Notes on Signature and Rough Path

Haosheng Zhou

Apr, 2025

Contents

Fundamentals of Signature

The question of interest is: how to characterize **paths** of processes (time series data)? In general, paths are infinite-dimensional objects denoted as $X : [0,T] \to \mathbb{R}^d$, and a natural way to reduce it to finite-dimensional objects is to consider the time discretized path $(X_{t_1}, ..., X_{t_n})$ where $t_1 < ... < t_n$. However, the choice of the discretization endpoints is subtle and one loses more information as the path gets rougher. For example, for the same set of three points on a path given as $(t_1, X_{t_1}), (t_2, X_{t_2}), (t_3, X_{t_3})$, a possible path of Lipschitz continuity interpolated by the three points has much fewer "restrictions" than a path that is $\frac{1}{2}$-Holder continuous (paths of BM).

Signatures provide a robust way to summarize features of a path. The signature $S(X)$ of path X has a countable dimension, and the truncation up to a certain level does not dramatically reduce the amount of information contained.

Signature for Paths of Finite Variation

For a path $X : [0,T] \to \mathbb{R}^d$ of finite variation, the signature is defined as

$$S_{0,T}(X) = (1, S_{0,T}^{(1)}, S_{0,T}^{(2)}, ...) \in T(\mathbb{R}^d), \tag{1}$$

where $S_{0,T}^{(i)} \in \mathbb{R}^{d^i}$ for $\forall i \in \mathbb{N}$, and $T(\mathbb{R}^d) := \{1\} \times \mathbb{R}^d \times \mathbb{R}^{d^2} \times ...$ is the tensor algebra. Here $S_{0,T}^{(i)}$ is called the i-th order signature on $[0,T]$ and consists of d^i path integrals of different orders. To raise some simple examples,

$$S_{0,T}^{(1)\ j} := \int_0^T \mathrm{d}X_t^j = X_T^j - X_0^j,\ \forall j \in [d]. \tag{2}$$

The first order signature consists of path integrals w.r.t. one of the d components, therefore

$$S_{0,T}^{(1)} := \left(X_T^1 - X_0^1, ..., X_T^d - X_0^d\right) \in \mathbb{R}^d. \tag{3}$$

For the second order signature $S_{0,T}^{(2)}$, the order of integration matters:

$$S_{0,T}^{(2)\ i,j} := \int_0^T \int_0^t \mathrm{d}X_s^i \, \mathrm{d}X_t^j = \int_0^T (X_t^i - X_0^i) \, \mathrm{d}X_t^j,\ \forall (i,j) \in [d]^2. \tag{4}$$

Therefore,

$$S_{0,T}^{(2)} := \left(S_{0,T}^{(2)\ 1,1}, ..., S_{0,T}^{(2)\ d,d}\right) \in \mathbb{R}^{d^2}. \tag{5}$$

In general, let $I := \{(i_1, ..., i_k) \in [d]^k\}$ denote the collection of words, then

$$S_{0,T}^{(k)\ i_1,...,i_k} := \int_0^T \int_0^{t_k} ... \int_0^{t_2} \mathrm{d}X_{t_1}^{i_1} ... \, \mathrm{d}X_{t_{k-1}}^{i_{k-1}} \, \mathrm{d}X_{t_k}^{i_k} \tag{6}$$

is one component of the k-th order signature indexed by $(i_1, ..., i_k) \in I$. Clearly, $|I| = d^k$ aligns with the dimension of $S_{0,T}^{(k)}$. Since the paths are of finite variation, the integrals are well-defined in the Lebesgue-Stieljes sense.

Example 1. *Consider the one-dimensional path $X_t = t$, then*

$$S^{(1)}_{0,T} = \int_0^T \mathrm{d}t = T, \tag{7}$$

$$S^{(2)}_{0,T} = \int_0^T S^{(1)}_{0,t}\,\mathrm{d}t = \frac{T^2}{2}, \tag{8}$$

$$S^{(n)}_{0,T} = \int_0^T S^{(n-1)}_{0,t}\,\mathrm{d}t = \frac{T^n}{n!}, \tag{9}$$

which have polynomial-like structures.

Example 2. *Consider any one-dimensional path X with finite variation,*

$$S^{(1)}_{0,T} = \int_0^T \mathrm{d}X_t = X_T - X_0, \tag{10}$$

$$S^{(2)}_{0,T} = \int_0^T S^{(1)}_{0,t}\,\mathrm{d}X_t = \frac{(X_T - X_0)^2}{2}, \tag{11}$$

$$S^{(n)}_{0,T} = \int_0^T S^{(n-1)}_{0,t}\,\mathrm{d}X_t = \frac{(X_T - X_0)^n}{n!}, \tag{12}$$

which still exhibit polynomial-like structures. When $d = 1$, the signatures are nothing but polynomials of $X_T - X_0$, the differences between the starting and ending points of the paths, which are not of much interests. As a result, we shall mainly focus on the applications of signatures when $d > 1$.

Example 3. *Consider the path X of a 1D BM $\{W_t\}$. Although this path does not have finite variation, the integrals can still be defined in the sense of stochastic integrals (e.g., in Ito or Stratonovich sense) so the signature can be similarly defined under stochastic integrals. Due to chain rule, the second order signature of BM path under Stratonovich integration is*

$$S^{(2)}_{0,T} = \int_0^T (W_t - W_0) \circ \mathrm{d}W_t = \frac{1}{2}W_T^2, \tag{13}$$

while the second order signature of BM path under Ito integration is

$$S^{(2)}_{0,T} = \int_0^T (W_t - W_0)\,\mathrm{d}W_t = \frac{1}{2}W_T^2 - \frac{T}{2}. \tag{14}$$

Clearly, when the path is rough, the values of the signature depend on the notion of integration adopted.

Since signatures are only of interest when $d > 1$, we provide a geometric interpretation for the case $d = 2$. Recall that by Green's theorem, if C is a smooth closed curve,

$$\oint_C -y\,\mathrm{d}x + x\,\mathrm{d}y = 2\iint_D \mathrm{d}x\,\mathrm{d}y, \tag{15}$$

where D is the interior of C that follows the same orientation. An analogue tells us that

$$\frac{S_{0,T}^{(2)\ 1,2} - S_{0,T}^{(2)\ 2,1}}{2} = \frac{1}{2}\int_0^T -(X_t^2 - X_0^2)\,\mathrm{d}X_t^1 + (X_t^1 - X_0^1)\,\mathrm{d}X_t^2 \tag{16}$$

is just the **Levy area** of the path X, i.e., the signed area formed by the path X and the straight line connecting (X_0^1, X_0^2) and (X_T^1, X_T^2).

Properties of Signature for Paths of Finite Variation

Theorem 1 (Factorial Decay)**.** *If X is a path of finite variation V on $[0,T]$, then*

$$|S_{0,T}^{(k)\ i_1,\dots,i_k}| \leq \frac{V^k}{k!},\ \forall (i_1,\dots,i_k) \in I,\ \forall k. \tag{17}$$

Proof. By triangle inequality for integrals,

$$|S_{0,T}^{(k)\ i_1,\dots,i_k}| \leq \int_{0<t_1<t_2<\dots<t_k<T} |\,\mathrm{d}X_{t_1}^{i_1}|\dots|\,\mathrm{d}X_{t_k}^{i_k}| = \int_{0<t_1<t_2<\dots<t_k<T} \mathrm{d}\mu_1(t_1)\dots\mathrm{d}\mu_k(t_k), \tag{18}$$

where each measure μ_i on $[0,T]$ has total mass at most V. This concludes the proof. □

The factorial decay of signatures ensures the convergence of $\sum_{k=0}^{\infty} S_{0,T}^{(k)}$ in the tensor algebra, which could be understood as an analytic function in terms of the path X. Also, it implies that the truncation of signatures up to a certain order does not drop too much information of the path.

Theorem 2 (Time Reparameterization Invariance)**.** *If X is a path of finite variation on $[0,T]$, and $\psi : [0,T] \to [0,T]$ is a continuous surjective non-decreasing function that reparameterizes the path, i.e.,*

$$\tilde{X}_t := X_{\psi(t)}, \tag{19}$$

then $S_{0,T}(X) = S_{0,T}(\tilde{X})$.

Proof. Consider the k-th order signature for path X:

$$S_{0,T}^{(k)\ i_1,\dots,i_k} = \int_{0<t_1<t_2<\dots<t_k<T} \mathrm{d}X_{t_1}^{i_1}\dots\mathrm{d}X_{t_k}^{i_k}, \tag{20}$$

and conduct a change of variables $t_j = \psi(s_j)$ for $\forall j \in [k]$:

$$S_{0,T}^{(k)\ i_1,\dots,i_k} = \int_{0<s_1<s_2<\dots<s_k<T} \mathrm{d}X_{\psi(s_1)}^{i_1}\dots\mathrm{d}X_{\psi(s_k)}^{i_k} = \int_{0<s_1<s_2<\dots<s_k<T} \mathrm{d}\tilde{X}_{s_1}^{i_1}\dots\mathrm{d}\tilde{X}_{s_k}^{i_k}, \tag{21}$$

which is equal to the k-th order signature for path $\tilde{X}$. □

This is a basic requirement for signatures to be useful, since time reparameterization only changes the speed moving along the path but not the path itself.

Theorem 3 (Time Reversal)**.** *If X is a path of finite variation on $[0,T]$, and $\overline{X}$ is its time-reversal, i.e., $\overline{X}_t = X_{T-t}$, then*

$$S^{(k)\ i_1,...,i_k}_{a,b}(X) = (-1)^k S^{(k)\ i_k,...,i_1}_{T-b,T-a}(\overline{X}),\ \forall k,\ \forall (i_1,...,i_k),\ \forall 0 \le a < b \le T. \tag{22}$$

In particular, take $a = 0, b = T$ to get

$$S^{(k)\ i_1,...,i_k}_{0,T}(X) = (-1)^k S^{(k)\ i_k,...,i_1}_{0,T}(\overline{X}),\ \forall k,\ \forall (i_1,...,i_k). \tag{23}$$

Proof. Prove by induction. Assume it holds for signatures of order at most $k-1$, when the order is k:

$$S^{(k)\ i_1,...,i_k}_{a,b}(X) = \int_a^b S^{(k-1)\ i_1,...,i_{k-1}}_{a,t}(X)\,\mathrm{d}X^{i_k}_t\ \ (s = T-t) \tag{24}$$

$$= \int_{T-a}^{T-b} S^{(k-1)\ i_1,...,i_{k-1}}_{a,T-s}(X)\,\mathrm{d}X^{i_k}_{T-s} \tag{25}$$

$$= (-1)^k \int_{T-b}^{T-a} S^{(k-1)\ i_{k-1},...,i_1}_{s,T-a}(\overline{X})\,\mathrm{d}\overline{X}^{i_k}_s \tag{26}$$

$$= (-1)^k \int_{T-b}^{T-a} \left(\int_{s<t_{k-1}<...<t_1<T-a} \mathrm{d}\overline{X}^{i_{k-1}}_{t_{k-1}} ... \,\mathrm{d}\overline{X}^{i_1}_{t_1} \right) \mathrm{d}\overline{X}^{i_k}_s. \tag{27}$$

Apply Fubini's theorem to get

$$S^{(k)\ i_1,...,i_k}_{a,b}(X) = (-1)^k \int_{T-b<s<t_{k-1}<...<t_1<T-a} \mathrm{d}\overline{X}^{i_k}_s\,\mathrm{d}\overline{X}^{i_{k-1}}_{t_{k-1}} ... \,\mathrm{d}\overline{X}^{i_1}_{t_1} = (-1)^k S^{(k)\ i_k,...,i_1}_{T-b,T-a}(\overline{X}). \tag{28}$$

□

The signatures of the original path and its time reversal are different by a sign $(-1)^k$ after reversing the order of the indices.

Theorem 4 (Uniqueness)**.** *The signature $S(X)$ uniquely determines the path X of finite variation, up to the starting point of the path, time reparameterization and a tree-like equivalence.*

This is the main reason why the quantity gets the name "signature". The tree-like equivalence will be mentioned later.

Theorem 5 (Shuffle Product)**.** *For multi-indices $I \in [d]^k$, $J \in [d]^m$, define $I \sqcup J$ as the collection of shuffles of (I, J), which are multi-indices of lengths $k+m$ where the orders in I and J are respectively preserved. If X is a path of finite variation on $[0,T]$, then*

$$S^{(k)\ I}_{0,T} \cdot S^{(m)\ J}_{0,T} = \sum_{K \in I \sqcup J} S^{(k+m)\ K}_{0,T}. \tag{29}$$

Proof. Prove by induction. Write the product of signatures as a double integral by Fubini's theorem with the integration domain $(s,t) \in [0,T]^2$. Split the domain into two parts based on $s < t$ or $s \geq t$ and use induction hypothesis. The proof is concluded by using the structure of $I \sqcup J$. $\square$

It is important to understand what the equation is telling us.

