle \, dY_t\right] + \mathbb{E}[Y_0] \tag{5.4}$$

where the first two limit signs may be interchanged.

Proof. To ease notation, assume that $Y_0 = 0$. Since

$$|Y_{\tau_l\wedge T} - Y_{\tau_l\wedge S}| \leq \sup_{|t-s|\leq T-S} |Y_t - Y_s| \to 0$$

for every l as $\kappa \to \infty$ and

$$\lim_{K\to\infty} \sup_{|l|+\deg(l)\leq K} \mathbb{E}[Y_{\tau_l\wedge\widehat{S}}] = \sup_{l\in T((\mathbb{R}^{1+d})^*)} \mathbb{E}[Y_{\tau_l\wedge\widehat{S}}],$$

with $\widehat{S}$ being either S or T, it follows that

$$\sup_{l\in T((\mathbb{R}^{1+d})^*)} \mathbb{E}[Y_{\tau_l\wedge T}] = \lim_{\kappa\to\infty}\lim_{K\to\infty}\sup_{|l|+\deg(l)\leq K} \mathbb{E}[Y_{\tau_l\wedge S}] = \lim_{K\to\infty}\lim_{\kappa\to\infty}\sup_{|l|+\deg(l)\leq K} \mathbb{E}[Y_{\tau_l\wedge S}]. \tag{5.5}$$

Now fix κ, K and l with $|l| + \deg(l) \leq K$. Recall the estimate

$$\left|\int_0^T f(s)\, dg(s)\right| \leq T\|f'\|_\infty \|g\|_\infty + |f(T)g(T) - f(0)g(0)|.$$

Note that

$$\exp\left(-\int_0^t \langle l, \widehat{\mathbb{X}}^{<\infty}_{0,s}\rangle^2 \, ds\right) = \exp(-\langle (l \sqcup\!\sqcup l)1, \widehat{\mathbb{X}}^{<\infty}_{0,t}\rangle).$$

Fix N. Then

$$\left|\mathbb{E}\left[\int_0^S \exp(-\langle (l \sqcup\!\sqcup l)1, \widehat{\mathbb{X}}^{<\infty}_{0,t}\rangle)\, dY_t\right] - \mathbb{E}\left[\int_0^S \langle \exp^{\sqcup\!\sqcup}(-(l \sqcup\!\sqcup l)1), \widehat{\mathbb{X}}^{\leq N}_{0,t}\rangle \, dY_t\right]\right|$$
$$\leq (1+T)\mathbb{E}\left[\|Y\|_\infty \|\exp(-\langle (l \sqcup\!\sqcup l)1, \widehat{\mathbb{X}}^{<\infty}_{0,\cdot}\rangle) - \langle \exp^{\sqcup\!\sqcup}(-(l \sqcup\!\sqcup l)1), \widehat{\mathbb{X}}^{\leq N}_{0,\cdot}\rangle\|_{\mathcal{C}^1[0,S]}\right].$$

Using Lemma 5.2,

$$\|\exp(-\langle (l \sqcup\!\sqcup l)1, \widehat{\mathbb{X}}^{<\infty}_{0,\cdot}\rangle) - \langle \exp^{\sqcup\!\sqcup}(-(l \sqcup\!\sqcup l)1), \widehat{\mathbb{X}}^{\leq N}_{0,\cdot}\rangle\|_{\infty;[0,S]} \leq 4 \sup_{t\in[0,S]} \frac{(|(l \sqcup\!\sqcup l)1||\widehat{\mathbb{X}}^{\leq 2\deg(l)+1}_{0,t}|)^M}{M!}.$$

where $M = \lfloor N/(2\deg(l)+1)\rfloor + 1$ provided N is sufficiently large. Clearly, $|(l \sqcup l)1| \le C_K$. Using Lyons' Extension theorem [LCL07, Theorem 3.7], we can estimate

$$\sup_{t\in[0,S]} |\widehat{\mathbb{X}}^{\le 2\deg(l)+1}_{0,t}| \le \|\widehat{\mathbb{X}}^{\le 2\deg(l)+1}\|_{p-\mathrm{var};[0,S]} \le C(1+\|\widehat{\mathbb{X}}\|_{p-\mathrm{var};[0,S]})^{2K+1} = C(1+\kappa)^{2K+1}.$$

Therefore, we obtain an estimate of the form

$$\|\exp(-\langle (l\sqcup l)1, \widehat{\mathbb{X}}^{<\infty}_{0,\cdot}\rangle) - \langle \exp^{\sqcup}(-(l\sqcup l)1), \widehat{\mathbb{X}}^{\le N}_{0,\cdot}\rangle\|_{\infty;[0,S]} \le \frac{C^M}{M!}$$

for a deterministic constant C.

Next, we consider the derivatives. Set $\widetilde{l} = -(l \sqcup l)$ and assume $\widetilde{l} = \lambda_1 w_1 + \ldots + \lambda_k w_k$. Clearly,

$$\frac{d}{dt}\exp(\langle \widetilde{l}1, \widehat{\mathbb{X}}^{<\infty}_{0,t}\rangle) = \langle \widetilde{l}, \widehat{\mathbb{X}}^{<\infty}_{0,t}\rangle \exp(\langle \widetilde{l}1, \widehat{\mathbb{X}}^{<\infty}_{0,t}\rangle)$$

and Lemma 5.4 shows that

$$\frac{d}{dt}\langle \exp^{\sqcup}(\widetilde{l}1), \widehat{\mathbb{X}}^{\le N}_{0,t}\rangle = \sum_{i=1}^{k} \langle \lambda_i w_i, \widehat{\mathbb{X}}^{<\infty}_{0,t}\rangle \langle \exp^{\sqcup}(\widetilde{l}1), \widehat{\mathbb{X}}^{\le N-\deg(w_i)-1}_{0,t}\rangle.$$

Thus for $t \in [0,S]$,

$$\begin{aligned}&\left|\frac{d}{dt}\left(\exp(\langle \widetilde{l}1, \widehat{\mathbb{X}}^{<\infty}_{0,t}\rangle) - \langle \exp^{\sqcup}(\widetilde{l}1), \widehat{\mathbb{X}}^{\le N}_{0,t}\rangle\right)\right| \\ &\le \sum_{i=1}^{k} |\langle \lambda_i w_i, \widehat{\mathbb{X}}^{<\infty}_{0,t}\rangle| \left|\exp(\langle \widetilde{l}1, \widehat{\mathbb{X}}^{<\infty}_{0,t}\rangle) - \langle \exp^{\sqcup}(\widetilde{l}1), \widehat{\mathbb{X}}^{\le N-\deg(w_i)-1}_{0,t}\rangle\right|.\end{aligned}$$

Using Lyons' Extension theorem,

$$|\langle \lambda_i w_i, \widehat{\mathbb{X}}^{<\infty}_{0,t}\rangle| \le C|\lambda_i| \|\widehat{\mathbb{X}}\|^{\deg(w_i)}_{p-\mathrm{var};[0,S]} \le C|\lambda_i|$$

for a deterministic constant $C > 0$. Lemma 5.2 implies that for N sufficiently large,

$$\left|\exp(\langle \widetilde{l}1, \widehat{\mathbb{X}}^{<\infty}_{0,t}\rangle) - \langle \exp^{\sqcup}(\widetilde{l}1), \widehat{\mathbb{X}}^{\le N-\deg(w_i)-1}_{0,t}\rangle\right| \le \frac{C^M}{M!}$$

for a deterministic constant $C > 0$ and $M \to \infty$ as $N \to \infty$. It follows that also

$$\left\|\frac{d}{dt}\left(\exp(-\langle (l\sqcup l)1, \widehat{\mathbb{X}}^{<\infty}_{0,t}\rangle) - \langle \exp^{\sqcup}(-(l\sqcup l)1), \widehat{\mathbb{X}}^{\le N}_{0,t}\rangle\right)\right\|_{\infty;[0,S]} \le \frac{C^M}{M!}.$$

This implies that

$$\sup_{|l|+\deg(l)\le K}\left|\mathbb{E}\left[\int_0^S \exp(-\langle (l\sqcup l)1, \widehat{\mathbb{X}}^{<\infty}_{0,t}\rangle)\,dY_t\right] - \mathbb{E}\left[\int_0^S \langle \exp^{\sqcup}(-(l\sqcup l)1), \widehat{\mathbb{X}}^{\le N}_{0,t}\rangle\,dY_t\right]\right| \to 0$$

as $N \to \infty$ and, in particular,

$$\begin{aligned}&\lim_{N\to\infty}\sup_{|l|+\deg(l)\le K}\mathbb{E}\left[\int_0^S \langle \exp^{\sqcup}(-(l\sqcup l)1), \widehat{\mathbb{X}}^{\le N}_{0,t}\rangle\,dY_t\right] \\ &= \sup_{|l|+\deg(l)\le K}\mathbb{E}\left[\int_0^S \exp(-\langle (l\sqcup l)1, \widehat{\mathbb{X}}^{<\infty}_{0,t}\rangle)\,dY_t\right] = \sup_{|l|+\deg(l)\le K}\mathbb{E}[Y_{\tau^r_l\wedge S}].\end{aligned}$$

Together with (5.5), this proves (5.4). □

Often, one is interested to solve the stopping problem for specific functionals of the underlying process X. In the next corollary, we consider a particular example. To simplify the exposition, we will consider the case $d = 1$ only. The generalization to arbitrary dimensions d is straightforward.