Example 4. *Consider the case $d = 2, k = m = 1$, and $I = \{1\}, J = \{2\}$, the shuffle product is $I \sqcup J = \{(1,2),(2,1)\}$. The LHS is*

$$S_{0,T}^{(1)\ 1} \cdot S_{0,T}^{(1)\ 2} = (X_T^1 - X_0^1) \cdot (X_T^2 - X_0^2). \tag{30}$$

The RHS is, by integration by parts,

$$S_{0,T}^{(2)\ 1,2} + S_{0,T}^{(2)\ 2,1} = \int_0^T (X_t^1 - X_0^1)\,\mathrm{d}X_t^2 + \int_0^T (X_t^2 - X_0^2)\,\mathrm{d}X_t^1 \tag{31}$$

$$= X_T^1 X_T^2 - X_0^1 X_0^2 - X_0^1(X_T^2 - X_0^2) - X_0^2(X_T^1 - X_0^1) \tag{32}$$

$$= X_T^1 X_T^2 - X_0^1 X_T^2 - X_0^2 X_T^1 + X_0^1 X_0^2 = S_{0,T}^{(1)\ 1} \cdot S_{0,T}^{(1)\ 2}. \tag{33}$$

As a result, ***the shuffle product identity is essentially integration by parts****. For rougher paths, the shuffle product identity not necessarily holds due to correction terms in stochastic integrals. From the proof, we can also see why the conclusion generally fails: one can never use Fubini-type theorems to interchange two stochastic integrals.*

Example 5. *Consider the case $d = 2, k = 2, m = 1$, and $I = \{1,2\}, J = \{1\}$, the shuffle product is $I \sqcup J = \{(1,1,2),(1,1,2),(1,2,1)\}$ (note that repeated elements are allowed). The conclusion is*

$$S_{0,T}^{(2)\ 1,2} \cdot S_{0,T}^{(1)\ 1} = 2S_{0,T}^{(3)\ 1,1,2} + S_{0,T}^{(3)\ 1,2,1}. \tag{34}$$

Example 6. *Set $X_t = tx$ for a fixed $x \in \mathbb{R}^d$ and $T = 1$. Calculate*

$$S_{0,1}^{(k)\ I} = \int_{0<t_1<\ldots<t_k<1} \mathrm{d}X_{t_1}^{i_1} \ldots \mathrm{d}X_{t_k}^{i_k} = x_{i_1} \ldots x_{i_k} \int_{0<t_1<\ldots<t_k<1} \mathrm{d}t_1 \ldots \mathrm{d}t_k = \frac{x_1^{\beta_1} \ldots x_d^{\beta_d}}{k!}, \tag{35}$$

where $\beta_j^I = |\{i_p : i_p = j,\ \forall p \in [k]\}|$ for $\forall j \in [d]$. In this case, the order of indices within I does not matter, and only the number of appearance matters.

The shuffle product identity tells us

$$\frac{x_1^{\beta_1^I} \ldots x_d^{\beta_d^I}}{k!} \cdot \frac{x_1^{\beta_1^J} \ldots x_d^{\beta_d^J}}{m!} = \sum_{K \in I \sqcup J} \frac{x_1^{\beta_1^K} \ldots x_d^{\beta_d^K}}{(k+m)!}. \tag{36}$$

Clearly, each $K \in I \sqcup J$ satisfies $\beta_j^K = \beta_j^I + \beta_j^J$ for $\forall j \in [d]$. Moreover, $|I \sqcup J| = \binom{m+k}{k}$. As a result,

$$\frac{x_1^{\beta_1^I} \ldots x_d^{\beta_d^I}}{k!} \cdot \frac{x_1^{\beta_1^J} \ldots x_d^{\beta_d^J}}{m!} = \binom{m+k}{k} \frac{x_1^{\beta_1^I + \beta_1^J} \ldots x_d^{\beta_d^I + \beta_d^J}}{(k+m)!}, \tag{37}$$

which is just the multiplication of polynomials. This again shows the polynomial-like structure of signatures, i.e., the product of signatures is a linear combination of signatures of higher orders.

The following Chen's identity allows computations of signatures on different time intervals.

Theorem 6 (Chen's Identity)**.** *For a path X of finite variation, for $\forall 0 < s < T$ and $\forall (i_1, ..., i_k) \in [d]^k$,*

$$S_{0,T}^{(k)\ (i_1,\dots,i_k)} = \sum_{m=0}^{k} S_{0,s}^{(m)\ i_1,\dots,i_m} \cdot S_{s,T}^{(k-m)\ i_{m+1},\dots,i_k}. \tag{38}$$

Proof. Prove by induction. The proof only uses the property of integrals that $\int_0^T = \int_0^s + \int_s^T$. $\square$

Chen's identity is a fundamental consistency condition for signatures on different time intervals. Different from the shuffle product identity and the time reversal which is based on **Fubini's theorem (which fails for stochastic integrals)**, the proof of Chen's identity only uses the splitting of integration domains (which holds for almost all useful notions of integrals). As a result, for extensions to signatures of rougher paths, we might lose the shuffle product identity but we always hope to maintain Chen's identity.

Example 7. *Consider the path X in $\mathbb{R}^2$ with $T = 2$. In time $[0, 1]$, the path goes from $(3, 0)$ to $(3, 1)$ in uniform speed, while in time $[1, 2]$, the path goes from $(3, 1)$ to $(0, 1)$ in uniform speed. We wish to calculate $S_{0,2}^{(2)\ 1,2}$ and $S_{0,2}^{(2)\ 2,1}$ through Chen's identity.*

Naturally, one would split the time horizon into two intervals at time 1.

$$\begin{aligned} S_{0,2}^{(2)\ 1,2} &= S_{0,1}^{(0)} \cdot S_{1,2}^{(2)\ 1,2} + S_{0,1}^{(1)\ 1} \cdot S_{1,2}^{(1)\ 2} + S_{0,1}^{(2)\ 1,2} \cdot S_{1,2}^{(0)} && (39)\\ &= 1 \cdot 0 + 0 \cdot 0 + 0 \cdot 1 = 0. && (40)\end{aligned}$$

On the other hand,

$$\begin{aligned} S_{0,2}^{(2)\ 2,1} &= S_{0,1}^{(0)} \cdot S_{1,2}^{(2)\ 2,1} + S_{0,1}^{(1)\ 2} \cdot S_{1,2}^{(1)\ 1} + S_{0,1}^{(2)\ 2,1} \cdot S_{1,2}^{(0)} && (41)\\ &= 1 \cdot 0 + 1 \cdot (-3) + 0 \cdot 1 = -3. && (42)\end{aligned}$$

As a result, $\frac{S_{0,2}^{(2)\ 1,2} - S_{0,2}^{(2)\ 2,1}}{2} = \frac{3}{2}$ matches with the Levy area of the path.

The following universal approximation theorem is what makes signatures useful in machine learning contexts.

Theorem 7 (Universal Approximation)**.** *Let $V^1([0, T], \mathbb{R}^d)$ be the collection of paths on $\mathbb{R}^d$ of finite variation on time horizon $[0, T]$ quotient tree-like structures, equipped with the norm $\|X\|_{BV}$ as the total variation of $\forall X \in V^1([0, T], \mathbb{R}^d)$. For $\forall \varepsilon > 0$, $\forall K \subset V^1([0, T], \mathbb{R}^d)$ as a compact subset and any $g : K \to \mathbb{R}$ continuous, there exists a linear functional $L : T(\mathbb{R}^d) \to \mathbb{R}$ such that*

$$|g(X) - L(S(X))| < \varepsilon,\ \forall X \in K. \tag{43}$$

In addition, this L depends on a finite truncation of the signature. There exists a truncation level $N = N(\varepsilon, g, K)$ and $l_k \in \mathbb{R}^{d^k}$ such that

$$L(S(X)) = \sum_{k=0}^{N} \langle l_k, S_{0,T}^{(k)} \rangle. \tag{44}$$

Proof. The quotient space implies the uniqueness of paths, i.e., $S(X) = S(Y)$ iff $X = Y$, which guarantees the point-separating property. As a result, $S(K)$, the collection of signatures for paths in K, forms a point-separating algebra of continuous functions that contains 1, which is dense in $C(K)$, as implied by the Stone-Weierstrass theorem. $\square$

Remark. *It is necessary to consider the quotient space since* ***signatures are invariant under constant translation of the path*** X_0 ***and time reparameterization****. In addition, if one considers concatenating a path $X_{[0,T]}$ and its own time-reversal $\overline{X}_{[0,T]}$ to form a new path Y, Y goes along X on $[0,T]$ and goes along $\overline{X}$ on $[T, 2T]$. In this case, we calculate the signature through Chen's identity:*

$$\begin{aligned} S_{0,2T}^{(k)\ i_1,\dots,i_k}(Y) &= \sum_{m=0}^{k} S_{0,T}^{(m)\ i_1,\dots,i_m}(Y) \cdot S_{T,2T}^{(k-m)\ i_{m+1},\dots,i_k}(Y) && (45)\\ &= \sum_{m=0}^{k} S_{0,T}^{(m)\ i_1,\dots,i_m}(X) \cdot S_{0,T}^{(k-m)\ i_{m+1},\dots,i_k}(\overline{X}) && (46)\\ &= \sum_{m=0}^{k} S_{0,T}^{(m)\ i_1,\dots,i_m}(X) \cdot (-1)^{k-m} S_{0,T}^{(k-m)\ i_k,\dots,i_{m+1}}(X) = 0,\ \forall k \geq 1. && (47) \end{aligned}$$

The last equation follows from the lemma below. As a result, the signature of Y is the same as that of a constant path $Z_t \equiv z$, i.e., $S(Y) = (1, 0, 0, \ldots)$. Such a pattern is called ***"tree-like"*** *and naturally appears when discussing the uniqueness of signatures. However, it prevents the point-separating property and needs to be removed when applying Stone-Weierstrass.*

Lemma 1. *For multi-index $L = (l_1, \ldots, l_k)$, denote $a_L := a_{l_1 l_2 \ldots l_k}$, then*

$$\sum_{m=0}^{k} (-1)^{k-m} \sum_{L \in (i_1,\dots,i_m) \sqcup\!\sqcup (i_k,\dots,i_{m+1})} a_L = 0,\ \forall k \geq 1. \tag{48}$$

Proof. Denote p_L as the coefficient of a_L on the LHS. Only need to prove $p_L = 0,\ \forall L$. We start with observing

$$\sum_{L} p_L = \sum_{m=0}^{k} (-1)^{k-m} \binom{k}{m} = 0. \tag{49}$$

Clearly, the first index in L is either i_1 or i_k. Compute by cases:

$$\sum_{L: l_1 = i_1} p_L = \sum_{m=1}^{k} (-1)^{k-m} \binom{k-1}{m-1} = 0, \quad \sum_{L: l_1 = i_k} p_L = \sum_{m=0}^{k-1} (-1)^{k-m} \binom{k-1}{m} = 0. \tag{50}$$

By discussing the second until the k-th index, an induction argument shows $p_L = 0,\ \forall L$. $\square$

For probability measures on the path space, expected signatures serve as moments. As an analogue to the moment method in Euclidean spaces, i.e., two probability measures supported on a compact set of $\mathbb{R}^d$ with the same moments are equal, we can extend this result in terms of signatures.

Theorem 8 (Expected Signature as Moment)**.** *If $\mathbb{P}$ and $\mathbb{Q}$ are probability measures on a compact set $K \subset V^1([0,T],\mathbb{R}^d)$ and they share the same expected signatures of all orders, i.e.,*

$$\mathbb{E}_{X\sim\mathbb{P}} S_{0,T}^{(k)\ I} = \mathbb{E}_{X\sim\mathbb{Q}} S_{0,T}^{(k)\ I},\ \forall k,\ \forall I, \tag{51}$$

then $\mathbb{P} = \mathbb{Q}$.

Proof. Directly follows from the universal approximation theorem of signatures. □

Applications of Signatures of Paths of Finite Variation

The first example illustrates the connection between signatures and ODEs.

Example 8. *Signatures detect and amplify the differences in paths. Given two paths X and $\tilde{X}$ in $\mathbb{R}^d$ that are uniformly close to each other, i.e., $\sup_{t\in[0,T]} \|X_t - \tilde{X}_t\| < \varepsilon$, the normal testing scheme fails, where one calculates the pointwise differences between paths at a selection of times $t_1, ..., t_n \in [0,T]$. Instead, an idea is to consider the solutions to the ODEs induced by the two paths:*

$$\mathrm{d}Y_t = V(Y_t)\,\mathrm{d}X_t, \quad \mathrm{d}\tilde{Y}_t = V(\tilde{Y}_t)\,\mathrm{d}\tilde{X}_t, \tag{52}$$

sharing the same initial condition $Y_0 = \tilde{Y}_0 = y_0$, where Y is a path in $\mathbb{R}^e$ and $V : \mathbb{R}^e \to \mathbb{R}^{e\times d}$ has the representation $V = (V_1, ..., V_d)$, where $V_i : \mathbb{R}^e \to \mathbb{R}^e$. It turns out that the difference $\|Y_T - \tilde{Y}_T\|$ significantly amplifies the path difference between X and $\tilde{X}$, which is hard to observe from pointwise differences $\|X_{t_i} - \tilde{X}_{t_i}\|$.