Corollary 5.6. *Assume $d = 1$ and that*

$$Y_t = G(X_t) + \int_0^t L(X_s)\, ds$$

for polynomials G and L. Then

$$\begin{aligned}&\sup_{l \in T((\mathbb{R}^{1+d})^*)} \mathbb{E}[Y_{\tau_l^r \wedge T}] \\ &= \lim_{\kappa\to\infty}\lim_{K\to\infty}\lim_{N\to\infty} \sup_{|l|+\deg(l)\le K} \langle (\exp^{\sqcup\!\sqcup}(-(l \sqcup\!\sqcup l)1) \sqcup\!\sqcup G'^{\sqcup\!\sqcup}(2))2 + (\exp^{\sqcup\!\sqcup}(-(l \sqcup\!\sqcup l)1) \sqcup\!\sqcup L^{\sqcup\!\sqcup}(2))1, \mathbb{E}[\widehat{\mathbb{X}}^{\le N}_{0,S}]\rangle \\ &\quad + \mathbb{E}[Y_0].\end{aligned}$$

In particular, if $d = 1$ and $X_0 = 0$,

$$\sup_{l \in T((\mathbb{R}^2)^*)} \mathbb{E}[X_{\tau_l^r \wedge T}] = \lim_{\kappa\to\infty}\lim_{K\to\infty}\lim_{N\to\infty} \sup_{|l|+\deg(l)\le K} \langle \exp^{\sqcup\!\sqcup}(-(l \sqcup\!\sqcup l)1)2, \mathbb{E}[\widehat{\mathbb{X}}^{\le N}_{0,S}]\rangle. \tag{5.6}$$

Proof. We have

$$\begin{aligned}\int_0^S \langle \exp^{\sqcup\!\sqcup}(-(l \sqcup\!\sqcup l)1), \widehat{\mathbb{X}}^{\le N}_{0,t}\rangle\, dY_t &= \int_0^S \langle \exp^{\sqcup\!\sqcup}(-(l \sqcup\!\sqcup l)1), \widehat{\mathbb{X}}^{\le N}_{0,t}\rangle G'(X_t)\, dX_t \\ &\quad + \int_0^S \langle \exp^{\sqcup\!\sqcup}(-(l \sqcup\!\sqcup l)1), \widehat{\mathbb{X}}^{\le N}_{0,t}\rangle L(X_t)\, dt.\end{aligned}$$

Since G' is a polynomial,

$$\begin{aligned}\int_0^S \langle \exp^{\sqcup\!\sqcup}(-(l \sqcup\!\sqcup l)1), \widehat{\mathbb{X}}^{\le N}_{0,t}\rangle G'(X_t)\, dX_t &= \int_0^S \langle \exp^{\sqcup\!\sqcup}(-(l \sqcup\!\sqcup l)1), \widehat{\mathbb{X}}^{\le N}_{0,t}\rangle \langle G'^{\sqcup\!\sqcup}(2), \widehat{\mathbb{X}}^{<\infty}_{0,t}\rangle\, dX_t \\ &= \int_0^S \langle \pi_{\le N}(\exp^{\sqcup\!\sqcup}(-(l \sqcup\!\sqcup l)1)) \sqcup\!\sqcup G'^{\sqcup\!\sqcup}(2), \widehat{\mathbb{X}}^{<\infty}_{0,t}\rangle\, dX_t \\ &= \int_0^S \langle \pi_{\le N+\deg(G')}(\exp^{\sqcup\!\sqcup}(-(l \sqcup\!\sqcup l)1) \sqcup\!\sqcup G'^{\sqcup\!\sqcup}(2)), \widehat{\mathbb{X}}^{<\infty}_{0,t}\rangle\, dX_t \\ &= \langle (\exp^{\sqcup\!\sqcup}(-(l \sqcup\!\sqcup l)1) \sqcup\!\sqcup G'^{\sqcup\!\sqcup}(2))2, \widehat{\mathbb{X}}^{\le N+\deg(G')+1}_{0,S}\rangle.\end{aligned}$$

Similarly, since V is a polynomial,

$$\int_0^S \langle \exp^{\sqcup\!\sqcup}(-(l \sqcup\!\sqcup l)1), \widehat{\mathbb{X}}^{\le N}_{0,t}\rangle L(X_t)\, dt = \langle (\exp^{\sqcup\!\sqcup}(-(l \sqcup\!\sqcup l)1) \sqcup\!\sqcup L^{\sqcup\!\sqcup}(2))1, \widehat{\mathbb{X}}^{\le N+\deg(V)+1}_{0,S}\rangle.$$

Taking expectation, we obtain

$$\begin{aligned}\mathbb{E}\left[\int_0^S \langle \exp^{\sqcup\!\sqcup}(-(l \sqcup\!\sqcup l)1), \widehat{\mathbb{X}}^{\le N}_{0,t}\rangle\, dY_t\right] &= \langle (\exp^{\sqcup\!\sqcup}(-(l \sqcup\!\sqcup l)1) \sqcup\!\sqcup G'^{\sqcup\!\sqcup}(2))2, \mathbb{E}[\widehat{\mathbb{X}}^{\le N+\deg(G')+1}_{0,S}]\rangle \\ &\quad + \langle (\exp^{\sqcup\!\sqcup}(-(l \sqcup\!\sqcup l)1) \sqcup\!\sqcup L^{\sqcup\!\sqcup}(2))1, \mathbb{E}[\widehat{\mathbb{X}}^{\le N+\deg(V)+1}_{0,S}]\rangle.\end{aligned}$$

Using Theorem 5.5, we can deduce the result. □

Remark 5.7. *Note that similar formulas are available whenever Y is roughly given as a polynomial of the signature. We restrict ourselves to a representative class of examples below. We note here that payoffs of American options usually cannot be exactly represented in such a way. In particular, for the standard American (put) option, we have $Y_t = (K - X_t)_+$ for some $K > 0$, where X denotes the underlying asset price process. If we want to price American options using signature stopping rules, we have two possible remedies. We can approximate the payoff function by polynomials, which would allow us to directly apply Corollary 5.6. Alternatively, we can attach Y to the path X, i.e., consider $\widetilde{X}_t \equiv (t, X_t, Y_t)$. Then, the corollary applies trivially, but at the price of increasing the dimension of the state space. The same strategy also works for more complicated functionals of the rough path $\widehat{\mathbb{X}}$. For instance, Y can be of the form $Y_t = g(t, \widetilde{Y}_t)$ where $\widetilde{Y}$ solves a rough differential equation*

$$d\widetilde{Y}_t = b(\widetilde{Y}_t)\,dt + \sigma(\widetilde{Y}_t)\,d\mathbb{X}_t.$$

If g is sufficiently smooth, Y is controlled by $\mathbb{X}$ *(cf. [FH14]) which guarantees that $\widetilde{X}_t \equiv (t, X_t, Y_t)$ can be lifted to a rough paths valued process.*

Remark 5.8 (A note on numerical implementation)**.** *For actual applications, the final formulas presented in Theorem 5.5 and Corollary 5.6 need to be implemented. In this context, several approximation steps will routinely apply.*

1. *The processes X, Y and, possibly, $\widetilde{Y}$ (see Remark 5.7) need to be simulated, which may require discretization of stochastic / rough differential equations.*
2. *Given paths of X, we need to compute truncated signatures $\mathbb{X}^{\leq N}_{0,t}$. Fortunately, packages for this task are readily available. See, for instance, the* iisignature *library [RG20].*
3. *If the setting allows to pose the optimization problem in terms of the expected signature only, as in the context of Corollary 5.6, then we next need to compute the expected truncated signature $\mathbb{E}[\mathbb{X}^{\leq N}_{0,T}]$. This can be done with the Monte Carlo method. The big advantage is, of course, that we only need to compute the expected signature once, and can then apply the optimization algorithm of our choice to a deterministic optimization problem.*
4. *Alternatively, if we rely on the formula of Theorem 5.5, then we need to apply MC simulation for each iteration of the optimization algorithm.*
5. *Finally, as already alluded to above, we solve an optimization problem in l restricted to a compact subset of $T(V^*)$ as seen in Theorem 5.5 or Corollary 5.6, respectively. We apply general state-of-the-art iterative optimization algorithms.*

Remark 5.9. *The method presented in Theorem 5.5 actually computes not just the value of the optimal stopping problem, but it also provides a (randomized) stopping time corresponding to this value. Indeed, suppose that $l^* = l^*_{N,K,\kappa} \in T(V^*)$ is a maximizer for the maximization problem on the r.h.s. of* (5.4)*. Recalling Definition 4.1, let $Z \sim Exp(1)$ denote a r.v. independent of X, Y, and set*

$$\tau^r_{l^*} := \inf\left\{ t \in [0, T \wedge S] \ : \ \int_0^t \langle l^*, \mathbb{X}^{<\infty}_{0,s}\rangle ds \geq Z \right\}.$$