It is natural to investigate such ODEs

$$Y_t = Y_0 + \sum_{i=1}^{d} \int_0^T V_i(Y_t)\,\mathrm{d}X_t^i, \tag{53}$$

and show their connections with signatures. The solution to ODEs are typically constructed through Picard iterations. Let Y_t^k denote the solution constructed in the k-th Picard iteration, then

$$Y_t^k := y_0 + \sum_{i=1}^{d} \int_0^t V_i(Y_s^{k-1})\,\mathrm{d}X_s^i, \quad Y_t^0 \equiv y_0. \tag{54}$$

Assume that all V_i are linear, we see that

$$Y_t^1 = y_0 + \sum_{i=1}^{d} V_i(y_0) \int_0^t \mathrm{d}X_s^i = y_0 + \sum_{i=1}^{d} V_i(y_0) \cdot S_{0,t}^{(1)\ i}(X). \tag{55}$$

$$Y_t^2 = y_0 + \sum_{i_2=1}^{d} \int_0^t V_{i_2}\left(y_0 + \sum_{i_1=1}^{d} V_{i_1}(y_0) \cdot S_{0,s}^{(1)\ i_1}(X)\right) \mathrm{d}X_s^{i_2} \tag{56}$$

$$= y_0 + \sum_{i_2=1}^{d} \int_0^t \left[V_{i_2}(y_0) + \sum_{i_1=1}^{d} V_{i_2}\left(V_{i_1}(y_0)\right) S_{0,s}^{(1)\ i_1}(X)\right] \mathrm{d}X_s^{i_2} \tag{57}$$

$$= y_0 + \sum_{i_2=1}^{d} V_{i_2}(y_0) \int_0^t \mathrm{d}X_s^{i_2} + \sum_{i_2=1}^{d}\sum_{i_1=1}^{d} V_{i_2}\left(V_{i_1}(y_0)\right) \int_0^t S_{0,s}^{(1)\ i_1}(X)\,\mathrm{d}X_s^{i_2} \tag{58}$$

$$= y_0 + \sum_{i_2=1}^{d} V_{i_2}(y_0) S_{0,t}^{(1)\ i_2}(X) + \sum_{i_2=1}^{d}\sum_{i_1=1}^{d} V_{i_2}\left(V_{i_1}(y_0)\right) S_{0,t}^{(2)\ i_1,i_2}(X). \tag{59}$$

In general,

$$Y_t^n = y_0 + \sum_{k=1}^{n} \sum_{i_1,\ldots,i_k=1}^{d} V_{i_k} \circ \ldots \circ V_{i_1}(y_0) \cdot S_{0,t}^{(k)\ i_1,i_2,\ldots,i_k}(X). \tag{60}$$

The n-th term in the Picard iteration is a linear combination of the signatures up to the n-th order. This aligns with the motivation of signatures we currently have, as objects that are good at identifying the paths. The example shows the natural appearance of iterated integrals in the Picard iteration, and the power of the ODE test for path identification essentially comes from the power of signatures.

In machine learning contexts, one typically collects sequential data (time series) and formulates it as paths. The signatures of paths are then used to build algorithms for prediction/classification. In terms of **formulating (discretely collected) data as (continuous) paths**, there are several popular approaches to take:

- Linear interpolation. Assume we have a sequence of two data points $(x_0, y_0), (x_1, y_1)$ in $\mathbb{R}^2$, linear interpolation provides $X_t = (1-t)(x_0, y_0) + t(x_1, y_1)$ on $[0, 1]$.

- Time augmentation. Augment the data $(x_0, y_0), (x_1, y_1)$ into $(0, x_0, y_0), (1, x_1, y_1)$, with the first component recording the time t. Linear interpolation can be carried out after time augmentation. This fits with the situation where the time parameterization matters, e.g., the changes of stock prices as time goes by. For example, let the one-dimensional path $X_t = t\mathbb{I}_{0<t<1} + (2-t)\mathbb{I}_{1<t<2}$ be interpreted as the stock price. Signatures won't tell the difference from $\tilde{X}_t = 2t\mathbb{I}_{0<t<\frac{1}{2}} + \frac{4-2t}{3}\mathbb{I}_{\frac{1}{2}<t<2}$ without time augmentation. However, X has the maximum price at time 1 while $\tilde{X}$ has the maximum price at time $\frac{1}{2}$, which matters a lot for trading. In brief, **time augmentation breaks time reparameterization invariance and removes tree-like structures** due to the strict monotonicity of $t \mapsto t$, but the signature is still invariant under a constant translation.

- Lead-lag augmentation. Augment the data $(x_0, y_0), (x_1, y_1), ..., (x_k, y_k)$ into $(x_n, y_n, ..., x_{n-k}, y_{n-k})$ to record the history. This breaks time reparameterization invariance and takes into account **autocorrelations** (time variability). Consider the second-order signature of lead-lag augmentation (X_t, X_{t-1}), where X is a path in $\mathbb{R}$,

$$S_{0,T}^{(2)\ 1,2}(X_\cdot, X_{\cdot-1}) = \int_0^T (X_t - X_0)\,\mathrm{d}X_{t-1}. \tag{61}$$

captures the autocorrelation of time lag 1. When the data is one-dimensional, signatures provide trivial information (polynomials of $X_T - X_0$), hence lead-lag augmentation is a nice operation that overcomes this issue and maintains information of temporal dependencies.

- Basepoint augmentation. Augment the data $(x_0, y_0), (x_1, y_1), ..., (x_k, y_k)$ into $(x_0, y_0, x_1, y_1), ..., (x_0, y_0, x_k, y_k)$, with the first component recording the basepoint. This breaks basepoint invariance of signatures and adapts to the situation where a constant translation in the path matters, e.g., change of trading strategies based on different absolute prices.

- Rough path interpolation. When there is prior knowledge on path regularity, one may use rougher paths to interpolate discrete data after augmentations. For example, one may use Brownian bridges for stock price data, and fractional Brownian bridges with a given Hurst index for stochastic volatilities in a fractional environment, etc.

Which augmentation to use depends on what invariance property of signature one hopes to break and what extra information one hopes to maintain. Typically a combination of different augmentations are used in practice. Which interpolation to use depends on the prior knowledge of path regularity and the specific modeling object.

Example 9. *Online classification of handwritten digits. Each training sample contains a collection of coordinates* $\{(x_1^i, y_1^i), ..., (x_8^i, y_8^i)\}$ *of* 8 *sequentially ordered points in* $\mathbb{R}^2$*, sampled when one is writing a digit. Each training sample is labelled with the associated true digit. Linear interpolation is carried out to turn discrete data into continuous paths, based on which signatures are calculated (truncated up to an order no larger than* 10*-th typically). Afterwards, a LASSO is carried out based on signatures of the paths, which achieves a comparable prediction accuracy to neural networks classifiers.*

Similarly, there is examples for classification of the country of a stock. Each training sample consists of daily closing price and trading volume of a stock collected within a year (a time series of length 365*), and each sample is labeled by the country it belongs to. Using the same approach as above, the classifier also achieves a comparable accuracy.*

Example 10. *Hypothesis testing for probability measures on the path space. Given* $\mathbb{P}, \mathbb{Q}$ *as probability measures on the path space (infinite-dim), one hope to determine if* $\mathbb{P} = \mathbb{Q}$ *based on the independent empirical observations:* $X^1, ..., X^n$ *as sample paths from* $\mathbb{P}$*,* $Y^1, ..., Y^n$ *as sample paths from* $\mathbb{Q}$*.*

The approach is based on a metric on the space of path space measures

$$d(\mathbb{P}, \mathbb{Q}) := \sqrt{\|\mathbb{E}_{X\sim\mathbb{P}} S(X) - \mathbb{E}_{Y\sim\mathbb{Q}} S(Y)\|^2}. \tag{62}$$

The positivity of this metric follows from Theorem 8. Expanding the squares yields a representation

$$d^2(\mathbb{P},\mathbb{Q}) = \mathbb{E}_{(X,X')\sim\mathbb{P}\otimes\mathbb{P}}[K(X,X')] + \mathbb{E}_{(Y,Y')\sim\mathbb{Q}\otimes\mathbb{Q}}[K(Y,Y')] - 2\mathbb{E}_{(X,Y)\sim\mathbb{P}\otimes\mathbb{Q}}[K(X,Y)], \tag{63}$$

where the kernel

$$K(X,Y) := \sum_I \langle S^I(X)_{0,T}, S^I(Y)_{0,T}\rangle \tag{64}$$

can be computed if the signatures are truncated up to a certain level.

The test statistic $T_n(X^1,...,X^n,Y^1,...,Y^n)$ is an unbiased estimator for $d^2(\mathbb{P},\mathbb{Q})$:

$$T_n := \frac{1}{n(n-1)}\sum_{i\neq j} K(X^i,X^j) + \frac{1}{n(n-1)}\sum_{i\neq j} K(Y^i,Y^j) - \frac{2}{n^2}\sum_{i,j} K(X^i,Y^j). \tag{65}$$

One rejects $\mathbb{P}=\mathbb{Q}$ if $T_n > c$ is large enough. In addition, one can prove that the Type I error probability is less than α, uniformly for $\forall\mathbb{P},\mathbb{Q}$, if $c = 4\sqrt{-\frac{\log\alpha}{n}}$ (rejection region does not depend on the selection of measures).

This approach is different from sequential probability ratio test in the sense that it is ***model-free*** *(non-parametric), while SPRT requires knowing the dynamics that generates the paths to calculate likelihood ratios, and it only works for certain types of dynamics, e.g., paths of diffusions with different drifts.*

Fundamentals of Rough Path Theory

Construction of Rough Path Space

In the discussions above, we focus on signatures defined for paths of finite variation. However, most paths of interesting stochastic processes are rough and has infinite variation, for which the Lebesgue-Steiljes integral can not be defined. Here we focus on paths $X : [0, T] \to V$ in a Banach space V that are α-Holder continuous, i.e.,

$$\|X\|_\alpha := \sup_{t \neq s,\ t,s \in [0,T]} \frac{\|X_t - X_s\|}{|t-s|^\alpha} < \infty. \tag{66}$$

Denote C^α as the collection of all such α-Holder continuous paths, and $\|\cdot\|_\alpha$ is a semi-norm on C^α. Note that for BM sample paths, α can take any positive values strictly less than $\frac{1}{2}$. When $\alpha = 1$, the continuity is Lipschitz continuity and when $\alpha > 1$, the path must be constant.

When $\alpha \in [\frac{1}{2}, 1)$, stochastic integral can be constructed through an isometry induced by the quadratic variation process $t \mapsto \langle X, X \rangle_t$, which is increasing and has finite variation. However, when $\alpha < \frac{1}{2}$, no general integral theory guarantees the existence of the iterated integrals within the definition of signatures. In this case, one has Young's integral $\int_0^T Y_t \, \mathrm{d}X_t$ to be well-defined for $X \in C^\alpha$ and $Y \in C^\beta$ where $\alpha + \beta > 1$, which is problematic even for BM sample paths.

The idea of rough path theory is that, for $X \in C^\alpha$, one can construct an object $(X_{s,t}, \mathbb{X}^{(2)}_{s,t}, ..., \mathbb{X}^{(n)}_{s,t})$ for $\forall s < t$ that mimics the properties of the signatures for paths of finite variation. In particular, $X_{s,t} = X_t - X_s$ is the first-order signature (the path itself), while $\mathbb{X}^{(k)}_{s,t}$ is interpreted (not exactly) as the "signature" of the k-th order on time $[s, t]$. Therefore, one needs to postulate values for those "signatures" to make sure that they share the same behavior as signatures for paths of finite variation. From the basic understanding on signatures for paths of finite variation, we notice that Chen's identity describes the consistency of signatures on different time intervals, which is of essential importance. As a result, the following definition of rough path space is based on a natural extension of Chen's identity.

Definition. *For a path $X \in C^\alpha([0, T], \mathbb{R}^d)$ where $\alpha \in (0, 1]$, let $\Delta_T := \{(s, t) : 0 \leq s \leq t \leq T\}$ denote the collection of time sub-intervals.*

$$\mathbf{X} := (1, \mathbb{X}^{(1)}, \mathbb{X}^{(2)}, ..., \mathbb{X}^{\lfloor \frac{1}{\alpha} \rfloor}) \tag{67}$$

is called an α-Holder rough path associated with X if the following conditions hold:

1. *$\mathbb{X}^{(1)}_{s,t} = X_t - X_s,\ \forall (s, t) \in \Delta_T$ is the path itself.*
2. *Chen's identity holds: $\mathbb{X}^{(k)\ i_1,...,i_k}_{s,t} = \sum_{m=0}^k \mathbb{X}^{(m)\ i_1,...,i_m}_{s,u} \cdot \mathbb{X}^{(k-m)\ i_{m+1},...,i_k}_{u,t},\ \forall 0 \leq s \leq u \leq t \leq T,\ \forall k \leq \lfloor \frac{1}{\alpha} \rfloor$.*
3. *The Holder regularity requirement holds: for any $k \leq \lfloor \frac{1}{\alpha} \rfloor$, $\exists C > 0, \|\mathbb{X}^{(k)}_{s,t}\| \leq C(t-s)^{k\alpha},\ \forall (s, t) \in \Delta_T$.*

Denote by $\Omega^\alpha([0, T], \mathbb{R}^d)$ the collection of all α-Holder rough paths.