Noting that $\tau^r_{l^*}$ is a randomized stopping time, we can conclude that

$$\mathbb{E}\left[Y_{\tau^r_{l^*}\wedge T}\right] \leq \sup_{\tau\in\mathcal{S}} \mathbb{E}\left[Y_{\tau\wedge T}\right],$$

see Proposition 4.2. Hence, a Monte Carlo approximation of $\mathbb{E}\left[Y_{\tau^r_{l^*}\wedge T}\right]$ is a low-biased estimator for $\sup_{\tau\in\mathcal{S}} \mathbb{E}\left[Y_{\tau\wedge T}\right]$, provided that it is based on samples of Y that are independent of the samples used to construct l^*.

6 Outlook: The dual problem

The optimal stopping problem has a dual formulation which is of the form

$$\sup_{\tau \in \mathcal{S}} \mathbb{E}[Y_{\tau\wedge T}] = \inf_{M\in H_0^1} \mathbb{E}[\sup_{t\in[0,T]} (Y_t - M_t)] \tag{6.1}$$

where H_0^1 is the space of $(\mathcal{F}_t)$-martingales which satisfy $M_0 = 0$ and $\mathbb{E}[\|M\|_\infty] < \infty$, cf. [Rog02, Theorem 2.1]. We believe that our approach can also be applied to the dual problem, at least under some additional assumptions. In this section, we give a sketch of how we believe this could be done.

Assume that we can define a Brownian motion B on the filtration $(\mathcal{F}_t)$. We can parameterize a subclass of local martingales by the map

$$\Lambda_T \ni \theta \mapsto \int_0^\cdot \theta(\widehat{\mathbb{X}}|_{[0,s]})\, dB_s$$

where the integral is understood in Itō-sense. If we restrict ourselves to the subclass

$$\mathcal{T}_M := \left\{ \theta \in \mathcal{T} \,:\, \mathbb{E}\left[\sup_{t\in[0,T]} \left| \int_0^t \theta(\widehat{\mathbb{X}}|_{[0,s]})\, dB_s \right| \right] < \infty \right\},$$

we obtain a subclass of true martingales. We believe that this class of martingales is sufficiently large to close the duality gap in 6.1, i.e. that we can restrict the infimum over this class of martingales. One strategy to prove this is to show that every progressively measurable process can be approximated (in a certain way) by stochastic processes of the form $t \mapsto \theta(\widehat{\mathbb{X}}|_{[0,t]})$. To see this, we can use similar ideas as in Lemma 3.2 and Proposition 4.2. Using the Martingale Representation Theorem [KS91, Theorem 4.15] should then be sufficient to prove the equality in (6.1).

We would like to add B as a component to our underlying rough path process. To do this, we need to make sense of integrals of the form

$$\int_0^t X_s \circ dB_s \tag{6.2}$$

pathwise. The symbol $\circ$ indicates that these integrals should satisfy the usual rules of calculus, otherwise their products can not be described with the shuffle algebra and the signature would not be a group-like element. In case of X being a semi-martingale, we can use Itō's theory of stochastic integration to define such integrals. However, for general rough paths valued processes $\mathbb{X}$, there is no canonical notion of a joint integral we can use here. To illustrate this issue, let us consider the case of X being a fractional Brownian motion with Hurst index $H \in (0,1)$ in which case we can assume that B is the Brownian motion used to define X through its kernel representation. In the case $H > 1/2$, we can use Young's integration theory to make sense of 6.2. Note that this integral will coincide with the Itō-integral now. For $H < 1/2$, things are getting much more complicated. In fact, one can show that the naive Riemann sum approximation to (6.2) does not converge in this case. However, it turns out that a suitable renormalization procedure can be used to obtain a candidate for 6.2 for $H < 1/2$, too, but this will require much more work [BFG+20]. To keep the discussion simple, we will assume from now on that X is a one-dimensional continuous stochastic process with sample paths of finite p-variation for some $p \in [1,2)$. This covers, for instance, the fractional Brownian motion with Hurst parameter $H > 1/2$. In this case, $\widehat{\mathbb{X}}$ denotes the joint rough paths lift of the process $t \mapsto \widehat{X}_t = (t, X_t, B_t)$ which is then canonically defined.

We aim to find a sufficient condition under which $\mathcal{T}_{\text{sig}} \subset \mathcal{T}_M$. This will follow if we can prove that

$$\mathbb{E}\left[\sup_{t\in[0,T]}\left|\int_0^t \langle l, \widehat{\mathbb{X}}^{<\infty}_{0,s}\rangle\, dB_s\right|\right] < \infty$$

for every $l \in T((\mathbb{R}^{1+2})^*)$. This, in turn, can be deduced if

$$\mathbb{E}\left[\sup_{t\in[0,T]}\left|\int_{0<t_1<\ldots<t_k<t} d\widehat{X}_{t_1}\otimes\cdots\otimes d\widehat{X}_{t_k}\right|\right] < \infty$$

for every $k \in \mathbb{N}$. From Lyons' Extension theorem, this is the case if

$$\mathbb{E}\left[\|\widehat{\mathbb{X}}\|^k_{q-\text{var}}\right] < \infty$$

for every $k \in \mathbb{N}$ where $q > 2$ is chosen such that $\frac{1}{p} + \frac{1}{q} > 1$. From the estimate

$$\|\widehat{\mathbb{X}}\|_{q-\text{var}} \lesssim \|\mathbb{B}\|_{q-\text{var}} + \|X\|_{p-\text{var}}$$

where $\mathbb{B}$ is the Stratonovich lift of the Brownian motion, we see that if $\mathbb{E}[\|X\|^k_{p-\text{var}}] < \infty$ for every $k \geq 1$, we can conclude that $\mathcal{T}_{\text{sig}} \subset \mathcal{T}_M$. Note that this holds in case of a fractional Brownian motion.

In the remaining parts, we assume that $Y = X$. We want to calculate

$$\inf_{\theta\in\mathcal{T}_{\text{sig}}} \mathbb{E}\left[\sup_{t\in[0,T]} (X_t - \int_0^t \theta(\widehat{\mathbb{X}}|_{[0,s]})\, dB_s)\right].$$

Note that

$$\begin{aligned}\inf_{\theta\in\mathcal{T}_{\text{sig}}} \mathbb{E}\left[\sup_{t\in[0,T]} (X_t - \int_0^t \theta(\widehat{\mathbb{X}}|_{[0,s]})\, dB_s)\right] &= \inf_{l\in T((\mathbb{R}^{1+2})^*)} \mathbb{E}\left[\sup_{t\in[0,T]} (X_t - \int_0^t \langle l, \widehat{\mathbb{X}}^{<\infty}_{0,s}\rangle\, dB_s)\right] \\ &= \inf_{l\in T((\mathbb{R}^{1+2})^*)} \mathbb{E}\left[\sup_{t\in[0,T]} \langle 2 - l3, \widehat{\mathbb{X}}^{<\infty}_{0,t}\rangle\right].\end{aligned}$$

For given $\kappa > 0$, we can consider the stopping time

$$S = \inf\{t \geq 0 \,:\, \|\widehat{\mathbb{X}}\|_{p-var;[0,t]} \geq \kappa\} \wedge T$$

and prove that

$$\begin{aligned}\inf_{l\in T((\mathbb{R}^{1+2})^*)} \mathbb{E}\left[\sup_{t\in[0,T]} \langle 2 - l3, \widehat{\mathbb{X}}^{<\infty}_{0,t}\rangle\right] &= \lim_{\kappa\to\infty}\lim_{K\to\infty}\inf_{|l|+\deg(l)\leq K} \mathbb{E}\left[\sup_{t\in[0,S]} \langle 2 - l3, \widehat{\mathbb{X}}^{<\infty}_{0,t}\rangle\right] \\ &= \lim_{K\to\infty}\lim_{\kappa\to\infty}\inf_{|l|+\deg(l)\leq K} \mathbb{E}\left[\sup_{t\in[0,S]} \langle 2 - l3, \widehat{\mathbb{X}}^{<\infty}_{0,t}\rangle\right].\end{aligned}$$

To approximate the supremum, we may use that for a continuous function f,

$$\sup_{t\in[0,T]} f(t) \approx \frac{1}{n}\log\left(\frac{1}{T}\int_0^T \exp(nf(t))dt\right) \approx \frac{1}{n}\log\left(\int_0^T \exp(nf(t))dt\right)$$

for large n. Since we are interested in upper bounds, we could also use the more precise estimate

$$\sup_{t\in[0,T]} f(t) \le \frac{1}{n}\log\left(\frac{1}{T}\int_0^T \exp(nf(t))dt\right) + \frac{1}{n}\log n + \|f\|_\alpha T^\alpha n^{-\alpha}$$

which holds for every $n \in \mathbb{N}$. In any case, we expect that

$$\begin{aligned}\sup_{t\in[0,S]} \langle 2 - l3, \widehat{\mathbb{X}}^{<\infty}_{0,t}\rangle &\approx \frac{1}{n}\log\left(\int_0^S \exp(n\langle 2 - l3, \widehat{\mathbb{X}}^{<\infty}_{0,t}\rangle)\, dt\right)\\ &\approx \frac{1}{n}\log\left(\langle \exp^{\sqcup\!\sqcup}(n(2 - l3))1, \widehat{\mathbb{X}}^{<\infty}_{0,S}\rangle\right)\end{aligned}$$