Remark. *The rough path object $\mathbf{X}$ is a **lifting** of the path X, and Chen's identity serves as the defining relationship of signatures. As a result, signatures are not defined analytically for rough paths (as a limit of discrete sum), but* *algebraically. Condition 3 follows from the intuition that if $Y \in C^\beta, X \in C^\alpha$ and Young's integral $\int_s^t (Y_u - Y_s)\,\mathrm{d}X_u$ is well-defined, then $\exists C > 0, \int_s^t (Y_u - Y_s)\,\mathrm{d}X_u \le C|t-s|^{\alpha+\beta}$ for $\forall (t,s) \in \Delta_T$. Note that $\mathbf{X}$ is basepoint invariant, i.e., it does not contain any information of X_0.*

Remark. *Note that $\mathbb{X}_{s,t}^{(k)} = \int_s^t \mathbb{X}_{s,u}^{(k-1)}\,\mathrm{d}X_u$ is a function in both variables s and t. The integrand and integration domain both changes as s changes, which means that, condition 3 does **NOT** imply that $\mathbb{X}_{0,t}^{(k)}$ is $k\alpha$-Holder continuous in t! For example, the second-order signature for Ito rough path of BM is $\mathbb{X}_{0,t}^{(2)} = \frac{W_t^2 - t}{2}$, which is NOT α-Holder continuous in t for any $\alpha < 1$ (actually it only holds for any $\alpha < \frac{1}{2}$), but condition 3 holds for $\mathbb{X}_{s,t}^{(2)} = \int_s^t (W_u - W_s)\,\mathrm{d}W_u = \frac{(W_t - W_s)^2}{2} - \frac{t-s}{2}$.*

The following theorem proves that, given the construction of the rough path space, signatures of higher orders are uniquely determined such that Chen's identity holds. As a result, only signatures up to order $\lfloor \frac{1}{\alpha} \rfloor$ have to be postulated, and those of higher orders can be defined through Young's integral.

Theorem 9 (Lyons)**.** *Given a rough path object $\mathbf{X} \in \Omega^\alpha([0,T], \mathbb{R}^d)$, there exists a unique signature extension $(1, \mathbb{X}^{(1)}, \mathbb{X}^{(2)}, ..., \mathbb{X}^{\lfloor \frac{1}{\alpha} \rfloor}, \mathbb{X}^{\lfloor \frac{1}{\alpha} \rfloor + 1}, ...)$ such that Chen's identity holds.*

Proof. For $k > \lfloor \frac{1}{\alpha} \rfloor$, define recursively by Young's integral $\mathbb{X}_{s,t}^{(k)} := \int_s^t \mathbb{X}_{s,u}^{(k-1)}\,\mathrm{d}X_u$. Since $(k-1)\alpha + \alpha > 1$, Young's integral is well-defined. Check that Chen's identity still holds. □

In the following context, we only consider the case where $\frac{1}{3} < \alpha \le \frac{1}{2}$ without specification, i.e., the rough path object has the form $\mathbf{X} = (1, \mathbb{X}^{(1)}, \mathbb{X}^{(2)})$, whose existence is guaranteed (Lyons, Victoir). For convenient notations, we denote $\mathbf{X} := (X, \mathbb{X}) \in \Omega^\alpha([0,T], \mathbb{R}^d)$, where X is the path and $\mathbb{X}$ is the second order signature. In this case, Chen's identity implies

$$\mathbb{X}_{s,t}^{i,j} = \mathbb{X}_{s,u}^{i,j} + X_{s,u}^i \cdot X_{u,t}^j + \mathbb{X}_{u,t}^{i,j},\ \forall 0 \le s \le u \le t \le T,\ \forall (i,j) \in [d]^2. \tag{68}$$

Hence, X is additive in time while $\mathbb{X}$ is not. For notational purpose, we denote $X \in C^\alpha$ and $\mathbb{X} \in C^{2\alpha}$ for the Holder regularity condition in the definition of the rough path space, but we remind the readers once again that $\mathbb{X}_{s,t}$ depends on both s, t and should **NEVER** be understood as a 2α-Holder continuous path.

Example 11. *Consider the Ito rough path for BM sample paths B when $d = 2$. The second order signature*

$$\mathbb{B}_{s,t}^{\mathrm{Ito}} = \begin{bmatrix} \int_s^t B_u^1 - B_s^1\,\mathrm{d}B_u^1 & \int_s^t B_u^1 - B_s^1\,\mathrm{d}B_u^2 \\ \int_s^t B_u^2 - B_s^2\,\mathrm{d}B_u^1 & \int_s^t B_u^2 - B_s^2\,\mathrm{d}B_u^2 \end{bmatrix}. \tag{69}$$

The diagonal entries can be explicitly calculated via Ito's formula (trivial), while the off-diagonal entries involve area information of the two-dimensional BM (nontrivial). The fact that $\mathbb{B}^{\mathrm{Ito}} \in C^{2\alpha}$ depends on a Kolmogorov-type regularity criterion for random rough paths and is nontrivial. Nevertheless, note that Theorem 9 claims the uniqueness of the extension of a given rough path object that respects Chen's identity, but it does not claim the uniqueness of the rough path object itself. In fact, multiple rough path objects can be constructed for a same path. For example, one can easily define the Stratonovich rough path $\mathbb{B}_{s,t}^{\mathrm{Strat}}$ for BM sample paths under a different notion of stochastic integration. Rough path liftings can always be defined for general semi-martingales and Markov processes.

Example 12. *Gaussian rough path. Let X be the sample path of a Gaussian process in $\mathbb{R}^d$, for which*

$$\mathbb{X}_{s,t}^{i,j} := \lim_{n\to\infty}\sum_{i=1}^{n} X_{s,t_i}^i X_{t_i,t_{i+1}}^j,\ \forall i \neq j \tag{70}$$

is well-defined in the L^2 sense for a partition $\{t_i\}$ of $[s,t]$ (under regularity assumptions), while

$$\mathbb{X}_{s,t}^{i,i} := \frac{1}{2}(X_{s,t}^i)^2,\ \forall i. \tag{71}$$

Such a construction provides a geometric rough path. Examples include fBM with Hurst index $H > \frac{1}{4}$, OU process and fOU process.

The following theorem comments on the uniqueness of rough path lifting for a given path.

Theorem 10. *If $(X, \mathbb{X}) \in \Omega^\alpha([0,T], \mathbb{R}^d)$, then $(X, \mathbb{X}') \in \Omega^\alpha([0,T], \mathbb{R}^d)$ iff there exists a 2α-Holder continuous function $G : [0,T] \to \mathbb{R}^{d\times d}$, such that*

$$\mathbb{X}'_{s,t} = \mathbb{X}_{s,t} + G(t) - G(s). \tag{72}$$

Proof. If $\mathbb{X}'$ is given as above, check Chen's identity:

$$\mathbb{X}'_{s,t} = \mathbb{X}_{s,u} + X_{s,u}\otimes X_{u,t} + \mathbb{X}_{u,t} + G(t) - G(u) + G(u) - G(s) \tag{73}$$

$$= \mathbb{X}'_{s,u} + X_{s,u}\otimes X_{u,t} + \mathbb{X}'_{u,t},\ \forall 0 \le s \le u \le t \le T. \tag{74}$$

Check Holder regularity: $\mathbb{X}'_{s,t} \in C^{2\alpha}$ since $\exists C > 0,\ |G(t) - G(s)| \le C|t-s|^{2\alpha}$.

Conversely, if $(X, \mathbb{X}') \in \Omega^\alpha([0,T], \mathbb{R}^d)$, take G as a function such that $G(t) - G(s) := \mathbb{X}'_{s,t} - \mathbb{X}_{s,t}$. By Chen's identity,

$$G(t) - G(s) = \mathbb{X}'_{s,t} - \mathbb{X}_{s,t} = \mathbb{X}'_{s,u} - \mathbb{X}_{s,u} + \mathbb{X}'_{u,t} - \mathbb{X}_{u,t} = G(u) - G(s) + G(t) - G(u). \tag{75}$$

Since $\mathbb{X}, \mathbb{X}' \in C^{2\alpha}$, G is 2α-Holder continuous. $\square$

Remark. *For paths of lower regularity, the difference in rough path liftings for a same path X is determined by a sequence of functions $\{G_k\}_{k=2,3,\ldots,\lfloor\frac{1}{\alpha}\rfloor}$, where each $G_k : [0,T] \to \mathbb{R}^{d^k}$ is $k\alpha$-Holder continuous.*

The shuffle product identity of signatures essentially comes from integration by parts, which does not hold for any rough path liftings, e.g., the Ito rough path of BM sample paths. In particular, those with the shuffle product identity are called (weakly) geometric rough paths.

Definition. $(X, \mathbb{X})$ *is called a (weakly) geometric rough path if shuffle product identity holds, i.e.,*

$$X_{s,t}^i \cdot X_{s,t}^j = \mathbb{X}_{s,t}^{i,j} + \mathbb{X}_{s,t}^{j,i},\ \forall (s,t) \in \Delta_T,\ \forall (i,j) \in [d]^2. \tag{76}$$

For any rough path lifting $(X, \mathbb{X})$, $[X] : [0,T] \to \mathbb{R}^{d\times d}$ is defined as

$$[X]^{i,j}_t := X^i_{0,t} \cdot X^j_{0,t} - (\mathbb{X}^{i,j}_{0,t} + \mathbb{X}^{j,i}_{0,t}), \tag{77}$$

and is called the rough bracket of $(X, \mathbb{X})$.

Remark. *A rough path $(X, \mathbb{X})$ is geometric iff $[X] \equiv 0$. When $[X] \equiv 0$, for any $0 \le s \le t \le T$,*

$$X^i_{s,t} \cdot X^j_{s,t} = (X^i_{0,t} - X^i_{0,s}) \cdot (X^j_{0,t} - X^j_{0,s}) \tag{78}$$

$$= \mathbb{X}^{i,j}_{0,t} + \mathbb{X}^{j,i}_{0,t} + \mathbb{X}^{i,j}_{0,s} + \mathbb{X}^{j,i}_{0,s} - (X^i_{0,t}X^j_{0,s} + X^i_{0,s}X^j_{0,t}) \tag{79}$$

$$= \mathbb{X}^{i,j}_{0,s} + X^i_{0,s}X^j_{s,t} + \mathbb{X}^{i,j}_{s,t} + \mathbb{X}^{j,i}_{0,s} + X^j_{0,s}X^i_{s,t} + \mathbb{X}^{j,i}_{s,t} + \mathbb{X}^{i,j}_{0,s} + \mathbb{X}^{j,i}_{0,s} - (X^i_{0,t}X^j_{0,s} + X^i_{0,s}X^j_{0,t}), \tag{80}$$

where we have used Chen's identity. Further calculations show that

$$X^i_{s,t} \cdot X^j_{s,t} = 2\mathbb{X}^{i,j}_{0,s} + \mathbb{X}^{i,j}_{s,t} + 2\mathbb{X}^{j,i}_{0,s} + \mathbb{X}^{j,i}_{s,t} - (X^i_{0,s}X^j_{0,s} + X^i_{0,s}X^j_{0,s}) \tag{81}$$

$$= \mathbb{X}^{i,j}_{s,t} + \mathbb{X}^{j,i}_{s,t}, \tag{82}$$

where we use $[X]_s = 0$ once more. Therefore, we have shown that $[X]_{0,t} = 0,\ \forall t \in [0,T]$ implies $[X]_{s,t} = 0,\ \forall (s,t) \in \Delta_T$, where

$$[X]^{i,j}_{s,t} := X^i_{s,t} \cdot X^j_{s,t} - (\mathbb{X}^{i,j}_{s,t} + \mathbb{X}^{j,i}_{s,t}). \tag{83}$$

In general, we have the following additive relationship: $[X]_{s,t} = [X]_{0,t} - [X]_{0,s}$.

Example 13. *Consider Ito and Stratonovich rough path liftings for BM sample path B. Clearly,*

$$[B]^{i,j}_t = B^i_t \cdot B^j_t - \left(\int_0^t B^i_s \circ \mathrm{d}B^j_s + \int_0^t B^j_s \circ \mathrm{d}B^i_s \right) = 0,\ \forall (i,j) \in [d]^2, \tag{84}$$

under Stratonovich integration. Therefore, $(B, \mathbb{B}^{\mathrm{Strat}})$ is geometric.

By Ito's formula,

$$\mathrm{d}(B^i_t \cdot B^j_t) = B^i_t\,\mathrm{d}B^j_t + B^j_t\,\mathrm{d}B^i_t,\ \forall i \ne j, \quad \mathrm{d}(B^i_t)^2 = 2B^i_t\,\mathrm{d}B^i_t + \mathrm{d}t,\ \forall i. \tag{85}$$

As a result, $[B]^{i,j}_t = t\mathbb{I}_{i=j}$, which implies $(B, \mathbb{B}^{\mathrm{Ito}})$ is not geometric. In addition, their signatures are different by a Lipschitz ($\alpha = 1$) function $[G(t)]_{i,j} = \frac{t}{2}\mathbb{I}_{i=j}$.

Remark. *Which rough path lifting to take depends on what property one hopes to impose, e.g., chain rule (geometric rough path), zero mean property (Ito integral), etc. Modifications can be made by changing the 2α-Holder continuous function G.*

Metric on Rough Path Space

To equip Ω^α with a topology, we define the following rough path metric, which has the natural form as the sum of the distance between paths and their signatures.