for large n. Let (P_m) be a series of polynomials which approximate the logarithm uniformly on compact sets. Then,

$$\log\left(\langle \exp^{\sqcup\!\sqcup}(n(2 - l3))1, \widehat{\mathbb{X}}^{<\infty}_{0,S}\rangle\right) \approx \langle P^{\sqcup\!\sqcup}_m(\exp^{\sqcup\!\sqcup}(n(2 - l3))1), \widehat{\mathbb{X}}^{<\infty}_{0,S}\rangle.$$

Eventually, we expect that

$$\inf_{M\in H^1_0} \mathbb{E}[\sup_{t\in[0,T]}(Y_t - M_t)] \approx \inf_{|l|+\deg(l)\le K} \langle P^{\sqcup\!\sqcup}_m(\exp^{\sqcup\!\sqcup}(n(2 - l3))1), \mathbb{E}[\widehat{\mathbb{X}}^{\le N}_{0,S}]\rangle$$

for large N, m, n, K and κ.

References

[AGG+18] Imanol Perez Arribas, Guy M Goodwin, John R Geddes, Terry Lyons, and Kate EA Saunders. A signature-based machine learning model for distinguishing bipolar disorder and borderline personality disorder. *Translational psychiatry*, 8(1):1–7, 2018.

[BFG+20] Christian Bayer, Peter K. Friz, Paul Gassiat, Jorg Martin, and Benjamin Stemper. A regularity structure for rough volatility. *Math. Finance*, 30(3):782–832, 2020.

[BTW20] Christian Bayer, Raúl Tempone, and Sören Wolfers. Pricing american options by exercise rate optimization. *Quantitative Finance*, pages 1–12, 2020.

[CF10] Rama Cont and David Fournie. A functional extension of the ito formula. *Comptes Rendus Mathematique*, 348(1-2):57–61, 2010.

[CL16] Ilya Chevyrev and Terry Lyons. Characteristic functions of measures on geometric rough paths. *The Annals of Probability*, 44(6):4049–4082, 2016.

[Dup19] Bruno Dupire. Functional Itô calculus. *Quant. Finance*, 19(5):721–729, 2019.

[Faw02] Thomas Fawcett. *Problems in stochastic analysis : connections between rough paths and non-commutative harmonic analysis*. PhD thesis, University of Oxford, 2002.

[FH14] Peter K. Friz and Martin Hairer. *A Course on Rough Paths with an introduction to regularity structures*, volume XIV of *Universitext*. Springer, Berlin, 2014.

[FV10] Peter K. Friz and Nicolas B. Victoir. *Multidimensional stochastic processes as rough paths*, volume 120 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010. Theory and applications.

[HL10] Ben Hambly and Terry Lyons. Uniqueness for the signature of a path of bounded variation and the reduced path group. *Annals of Mathematics*, pages 109–167, 2010.

[KLPA20] Jasdeep Kalsi, Terry Lyons, and Imanol Perez Arribas. Optimal execution with rough path signatures. *SIAM J. Financial Math.*, 11(2):470–493, 2020.

[KS91] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991.

[LCL07] Terry J. Lyons, Michael Caruana, and Thierry Lévy. *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, 2004, With an introduction concerning the Summer School by Jean Picard.

[RG20] Jeremy F Reizenstein and Benjamin Graham. Algorithm 1004: The iisignature library: Efficient calculation of iterated-integral signatures and log signatures. *ACM Transactions on Mathematical Software (TOMS)*, 46(1):1–21, 2020.

[Rog02] L. C. G. Rogers. Monte Carlo valuation of American options. *Math. Finance*, 12(3):271–286, 2002.

[Shi08] A. N. Shiryaev. *Optimal stopping rules*, volume 8 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2008. Translated from the 1976 Russian second edition by A. B. Aries, Reprint of the 1978 translation.

[Wiś94] Andrzej Wiśniewski. The structure of measurable mappings on metric spaces. *Proc. Amer. Math. Soc.*, 122(1):147–150, 1994.