Definition. *For $\forall (X,\mathbb{X}),(Y,\mathbb{Y}) \in \Omega^\alpha([0,T],\mathbb{R}^d)$, the rough path metric is defined as the sum of Holder semi-norms:*

$$d^\alpha((X,\mathbb{X}),(Y,\mathbb{Y})) := \|X-Y\|_\alpha + \|\mathbb{X}-\mathbb{Y}\|_{2\alpha} = \sup_{0\le s<t\le T} \frac{\|X_{s,t}-Y_{s,t}\|}{|t-s|^\alpha} + \sup_{0\le s<t\le T} \frac{\|\mathbb{X}_{s,t}-\mathbb{Y}_{s,t}\|}{|t-s|^{2\alpha}}. \tag{86}$$

The following theorem proves the completeness of the rough path space as a metric space.

Theorem 11. *The rough path space $(\Omega^\alpha, d^\alpha)$ is complete. The collection of weakly geometric rough path liftings $\mathrm{WG}(\Omega^\alpha)$ is a closed subspace of Ω^α.*

This topology is a really fine one, under which continuity is easier to obtain. The Ito map that maps a sample point ω to the path of the solution of an SDE under ω, is continuous under the rough path topology, but is not continuous w.r.t. ANY topology on the path space (Lyons).

Example 14. *Consider a sequence of paths in $\mathbb{R}^2$, namely $X^n : t \mapsto (\frac{\cos(2\pi n^2 t)}{n}, \frac{\sin(2\pi n^2 t)}{n})$. Compute the second-order signature*

$$\begin{aligned}
\mathbb{X}^{n\ 1,1}_{s,t} &= \int_s^t (X^1_u - X^1_s)\,\mathrm{d}X^1_u && (87)\\
&= \int_s^t \frac{\cos(2\pi n^2 u)-\cos(2\pi n^2 s)}{n}\,\mathrm{d}\frac{\cos(2\pi n^2 u)}{n} && (88)\\
&= -\pi\int_s^t \sin(4\pi n^2 u)\,\mathrm{d}u - \frac{\cos(2\pi n^2 s)}{n}\cdot\frac{\cos(2\pi n^2 t)-\cos(2\pi n^2 s)}{n} && (89)\\
&= \frac{\cos(4\pi n^2 t)-\cos(4\pi n^2 s)}{4n^2} - \frac{\cos(2\pi n^2 s)}{n}\cdot\frac{\cos(2\pi n^2 t)-\cos(2\pi n^2 s)}{n} \to 0\ (n\to\infty). && (90)
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\mathbb{X}^{n\ 1,2}_{s,t} &= \int_s^t (X^1_u - X^1_s)\,\mathrm{d}X^2_u && (91)\\
&= \int_s^t \frac{\cos(2\pi n^2 u)-\cos(2\pi n^2 s)}{n}\,\mathrm{d}\frac{\sin(2\pi n^2 u)}{n} && (92)\\
&= 2\pi\int_s^t \cos^2(2\pi n^2 u)\,\mathrm{d}u - \frac{\cos(2\pi n^2 s)}{n}\cdot\frac{\sin(2\pi n^2 t)-\sin(2\pi n^2 s)}{n} && (93)\\
&= \pi(t-s) + \frac{\sin(4\pi n^2 t)-\sin(4\pi n^2 s)}{4n^2} - \frac{\cos(2\pi n^2 s)}{n}\cdot\frac{\sin(2\pi n^2 t)-\sin(2\pi n^2 s)}{n} \to \pi(t-s)\ (n\to\infty). && (94)
\end{aligned}$$

As observed, $X^n_{s,t}\to 0$, which is identical to the path that constantly stays at the origin. However, the limit of area information $\mathbb{X}^n_{s,t} \to \begin{bmatrix} 0 & \pi(t-s) \\ -\pi(t-s) & 0 \end{bmatrix}$ distinguishes itself from a constant path. The rough path lifting maintains path information in the limit $n\to\infty$ and does record how the convergence is happening.

Rough Integral

Now that we have defined the signatures (integral of X w.r.t. X itself) algebraically, it is natural to expect that the integral of Y w.r.t. X can be defined for the paths Y that locally looks like X, e.g., $Y_t = F(X_t)$ for a smooth enough F. For a given $\mathbf{X} = (X, \mathbb{X}) \in \Omega^\alpha([0,T], \mathbb{R}^d)$, the class of paths controlled by X is the collection of paths that are "locally similar" to X and is identified as the integrand for rough integrals.

Definition. *$Y \in C^\alpha([0,T], \mathbb{R}^m)$ is defined to be controlled by X if there exists $Y' \in C^\alpha$ such that the remainder term of Y on $[s,t]$,*

$$R^Y_{s,t} := Y_{s,t} - Y'_s X_{s,t} \in C^{2\alpha}. \tag{95}$$

Denote by $D^{2\alpha}_X([0,T], \mathbb{R}^d)$, the collection of pairs (Y, Y') defined above, as the space of paths controlled by X. Here Y' is called the Gubinelli derivative of Y w.r.t. X, and the controlled path space is equipped with the semi-norm $\|(Y,Y')\|_{X,2\alpha} := \|Y'\|_\alpha + \|R^Y\|_{2\alpha}$.

Remark. *In the definition above, the linear mapping $Y'_s : \mathbb{R}^d \to \mathbb{R}^m$ so the operation in the expression $Y'_s X_{s,t}$ shall be understood as function evaluation $Y'_s(X_{s,t})$. For the convenience of the notation, we omit the parenthesis.*

Example 15. *Consider a 2α-Holder continuous path ϕ_t, for which*

$$R^\phi_{s,t} = \phi_{s,t} \in C^{2\alpha}. \tag{96}$$

As a result, $(\phi, 0)$ is a path controlled by any $X \in C^\alpha$. Therefore, if the path regularity is strong enough, the derivative is naturally 0.

Example 16. *Consider $X \in C^\alpha$ and $Y_t = X_t - X_0$, for which*

$$R^Y_{s,t} = (X_t - X_0) - (X_s - X_0) - X_{s,t} = 0 \in C^{2\alpha}. \tag{97}$$

Therefore, $(X_\cdot - X_0, 1)$ is a path controlled by X. The derivative of X w.r.t. itself is constantly 1.

Example 17. *Consider $\mathbf{X} = (X, \mathbb{X}) \in \Omega^\alpha$ and $Y_t = \mathbb{X}_{0,t}$ taking values in $\mathbb{R}^{d\times d}$, for which*

$$R^Y_{s,t} = \mathbb{X}_{0,t} - \mathbb{X}_{0,s} - X_{0,s} \otimes X_{s,t} = \mathbb{X}_{s,t} \in C^{2\alpha}, \tag{98}$$

by Chen's identity and the definition of rough path lifting. As a result, Y'_s is a mapping $\mathbb{R}^d \to \mathbb{R}^{d\times d}$, whose action is determined by $X_{s,t} \mapsto X_{0,s} \otimes X_{s,t}$. As a shorthand notation, we say $(\mathbb{X}_{0,\cdot}, X_{0,\cdot})$ is a path controlled by X.

Example 18. *Consider $X \in C^\alpha$ and $Y_t = F(X_t)$, where $F : \mathbb{R}^d \to \mathbb{R}^m$* ***is second-order continuously differentiable****.*

$$R^Y_{s,t} = F(X_t) - F(X_s) - F'(X_s)(X_t - X_s) = \frac{1}{2} F''(X_\xi)(X_t - X_s)^2 \in C^{2\alpha}, \tag{99}$$

for some $\xi \in [s,t]$ by Taylor formula. Therefore, $(F(X), F'(X))$ is a path controlled by X, $Y'_s = F'(X_s)$.

Remark. *Without enough regularity of F, the conclusion above no longer holds, e.g., F could be a sample path of BM which is rough.*

Intuitively, the Gubinelli derivative has most properties of a classical derivative, and the remainder is required to be $C^{2\alpha}$ for the rough integral to be well-defined. The following example contains properties of the Gubinelli derivative.

Example 19. *If $(Y, Y') \in D_X^{2\alpha}$ is a path controlled by X, check that* ***chain rule*** *holds: $(F(Y), F'(Y) \circ Y') \in D_X^{2\alpha}$ for $F : \mathbb{R}^m \to \mathbb{R}^k$ which is second-order continuously differentiable. Here $\circ$ denotes function composition (matrix product). The dimension aligns since $F'(Y) : \mathbb{R}^m \to \mathbb{R}^k$ while $Y' : \mathbb{R}^d \to \mathbb{R}^m$ so $F'(Y) \circ Y' : \mathbb{R}^d \to \mathbb{R}^k$. By definition,*

$$
\begin{aligned}
R_{s,t}^{F(Y)} &= F(Y_t) - F(Y_s) - F'(Y_s) \circ Y_s'(X_t - X_s) && (100)\\
&= F'(Y_s)(Y_t - Y_s) + \frac{1}{2}F''(Y_\xi)(Y_t - Y_s)^2 - F'(Y_s) \circ Y_s'(X_t - X_s) && (101)\\
&= F'(Y_s)R_{s,t}^Y + \frac{1}{2}F''(Y_\xi)(Y_t - Y_s)^2, && (102)
\end{aligned}
$$

for some $\xi \in [s, t]$. By definition, both $R_{s,t}^Y$ and $(Y_t - Y_s)^2$ are $C^{2\alpha}$ and F', F'' are bounded, which concludes the proof.

Example 20. *If $(Y, Y'), (Z, Z') \in D_X^{2\alpha}$ are paths controlled by X (a path in $\mathbb{R}^d$), where Y takes values in $\mathbb{R}^k \to \mathbb{R}^m$ and Z takes values in $\mathbb{R}^k$, check that* ***product rule*** *holds: $(YZ, Y'Z + Y \circ Z') \in D_X^{2\alpha}$. Firstly we check the consistency of dimensions: $YZ \in \mathbb{R}^m$, while $Y' : \mathbb{R}^d \to (\mathbb{R}^k \to \mathbb{R}^m)$ and $Z' : \mathbb{R}^d \to \mathbb{R}^k$ so that $Y'Z : \mathbb{R}^d \to \mathbb{R}^m$ is defined through the action $Y'Z(x) := [Y'(x)](Z)$. On the other hand, $Y \circ Z' : \mathbb{R}^d \to \mathbb{R}^m$ is defined through function composition $Y \circ Z'(x) := Y(Z'(x))$. We remind the readers to distinguish between three operations: function evaluation (denoted as $(\cdot)$ or direct concatenation), function composition (matrix product denoted as $\circ$), and tensor product (denoted as $\otimes$), in the context in order to match dimensions.*

By definition,

$$
\begin{aligned}
R_{s,t}^{YZ} &= Y_tZ_t - Y_sZ_s - (Y_s'Z_s + Y_s \circ Z_s')(X_t - X_s) && (103)\\
&= [Y_tZ_t - Y_sZ_t - Y_s'(X_t - X_s)(Z_s)] + [Y_sZ_t - Y_sZ_s - Y_s \circ Z_s'(X_t - X_s)] && (104)\\
&= [(Y_s'(X_t - X_s) + R_{s,t}^Y)Z_t - Y_s'(X_t - X_s)(Z_s)] + [Y_s(Z_s'(X_t - X_s) + R_{s,t}^Z) - Y_s \circ Z_s'(X_t - X_s)] && (105)\\
&= Y_s'(X_t - X_s)Z_s'(X_t - X_s) + Y_s'(X_t - X_s)R_{s,t}^Z + R_{s,t}^Y Z_t + Y_s R_{s,t}^Z \in C^{2\alpha}. && (106)
\end{aligned}
$$

This concludes the proof.

With everything ready, we introduce the construction of rough integrals. The motivation is from a simple result for integrals w.r.t. smooth paths.

Theorem 12. *Let X be a path on $\mathbb{R}$ that is $(n+1)$-th order continuously differentiable, and $f : \mathbb{R} \to \mathbb{R}$ is $(n+1)$-th order continuously differentiable, then*

$$
\int_0^t f(X_s)\,\mathrm{d}X_s = \lim_{|\Delta| \to 0} \sum_{i=0}^{N-1} \sum_{k=0}^{n} \frac{f^{(k)}(X_{t_i})}{(k+1)!}(X_{t_{i+1}} - X_{t_i})^{k+1}, \tag{107}
$$

where Δ is a partition of $[0, t]$ with N intervals, and the partition gets finer $|\Delta| := \sup_i |t_{i+1} - t_i| \to 0$ as $N \to \infty$.

Proof. Apply Taylor expansion for $f(X_s)$ at t_i for any $s \in [t_i, t_{i+1}]$:

$$f(X_s) = \sum_{k=0}^{n} \frac{f^{(k)}(X_{t_i})}{k!}(X_s - X_{t_i})^k + \frac{f^{(n+1)}(X_{\xi_s})}{(n+1)!}(X_s - X_{t_i})^{n+1}, \tag{108}$$

for some $\xi_s \in [t_i, s]$ (intermediate value theorem + continuity of $t \mapsto X_t$). Integrate both sides on $[t_i, t_{i+1}]$:

$$\int_{t_i}^{t_{i+1}} f(X_s)\,\mathrm{d}X_s = \sum_{k=0}^{n} \frac{f^{(k)}(X_{t_i})}{(k+1)!}(X_{t_{i+1}} - X_{t_i})^{k+1} + \int_{t_i}^{t_{i+1}} \frac{f^{(n+1)}(X_{\xi_s})}{(n+1)!}(X_s - X_{t_i})^{n+1}\,\mathrm{d}X_s. \tag{109}$$

By the regularity condition, $t \mapsto f^{(n+1)}(X_t)$ is continuous on $[0, T]$, thus bounded by some constant $M > 0$. Therefore,

$$\left|\int_{t_i}^{t_{i+1}} \frac{f^{(n+1)}(X_{\xi_s})}{(n+1)!}(X_s - X_{t_i})^{n+1}\,\mathrm{d}X_s\right| \le \frac{M}{(n+1)!}\int_{t_i}^{t_{i+1}} |X_s - X_{t_i}|^{n+1}|\,\mathrm{d}X_s|. \tag{110}$$

By the intermediate value theorem, $X_s - X_{t_i} = \dot{X}_{\eta_s}(s - t_i)$ for some $\eta \in [t_i, s]$, whereas $t \mapsto \dot{X}_t$ is continuous on $[0, T]$, thus bounded by D. As a result,

$$\frac{M}{(n+1)!}\int_{t_i}^{t_{i+1}} |X_s - X_{t_i}|^{n+1}|\,\mathrm{d}X_s| \le \frac{MD^{n+2}}{(n+1)!}\int_{t_i}^{t_{i+1}} (s - t_i)^{n+1}\,\mathrm{d}s = \frac{MD^{n+2}}{(n+2)!}(t_{i+1} - t_i)^{n+2}. \tag{111}$$

Summing both sides w.r.t. i yields

$$\int_0^t f(X_s)\,\mathrm{d}X_s = \sum_{i=0}^{N-1}\sum_{k=0}^{n} \frac{f^{(k)}(X_{t_i})}{(k+1)!}(X_{t_{i+1}} - X_{t_i})^{k+1} + R_N(s), \tag{112}$$

where the remainder

$$|R_N(s)| \le \sum_{i=0}^{N-1} \frac{MD^{n+2}}{(n+2)!}(t_{i+1} - t_i)^{n+2} \le |\Delta|^{n+1}\frac{MD^{n+2}}{(n+2)!}\sum_{i=0}^{N-1}(t_{i+1} - t_i) \to 0 \ (N \to \infty,\ |\Delta| \to 0). \tag{113}$$

□

Remark. *This conclusion is not very interesting for a smooth path X with finite variation, since higher-order terms in the sum do not actually contribute to the limit. However, in the context of rough path, this conclusion provides a nice way to generalize the definition of integrals, where $\frac{f^{(k)}(X_{t_i})}{(k+1)!}(X_{t_{i+1}} - X_{t_i})^{k+1}$ has nontrivial contribution to the limit if the $(k+1)$-th order variation of X is not constantly zero. When $\frac{1}{3} < \alpha \le \frac{1}{2}$, the second variation is nontrivial while the third variation is trivial. Therefore, we shall take $n = 1$, which yields*

$$\int_0^t f(X_s)\,\mathrm{d}X_s = \lim_{|\Delta| \to 0}\sum_{i=0}^{N-1}\left[f(X_{t_i})(X_{t_{i+1}} - X_{t_i}) + \frac{f'(X_{t_i})}{2}(X_{t_{i+1}} - X_{t_i})^2\right]. \tag{114}$$

Since $(f(X), f'(X))$ is a path controlled by X for a sufficiently regular f, we naturally expect that this definition of rough integral w.r.t. X can be extended for integrands Y that are paths controlled by X. In this general case, we would replace f' with the Gubinelli derivative Y' and replace the quadratic part with the second-order signature $\mathbb{X}_{t_i,t_{i+1}}$. This yields the following definition of rough integrals.

Definition. *For $(X, \mathbb{X}) \in \Omega^\alpha([0,T], \mathbb{R}^d)$ and $(Y, Y') \in D_X^{2\alpha}$ as a path in $\mathbb{R}^d \to \mathbb{R}^m$, we hope to define the rough integral as*

$$\int_a^b (Y, Y') \,\mathrm{d}(X, \mathbb{X}) := \lim_{|\Delta| \to 0} \sum_{[s,t] \in \Delta_{a,b}} (Y_s X_{s,t} + Y'_s \mathbb{X}_{s,t}), \tag{115}$$

*where $\Delta_{a,b}$ is a partition of $[a,b] \subset [0,T]$. Note the **compensated Riemann-Stieljes sum**, which results from the roughness of the path. To clarify, this integral (if exists, which has not been checked yet) takes values in $\mathbb{R}^m$. The dimensions are consistent since $Y' : [\mathbb{R}^d \to (\mathbb{R}^d \to \mathbb{R}^m)] \cong (\mathbb{R}^{d\times d} \to \mathbb{R}^m)$ for linear mappings. The canonical isomorphism is defined on the space $\mathbb{R}^d \otimes \mathbb{R}^d$ through*

$$Y'(v \otimes w) := Y'(v)(w), \ \forall v, w \in \mathbb{R}^d, \tag{116}$$

and then linearly extended to the whole space $\mathbb{R}^{d\times d}$. By the universal property of tensor product, this mapping is well-defined, i.e., if a matrix has two different linear representations using tensor products, the mapping provides the same image. In later contexts, we refer to this canonical isomorphism without further explanation.

The next step is to prove that the rough integral is well-defined, which is based on the sewing lemma.

Lemma 2 (Sewing Lemma)**.** *Let*

$$C_2^{\alpha,\beta}([0,T], \mathbb{R}^d) := \left\{ \Theta : [0,T]^2 \to \mathbb{R}^d, \Theta \in C^\alpha, \sup_{s<u<t} \frac{|\Theta_{s,t} - \Theta_{s,u} - \Theta_{u,t}|}{|t-s|^\beta} < \infty \right\} \tag{117}$$

be the collection of double-time-index mappings whose additivity error is not too large. If $\forall 0 < \alpha < 1 < \beta$, there exists a unique linear mapping $I : C_2^{\alpha,\beta}([0,T], \mathbb{R}^d) \to C^\alpha([0,T], \mathbb{R}^d)$ such that for some constant $C > 0$,

$$(I\Theta)_0 = 0, \quad \|(I\Theta)_t - (I\Theta)_s - \Theta_{s,t}\| \le C|t-s|^\beta, \ \forall s, t \in [0,T]. \tag{118}$$

We shall understand $\Theta_{s,t}$ as local approximations of rough integrals on $[s,t]$, which is not necessarily additive. However, we can always find a unique linear mapping I that sews all local approximations into a global approximation.

Theorem 13. *For $(X, \mathbb{X}) \in \Omega^\alpha$ and $(Y, Y') \in D_X^{2\alpha}$, the rough integral $\int_0^t (Y, Y') \,\mathrm{d}(X, \mathbb{X})$ is well-defined and additive in time, i.e., $\int_0^s + \int_s^t = \int_0^t$, $\forall 0 \le s \le t \le T$.*

Proof. Identify $\Theta_{s,t} = Y_s \cdot X_{s,t} + Y'_s \cdot \mathbb{X}_{s,t}$ in the sewing lemma. By Chen's identity,

$$\Theta_{s,t} - \Theta_{s,u} - \Theta_{u,t} = -Y_{s,u} X_{u,t} - Y'_{s,u} \mathbb{X}_{u,t} + Y'_s (X_{s,u} \otimes X_{u,t}). \tag{119}$$

Use the canonical isomorphism and apply the definition of controlled path

$$|\Theta_{s,t}-\Theta_{s,u}-\Theta_{u,t}| = |-Y_{s,u}X_{u,t}-Y'_{s,u}\mathbb{X}_{u,t}+Y'_s(X_{s,u})(X_{u,t})| \tag{120}$$

$$= |-R^Y_{s,u}X_{u,t}-Y'_{s,u}\mathbb{X}_{u,t}| \le C|s-t|^{3\alpha} \tag{121}$$

for some constant $C>0$, since $R^Y,\mathbb{X}\in C^{2\alpha}$ and $X,Y'\in C^{\alpha}$. As a result $\beta=3\alpha>1$ in the sewing lemma, which guarantees the existence and uniqueness of the mapping I such that

$$(I\Theta)_0=0, \quad \|(I\Theta)_t-(I\Theta)_s-\Theta_{s,t}\| \le C|t-s|^{\beta},\ \forall s,t\in[0,T]. \tag{122}$$

Due to $\beta>1$, summing both sides w.r.t. all time sub-intervals within a partition yields

$$(I\Theta)_0=0, \quad (I\Theta)_t=\lim_{|\Delta|\to 0}\sum_{[t_i,t_{i+1}]\in\Delta_{0,t}}\Theta_{t_i,t_{i+1}},\ \forall t\in[0,T], \tag{123}$$

and the existence of the limit is justified. This enables us to define the rough integral in an additive way

$$\int_s^t (Y,Y')\,\mathrm{d}(X,\mathbb{X}) := (I\Theta)_t-(I\Theta)_s, \tag{124}$$

which is equivalent to

$$\int_s^t (Y,Y')\,\mathrm{d}(X,\mathbb{X}) = \lim_{|\Delta|\to 0}\sum_{[t_i,t_{i+1}]\in\Delta_{s,t}} Y_{t_i}X_{t_i,t_{i+1}}+Y'_{t_i}\mathbb{X}_{t_i,t_{i+1}}. \tag{125}$$

□

Remark. *For rougher paths, i.e., $\alpha\le\frac{1}{3}$, let $k=\lfloor\frac{1}{\alpha}\rfloor$ so that up to the k-th order variation of the path is nontrivial. As a result, one needs to truncate at level $n=k-1$ in the compensated Riemann-Stieljes sum $\sum_{j=0}^{n}\frac{f^{(j)}(X_{t_i})}{(j+1)!}(X_{t_{i+1}}-X_{t_i})^{j+1}$, which now contains k terms associated with $(X,\mathbb{X}^{(2)},...,\mathbb{X}^{(k)})$. To define rough integrals, one can extend the definition of controlled paths to a higher-order in an iterative way (higher-order Gubinelli derivatives)*

The difficulty is that, if the rough path is not geometric, then $(F(X),F'(X),...,F^{(k)}(X))$ might not be a path controlled by X even for a sufficiently regular F. As indicated by Theorem 10, one has the freedom to modify the j-th order signatures of a rough path by any $j\alpha$-Holder continuous function G_j for $j\in\{2,3,...,k\}$, without changing the first-level path X. When $k=2$ (discussed case), the single function G_2 can always be chosen such that the rough path becomes geometric, but this is not the case for $k\ge 3$. Therefore, the definition of rough integral does not require any geometric conditions for $\frac{1}{3}<\alpha\le\frac{1}{2}$ but geometricity is necessary for a general α.

Corollary 1. *For $(X,\mathbb{X})\in\Omega^{\alpha}$ and $(Y,Y')\in D_X^{2\alpha}$, $(\int_0^{\cdot}(Y,Y')\,\mathrm{d}(X,\mathbb{X}),Y)\in D_X^{2\alpha}$. The Gubinelli derivative of the rough integral is the integrand itself (Newton-Lebniz-type result).*

Proof. Let $Z_t:=\int_0^t(Y,Y')\,\mathrm{d}(X,\mathbb{X})$, so that

$$R^Z_{s,t}=Z_t-Z_s-Y_s(X_t-X_s)=\int_s^t(Y,Y')\,\mathrm{d}(X,\mathbb{X})-Y_s(X_t-X_s). \tag{126}$$

By sewing lemma, we get a powerful representation,

$$\int_s^t (Y, Y')\,\mathrm{d}(X, \mathbb{X}) = Y_s X_{s,t} + Y'_s \mathbb{X}_{s,t} + R_{s,t}, \quad R_{s,t} \in C^{3\alpha}. \tag{127}$$

Combining both parts yields

$$R^Z_{s,t} = Y'_s \mathbb{X}_{s,t} + R_{s,t} \in C^{2\alpha}, \tag{128}$$

which concludes the proof. □

Corollary 2. *For $(X, \mathbb{X}) \in \Omega^\alpha$ and Y as a 2α-Holder continuous path, it is clear that $(Y, 0) \in D_X^{2\alpha}$. In this case, $\int_s^t (Y, 0)\,\mathrm{d}(X, \mathbb{X}) = \int_s^t Y_u\,\mathrm{d}X_u$, where the LHS is rough integral and the RHS is Young's integral (exists since $2\alpha+\alpha = 3\alpha > 1$).* ***The definition of rough integral is consistent with Young's integral when the integrand has enough regularity.***

Proof. It directly follows from the definition that

$$\int_s^t (Y, 0)\,\mathrm{d}(X, \mathbb{X}) = \lim_{|\Delta|\to 0} \sum_{[t_i,t_{i+1}]\in\Delta_{s,t}} Y_{t_i} X_{t_i,t_{i+1}} = \int_s^t Y_u\,\mathrm{d}X_u. \tag{129}$$

□

Remark. *The high-level idea of rough integral: (i). local approximation on $[s, t]$ via compensated Riemann-Stieljes sum (ii). sewing lemma globalizes local approximations in a consistent way (iii). regularity of signatures and controlled paths guarantees the vanishing of the remainder. The introduction of the second-order signature into the compensated sum is mainly to ensure that different local approximations are showing certain intrinsic structures (additivity).*

Example 21. *Consider examples of rough integral*

$$\int_s^t (1, 0)\,\mathrm{d}(X, \mathbb{X})_u = \lim_{|\Delta|\to 0} \sum_{[t_i,t_{i+1}]\in\Delta_{s,t}} X_{t_i,t_{i+1}} = X_{s,t}. \tag{130}$$

By Chen's identity,

$$\int_s^t (X_{s,u}, 1)\,\mathrm{d}(X, \mathbb{X})_u = \lim_{|\Delta|\to 0} \sum_{[t_i,t_{i+1}]\in\Delta_{s,t}} X_{s,t_i} \otimes X_{t_i,t_{i+1}} + \mathbb{X}_{t_i,t_{i+1}} = \lim_{|\Delta|\to 0} \sum_{[t_i,t_{i+1}]\in\Delta_{s,t}} \mathbb{X}_{s,t_{i+1}} - \mathbb{X}_{s,t_i} = \mathbb{X}_{s,t}. \tag{131}$$

Example 22. *Consider the controlled path $(\mathbb{X}_{s,\cdot}, X_{s,\cdot}) \in D_X^{2\alpha}$, which is implied by Corollary 1 together with the example above. Note that $\mathbb{X}_{s,t} \in C^{2\alpha}$ but this does NOT imply that $t \mapsto \mathbb{X}_{s,t}$ is 2α-Holder continuous for any fixed s, hence $(\mathbb{X}_{s,\cdot}, 0) \notin D_X^{2\alpha}$. By the definition of rough integral,*

$$\mathbb{X}^{(3)}_{s,t} := \int_s^t (\mathbb{X}_{s,u}, X_{s,u})\,\mathrm{d}(X, \mathbb{X})_u \tag{132}$$

is well-defined and is understood as the third-order signature of the rough path. Higher-order signatures can be iteratively defined.