A Technical aspects of stopped rough paths

Lemma A.1. *The topology on Λ_T coincides with the final topology induced by the map $\varphi\colon [0,T] \times \widehat{\Omega}^p_T \to \Lambda_T$, $\varphi(t,\widehat{\mathbb{X}}) = \widehat{\mathbb{X}}|_{[0,t]}$. Moreover, Λ_T is Polish.*

Proof. A set $U \subset \Lambda_T$ is open with respect to the final topology if and only if $\varphi^{-1}(U)$ is open in $[0,T] \times \widehat{\Omega}^p_T$. One can easily check that φ is continuous for the topology induced by d, therefore $\varphi^{-1}(U)$ is open for every open set $U \subset \Lambda_T$. Now assume that $\varphi^{-1}(U)$ is open for a set $U \subset \Lambda_T$. Let $\mathbb{X}|_{[0,t]} \in U$ and choose $\mathbb{Y}|_{[0,s]} \in \Lambda_T$ with $d(\mathbb{X}|_{[0,t]}, \mathbb{Y}|_{[0,s]}) < \varepsilon$. Our goal is to prove that $\mathbb{Y}|_{[0,s]} \in U$ for ε chosen sufficiently small. Note that

$$\varphi^{-1}(\mathbb{X}|_{[0,t]}) = \{(t,\widetilde{\mathbb{X}}) \,:\, \widetilde{\mathbb{X}}|_{[0,t]} = \mathbb{X}|_{[0,t]}\}.$$

Assume $s \geq t$ first. Then $d(\mathbb{X}|_{[0,t]}, \mathbb{Y}|_{[0,s]}) < \varepsilon$ implies that

$$|t-s| < \varepsilon \quad \text{and} \quad d_{p-\mathrm{var};[0,s]}(\widetilde{\mathbb{X}}|_{[0,s]}, \mathbb{Y}|_{[0,s]}) < \varepsilon$$

where $\widetilde{\mathbb{X}}|_{[0,s]}$ is the stopped path defined on $[0,s]$ as explained in Definition 3.1. Let $\widetilde{\mathbb{X}} = \widetilde{\mathbb{X}}|_{[0,T]} \in \widehat{\Omega}^p_T$ be the stopped path defined on the whole time interval $[0,T]$. Since $(t,\widetilde{\mathbb{X}}) \in \varphi^{-1}(\mathbb{X}|_{[0,t]}) \subset \varphi^{-1}(U)$ and $\varphi^{-1}(U)$ is open, there is a $\delta > 0$ such that whenever $u \in (t-\delta, t+\delta)$ and $d_{p-\mathrm{var};[0,T]}(\widetilde{\mathbb{X}}, \widetilde{\mathbb{Y}}) < \delta$, we have $(u,\widetilde{\mathbb{Y}}) \in \varphi^{-1}(U)$. Choosing ε sufficiently small, we can assume that $s \in (t-\delta, t+\delta)$.

Define $\widetilde{\mathbb{Y}} = \widetilde{\mathbb{Y}}|_{[0,T]} \in \widehat{\Omega}^p_T$ as in Definition 3.1. Then $(s, \widetilde{\mathbb{Y}}) \in \varphi^{-1}(\mathbb{Y}|_{[0,s]})$ and

$$\begin{aligned} d_{p-\mathrm{var};[0,T]}(\widetilde{\mathbb{X}}, \widetilde{\mathbb{Y}}) &\leq C_p(d_{p-\mathrm{var};[0,s]}(\widetilde{\mathbb{X}}, \widetilde{\mathbb{Y}}) + d_{p-\mathrm{var};[s,T]}(\widetilde{\mathbb{X}}, \widetilde{\mathbb{Y}})) \\ &= C_p d_{p-\mathrm{var};[0,s]}(\widetilde{\mathbb{X}}|_{[0,s]}, \mathbb{Y}|_{[0,s]}) \leq C_p\varepsilon. \end{aligned}$$

Choosing ε small, we conclude $(s, \widetilde{\mathbb{Y}}) \in \varphi^{-1}(U)$ and thus $\mathbb{Y}|_{[0,s]} \in U$. For $s \leq t$, we can argue similarly which proves that both topologies indeed coincide. Concerning the second statement, separability follows from the separability of $[0,T] \times \widehat{\Omega}^p_T$ and the fact that φ is a continuous surjection. To prove that Λ_T is complete with respect to the metric d is straightforward and follows from the fact that $[0,T]$ and $\widehat{\Omega}^p_T$ are complete.

□

Corollary A.2. *Let Z be any topological space. A map $g\colon \Lambda_T \to Z$ is continuous if and only if the map $[0,T] \times \widehat{\Omega}^p_T \ni (t, \widehat{\mathbb{X}}) \mapsto g(\widehat{\mathbb{X}}|_{[0,t]}) \in Z$ is continuous.*

Proof. Follows from the universal property of the final topology. □

DOI 10.20347/WIAS.PREPRINT.2790