After defining the integration scheme, one might quantify the change in $F(t, X_t)$ caused by changes in time t, for a given function $F : [0, T] \times \mathbb{R}^d \to \mathbb{R}$. This result is an analogue to the Itô's formula in stochastic analysis, but in the pathwise sense (while the classical Itô's formula is not pathwise, but holds under the almost sure sense).

Theorem 14 (Rough Itô's formula)**.** *For $F : [0, T] \times \mathbb{R}^d \to \mathbb{R}^m$ which is $C^{1,3}$, and $(X, \mathbb{X}) \in \Omega^\alpha([0, T], \mathbb{R}^d)$,*

$$F(t, X_t) - F(s, X_s) = \int_s^t \partial_t F(u, X_u)\,\mathrm{d}u + \int_s^t \partial_x F(u, X_u)\,\mathrm{d}(X, \mathbb{X})_u + \frac{1}{2}\int_s^t \partial_{xx} F(u, X_u)\,\mathrm{d}[X]_u. \tag{133}$$

Proof. Prove for $m = 1$. Consider the Taylor expansion of $F(t, X_t)$ at time s:

$$\begin{aligned} F(t, X_t) &= F(s, X_s) + \partial_t F(s, X_s)(t - s) + \partial_x F(s, X_s) \cdot X_{s,t} + \frac{1}{2} X_{s,t}^T \cdot [\partial_{xx} F(s, X_s)] \cdot X_{s,t} + R_{s,t} & (134)\\ &= F(s, X_s) + \partial_t F(s, X_s)(t - s) + \partial_x F(s, X_s) \cdot X_{s,t} + \frac{1}{2}\mathrm{Tr}(\partial_{xx} F(s, X_s) \cdot (X_{s,t} \otimes X_{s,t})) + R_{s,t} & (135) \end{aligned}$$

where $R_{s,t} = O((t - s)^2)$. Applying this Taylor expansion on each interval within the partition $\Delta_{s,t}$ and taking $|\Delta| \to 0$ yields

$$\begin{aligned} F(t, X_t) - F(s, X_s) = \int_s^t \partial_t F(u, X_u)\,\mathrm{d}u &+ \lim_{|\Delta|\to 0} \sum_{[t_i,t_{i+1}]\in\Delta_{s,t}} \partial_x F(t_i, X_{t_i}) \cdot X_{t_i,t_{i+1}} & (136)\\ &+ \lim_{|\Delta|\to 0} \sum_{[t_i,t_{i+1}]\in\Delta_{s,t}} \frac{1}{2}\mathrm{Tr}(\partial_{xx} F(t_i, X_{t_i}) \cdot (X_{t_i,t_{i+1}} \otimes X_{t_i,t_{i+1}})), & (137) \end{aligned}$$

where the aggregation of the remainders vanish under the limit. By adding and subtracting the same term

$$\sum_{[t_i,t_{i+1}]\in\Delta_{s,t}} \mathrm{Tr}(\partial_{xx} F(t_i, X_{t_i}) \cdot \mathbb{X}_{t_i,t_{i+1}}), \tag{138}$$

the second term on the RHS becomes the rough integral $\int_s^t \partial_x F(u, X_u)\,\mathrm{d}(X, \mathbb{X})_u$, while the third term has the form

$$\begin{aligned} &\sum_{[t_i,t_{i+1}]\in\Delta_{s,t}} \mathrm{Tr}\left[\partial_{xx} F(t_i, X_{t_i}) \cdot \left(\frac{1}{2}(X_{t_i,t_{i+1}} \otimes X_{t_i,t_{i+1}}) - \mathbb{X}_{t_i,t_{i+1}}\right)\right] & (139)\\ &= \frac{1}{2}\sum_{[t_i,t_{i+1}]\in\Delta_{s,t}} \mathrm{Tr}\left[\partial_{xx} F(t_i, X_{t_i}) \cdot [X]_{t_i,t_{i+1}}\right] + \sum_{[t_i,t_{i+1}]\in\Delta_{s,t}} \mathrm{Tr}\left[\partial_{xx} F(t_i, X_{t_i}) \cdot \frac{\mathbb{X}_{t_i,t_{i+1}}^T - \mathbb{X}_{t_i,t_{i+1}}}{2}\right] & (140)\\ &= \frac{1}{2}\sum_{[t_i,t_{i+1}]\in\Delta_{s,t}} \mathrm{Tr}\left[\partial_{xx} F(t_i, X_{t_i}) \cdot [X]_{t_i,t_{i+1}}\right], & (141) \end{aligned}$$

where the first equation uses the definition of rough bracket. Taking the limit yields $\frac{1}{2}\int_s^t \partial_{xx} F(u, X_u)\,\mathrm{d}[X]_u$. This concludes the proof. ∎

Remark. *From Example 18, it is clear that* $(\partial_x F(\cdot, X.), \partial_{xx} F(\cdot, X.)) \in D_X^{2\alpha}$ *if* $\partial_x F$ *is* C^2 *in the second component, which justifies why* F *has to be* C^3 *in the space component. For simplicity, we omit the Gubinelli derivative component in the rough integral. Note that the third integral on the RHS is a Young's integral since*

$$\|[X]_t - [X]_s\| = \|[X]_{s,t}\| = \|X_{s,t} \otimes X_{s,t} - (\mathbb{X}_{s,t} + \mathbb{X}_{s,t}^T)\| \leq C|t-s|^{2\alpha}, \tag{142}$$

where the first equality has been proved in the previous context through Chen's identity. Actually, $\{[X]_t\}$ *is a* 2α*-Holder continuous path. Note that the definition of Young's integral w.r.t. the rough bracket involves the canonical isomorphism (when* $m = 1$*, it's the quadratic form).*

Example 23. *Consider the case where* X_t *is a BM sample path on* $\mathbb{R}$*, so that* $[X]_t = t$*. Rough Itô's formula applied on* $[0, t]$ *implies the following differential form*

$$\mathrm{d}F(t, X_t) = \partial_t F(t, X_t)\,\mathrm{d}t + \partial_x F(t, X_t)\,\mathrm{d}(X, \mathbb{X})_t + \frac{1}{2}\partial_{xx} F(t, X_t)\,\mathrm{d}t, \tag{143}$$

which aligns with the classical Itô's formula. Unfortunately, although rough Itô's formula is pathwise, when applied to BM sample path, it still holds under the almost sure sense, since the rough path lifting of BM exists in the almost sure sense (so rough Itô is not stronger than Itô when applying to BM sample paths). However, as we might have noticed, quadratic variation is defined in the sense of convergence in probability in the context of stochastic calculus, while the rough bracket is a pathwise deterministic object.

Rough integral is known to be continuous w.r.t. both the integrand and the integrator. To state the result, we first define a metric on the space of controlled path.

Definition. *For* $(X, \mathbb{X}), (\tilde{X}, \tilde{\mathbb{X}}) \in \Omega^\alpha$ *and* $(Y, Y') \in D_X^{2\alpha}, (\tilde{Y}, \tilde{Y}') \in D_{\tilde{X}}^{2\alpha}$*, define*

$$\|(Y, Y'); (\tilde{Y}, \tilde{Y}')\| := \|Y' - \tilde{Y}'\|_\alpha + \|R^Y - R^{\tilde{Y}}\|_{2\alpha}, \tag{144}$$

where R *is the remainder when defining the Gubinelli derivative.*

Theorem 15. *For* $(X, \mathbb{X}), (\tilde{X}, \tilde{\mathbb{X}}) \in \Omega^\alpha$ *and* $(Y, Y') \in D_X^{2\alpha}, (\tilde{Y}, \tilde{Y}') \in D_{\tilde{X}}^{2\alpha}$*, let* $(Z, Z') := (\int_0^\cdot (Y, Y')\,\mathrm{d}(X, \mathbb{X}), Y) \in D_X^{2\alpha}$ *and* $(\tilde{Z}, \tilde{Z}') := (\int_0^\cdot (\tilde{Y}, \tilde{Y}')\,\mathrm{d}(\tilde{X}, \tilde{\mathbb{X}}), \tilde{Y}) \in D_{\tilde{X}}^{2\alpha}$*, then*

$$\|(Z, Z'); (\tilde{Z}, \tilde{Z}')\| \leq C[d^\alpha(\mathbf{X}, \tilde{\mathbf{X}}) + \|Y_0' - \tilde{Y}_0'\| + T^\alpha \|(Y, Y'); (\tilde{Y}, \tilde{Y}')\|], \tag{145}$$

for some constants $C, T > 0$.

Under suitable metrics, rough integrals are close enough if the rough path liftings and the integrands are close enough. This is a property which does not hold for integration schemes in infinite-dimensional spaces (e.g., the path integral without including the signature), which once again shows the regularization effect of signatures.

Rough Differential Equation (RDE)

Since we have defined rough integrals and derived rough Itô's formula, we expect to extend the Itô's calculus for Markovian-type integrands to the rough setting. Firstly, we define the rough differential equation and clarify the notion of its solution.

Definition. *For $f \in C^2$ and $(X, \mathbb{X}) \in \Omega^\alpha$, the controlled path $(Y, Y') \in D_X^{2\alpha}$ is a solution to the RDE*

$$\mathrm{d}Y_t = f(Y_t)\,\mathrm{d}(X, \mathbb{X})_t, \quad Y_0 = y_0, \tag{146}$$

if

$$Y_t = y_0 + \int_0^t f(Y_s)\,\mathrm{d}(X, \mathbb{X})_s, \ \forall t \in [0, T]. \tag{147}$$

Note that the integral is well-defined due to Example 19.

The local well-posedness result of RDE is as follows.

Theorem 16. *Let $f \in C^3$, then there exists a unique local solution to the RDE near time 0.*

Proof. A standard invariant subspace and contraction mapping argument with Banach fixed point theorem. □

The stability of RDE solution is within our expectation.

Theorem 17. *Let $f \in C_b^3$ and $(Y, f(Y)) \in D_X^{2\alpha}$ be the unique local solution to the RDE*

$$\mathrm{d}Y_t = f(Y_t)\,\mathrm{d}(X, \mathbb{X})_t, \quad Y_0 = \xi, \tag{148}$$

let $(\tilde{Y}, f(\tilde{Y})) \in D_{\tilde{X}}^{2\alpha}$ be the unique local solution to the RDE with a perturbed rough path and initial condition

$$\mathrm{d}Y_t = f(Y_t)\,\mathrm{d}(\tilde{X}, \tilde{\mathbb{X}})_t, \quad Y_0 = \tilde{\xi}, \tag{149}$$

then

$$\|(Y, f(Y)); (\tilde{Y}, f(\tilde{Y}))\| \leq C_M(\|\xi - \tilde{\xi}\| + d^\alpha(\mathbf{X}, \tilde{\mathbf{X}})), \tag{150}$$

$$\|Y - \tilde{Y}\|_\alpha \leq C_M(\|\xi - \tilde{\xi}\| + d^\alpha(\mathbf{X}, \tilde{\mathbf{X}})), \tag{151}$$

for some constant $C_M > 0$.

Consider the Itô-Lyons map Φ that sends a rough path lifting $\mathbf{X}$ to Y, the solution to the RDE induced by $\mathbf{X}$. The theorem above proves that **the Itô-Lyons map is continuous under the rough path metric**. This once more shows the necessity of including signatures when paths are rough.

Lifting Controlled Paths to Rough Paths

One main problem with the RDE formulation is that $\mathbf{X}$ is a rough path while the solution Y is a controlled path, i.e., the solution is no longer guaranteed to be in the rough path space. For example, if $\mathbf{X}$ stands for a rough path as the source of noise and Y denotes the path of stock price (e.g., determined through a rough Black-Scholes model), one might want to naturally investigate hedging portfolios, whose dynamics involve terms like $\mathrm{d}Y_t$. Consequently, there might be problems defining such integrals w.r.t. the stock price path, which poses a problem to model financial behaviors under the rough setting. Motivated by that, we hope to construct a mapping from $D_X^{2\alpha}$ back to Ω^α while maintaining the fundamental consistency conditions.

Firstly, we define the integral of a controlled path w.r.t. another controlled path.

Theorem 18. *Let $(Y, Y'), (Z, Z') \in D_X^{2\alpha}$, where Y is a path in $\mathbb{R}^k \to \mathbb{R}^m$ as a linear mapping, and Z is a path in $\mathbb{R}^k$. Then*

$$\int (Y, Y')\,\mathrm{d}(Z, Z') := \lim_{|\Delta|\to 0} \sum_{[s,t]\in\Delta} Y_s Z_{s,t} + (Y'_s \otimes Z'_s)\mathbb{X}_{s,t} \tag{152}$$

is a well-defined integral that is additive in time and take values in $\mathbb{R}^m$. Here $Y' : \mathbb{R}^d \to (\mathbb{R}^k \to \mathbb{R}^m)$ and $Z' : \mathbb{R}^d \to \mathbb{R}^k$, so the action of $Y'_s \otimes Z'_s : \mathbb{R}^{d\times d} \to \mathbb{R}^m$ is defined through a well-defined linear extension of $v \otimes w \mapsto Y'_s(v)Z'_s(w) \in \mathbb{R}^m$ for $\forall v, w \in \mathbb{R}^d$. In other words, $Y'_s \otimes Z'_s$ is a tensor product of two linear operators.

Proof. Set $\Theta_{s,t} = Y_s Z_{s,t} + (Y'_s \otimes Z'_s)\mathbb{X}_{s,t}$ in the sewing lemma. By Chen's identity,

$$\Theta_{s,t} - \Theta_{s,u} - \Theta_{u,t} = -Y_{s,u}Z_{u,t} + (Y'_s \otimes Z'_s)(\mathbb{X}_{u,t} + X_{s,u} \otimes X_{u,t}) - (Y'_u \otimes Z'_u)\mathbb{X}_{u,t}. \tag{153}$$

Using the definition of the controlled path yields

$$\Theta_{s,t} - \Theta_{s,u} - \Theta_{u,t} = -Y'_s(X_{s,u})Z'_{s,u}(X_{u,t}) - Y'_s(X_{s,u})R^Z_{u,t} - R^Y_{s,u}Z'_u(X_{u,t}) - R^Y_{s,u}R^Z_{u,t} + (Y'_s \otimes Z'_s - Y'_u \otimes Z'_u)\mathbb{X}_{u,t}. \tag{154}$$

Since the remainders are $C^{2\alpha}$, all terms that contain the remainders are $C^{3\alpha}$. Besides, $Y'_s(X_{s,u})Z'_{s,u}(X_{u,t}) \in C^{3\alpha}$ since $Z' \in C^\alpha$, and $(Y'_s \otimes Z'_s - Y'_u \otimes Z'_u)(\mathbb{X}_{u,t}) \in C^{3\alpha}$ since $Y'_s \otimes Z'_s - Y'_u \otimes Z'_u = -Y'_s \otimes Z'_{s,u} - Y'_{s,u} \otimes Z'_u \in C^\alpha$. As a result,

$$\Theta_{s,t} - \Theta_{s,u} - \Theta_{u,t} \in C^{3\alpha}, \quad 3\alpha > 1. \tag{155}$$

We have checked that the condition of the sewing lemma holds. Using similar subsequent arguments as in Theorem 13 concludes the proof. □

Corollary 3. *Let $(Y, Y'), (Z, Z') \in D_X^{2\alpha}$, where Y is a path in $\mathbb{R}^k \to \mathbb{R}^m$ as a linear mapping and Z is a path in $\mathbb{R}^k$. Then*

$$\left(\int_0^\cdot (Y, Y')\,\mathrm{d}(Z, Z'), Y \circ Z'\right) \in D_X^{2\alpha}. \tag{156}$$

This is another Newton-Lebniz-type result for integrals of controlled paths w.r.t. controlled paths.

Proof. By the sewing lemma, we get the powerful representation

$$\int_s^t (Y,Y')_u \,\mathrm{d}(Z,Z')_u = Y_s Z_{s,t} + (Y'_s \otimes Z'_s)\mathbb{X}_{s,t} + R_{s,t}, \tag{157}$$

where $R_{s,t} \in C^{3\alpha}$. Clearly, the remainder for the Gubinelli derivative has the form

$$\int_s^t (Y,Y')_u \,\mathrm{d}(Z,Z')_u - Y_s \circ Z'_s(X_{s,t}) = Y_s R^Z_{s,t} + (Y'_s \otimes Z'_s)\mathbb{X}_{s,t} + R_{s,t} \in C^{2\alpha}, \tag{158}$$

which concludes the proof. $\square$

Remark. *Intuitively, the definition of the integral of a controlled path w.r.t. another controlled path guarantees the validity of the change of variable. Formally, $\int Y \,\mathrm{d}Z = \int Y \cdot Z' \,\mathrm{d}X$, which yields the Gubinelli derivative in the corollary above.*

With such an extension of rough integrals, we can postulate the signature values for the controlled path $(Y,Y') \in D_X^{2\alpha}$. The following theorem provides an explicit construction of the signature $\mathbb{Y}$, which guarantees that $(Y,\mathbb{Y})$ is a rough path lifting.

Theorem 19. *For any $(Y,Y') \in D_X^{2\alpha}$ as a path in $\mathbb{R}^\ell$, define the signature*

$$\mathbb{Y}_{s,t} := \int_s^t (Y_{s,u}, Y'_u) \,\mathrm{d}(Y,Y')_u \tag{159}$$

as the integral of a controlled path w.r.t. a controlled path taking values in $\mathbb{R}^{\ell\times\ell}$. Then $(Y,\mathbb{Y}) \in \Omega^\alpha$.

Proof. By the definition of rough path lifting, we have to check Chen's identity and Holder regularity. Note that in order to match dimensions, the integral is defined as

$$\int_s^t (Y_{s,u}, Y'_u) \,\mathrm{d}(Y,Y')_u = \lim_{|\Delta_{s,t}|\to 0} \sum_{[t_i,t_{i+1}]\in\Delta_{s,t}} Y_{s,t_i} \otimes Y_{t_i,t_{i+1}} + (Y'_{t_i} \otimes Y'_{t_i})\mathbb{X}_{t_i,t_{i+1}}. \tag{160}$$

For Chen's identity, check

$$\begin{aligned}
\mathbb{Y}_{s,t} &= \lim_{|\Delta_{s,t}|\to 0} \sum_{[t_i,t_{i+1}]\in\Delta_{s,t}} Y_{s,t_i} \otimes Y_{t_i,t_{i+1}} + (Y'_{t_i} \otimes Y'_{t_i})\mathbb{X}_{t_i,t_{i+1}} && (161)\\
&= \mathbb{Y}_{s,u} + \lim_{|\Delta_{s,t}|\to 0} \sum_{[t_i,t_{i+1}]\in\Delta_{s,t}, t_i\geq u} Y_{s,t_i} \otimes Y_{t_i,t_{i+1}} + (Y'_{t_i} \otimes Y'_{t_i})\mathbb{X}_{t_i,t_{i+1}} && (162)\\
&= \mathbb{Y}_{s,u} + \lim_{|\Delta_{s,t}|\to 0} \sum_{[t_i,t_{i+1}]\in\Delta_{s,t}, t_i\geq u} Y_{u,t_i} \otimes Y_{t_i,t_{i+1}} + (Y'_{t_i} \otimes Y'_{t_i})\mathbb{X}_{t_i,t_{i+1}} + Y_{s,u} \otimes Y_{t_i,t_{i+1}} && (163)\\
&= \mathbb{Y}_{s,u} + \mathbb{Y}_{u,t} + Y_{s,u} \otimes Y_{u,t}. && (164)
\end{aligned}$$

For Holder regularity, $Y \in C^\alpha$ by definition. By sewing lemma, we get the representation

$$\mathbb{Y}_{s,t} = (Y'_s \otimes Y'_s)\mathbb{X}_{s,t} + R_{s,t}, \tag{165}$$

where $R_{s,t} \in C^{3\alpha}$. As a result, $\mathbb{Y}_{s,t} \in C^{2\alpha}$, which concludes the proof. $\square$

Now that we know how to lift a controlled path back to a rough path, we focus on consistency results of this construction. The following corollary shows that this construction preserves the weakly geometric property.

Corollary 4. *For any $(X, \mathbb{X}) \in \Omega^\alpha([0,T], \mathbb{R}^d), (Y, Y') \in D_X^{2\alpha}$, $(Y, \mathbb{Y}) \in \Omega^\alpha$ takes values in $\mathbb{R}^d \to \mathbb{R}^m$ as linear mappings and is constructed as above. If $(X, \mathbb{X}) \in \mathrm{WG}(\Omega^\alpha)$, then $(Y, \mathbb{Y}) \in \mathrm{WG}(\Omega^\alpha)$.*

Proof. Check the compensated Riemann sum

$$\begin{aligned}
&Y_{s,t_i} \otimes Y_{t_i,t_{i+1}} + (Y'_{t_i} \otimes Y'_{t_i})\mathbb{X}_{t_i,t_{i+1}} + (Y_{s,t_i} \otimes Y_{t_i,t_{i+1}})^T + [(Y'_{t_i} \otimes Y'_{t_i})\mathbb{X}_{t_i,t_{i+1}}]^T && (166)\\
&= Y_{s,t_i} \otimes Y_{t_i,t_{i+1}} + (Y'_{t_i} \otimes Y'_{t_i})\mathbb{X}_{t_i,t_{i+1}} + (Y_{s,t_i} \otimes Y_{t_i,t_{i+1}})^T + (Y'_{t_i} \otimes Y'_{t_i})(\mathbb{X}^T_{t_i,t_{i+1}}) && (167)\\
&= Y_{s,t_i} \otimes Y_{t_i,t_{i+1}} + (Y_{s,t_i} \otimes Y_{t_i,t_{i+1}})^T + (Y'_{t_i} \otimes Y'_{t_i})(X_{t_i,t_{i+1}} \otimes X_{t_i,t_{i+1}}) && (168)\\
&= Y_{s,t_i} \otimes Y_{t_i,t_{i+1}} + Y_{t_i,t_{i+1}} \otimes Y_{s,t_i} + Y'_{t_i} X_{t_i,t_{i+1}} \otimes Y'_{t_i} X_{t_i,t_{i+1}}, && (169)
\end{aligned}$$

where we have been using the canonical isomorphism, the fact that $(X, \mathbb{X})$ is weakly geometric, and the flipping $(v \otimes w)^T = w \otimes v$. Note that $(Y'_{t_i} \otimes Y'_{t_i})\mathbb{X}_{t_i,t_{i+1}}$ is not a matrix product but a function evaluation! By the definition of controlled path,

$$\begin{aligned}
&Y_{s,t_i} \otimes Y_{t_i,t_{i+1}} + Y_{t_i,t_{i+1}} \otimes Y_{s,t_i} + Y'_{t_i} X_{t_i,t_{i+1}} \otimes Y'_{t_i} X_{t_i,t_{i+1}} && (170)\\
&= Y_{s,t_i} \otimes Y_{t_i,t_{i+1}} + Y_{t_i,t_{i+1}} \otimes Y_{s,t_i} + Y_{t_i,t_{i+1}} \otimes Y_{t_i,t_{i+1}} + R_{t_i,t_{i+1}} && (171)\\
&= Y_{t_{i+1}} \otimes Y_{t_{i+1}} - Y_{t_i} \otimes Y_{t_i} - Y_s \otimes Y_{t_i,t_{i+1}} - Y_{t_i,t_{i+1}} \otimes Y_s + R_{t_i,t_{i+1}} && (172)
\end{aligned}$$

where $R_{t_i,t_{i+1}} \in C^{3\alpha}$. Summing both sides w.r.t. $[t_i, t_{i+1}] \in \Delta_{s,t}$ and taking $|\Delta_{s,t}| \to 0$ yield

$$\mathbb{Y}_{s,t} + \mathbb{Y}^T_{s,t} = Y_t \otimes Y_t - Y_s \otimes Y_s - Y_s \otimes Y_{s,t} - Y_{s,t} \otimes Y_s = Y_{s,t} \otimes Y_{s,t}. \tag{173}$$

This concludes the proof. $\square$

The following corollary shows that this construction enables the calculation of rough brackets for controlled paths.

Corollary 5. *For any $(X, \mathbb{X}) \in \Omega^\alpha([0,T], \mathbb{R}^d)$, let $(Y, Y') \in D_X^{2\alpha}$ be a controlled path taking values in $\mathbb{R}^d \to \mathbb{R}^m$ as linear mappings, so that $(Z, Z') := (\int_0^\cdot (Y, Y')\, \mathrm{d}(X, \mathbb{X}), Y) \in D_X^{2\alpha}$ (cf. Corollary 1). For the rough path lifting $(Z, \mathbb{Z}) \in \Omega^\alpha([0,T]; \mathbb{R}^m)$ defined as above,*

$$[Z]_{s,t} = \int_s^t (Y_u \otimes Y_u)\, \mathrm{d}[X]_u. \tag{174}$$

Proof. We follow the same strategy and check the compensated Riemann sum. Since $Z' = Y$, and by the definition of the controlled path,

$$
\begin{aligned}
&Z_{s,t_i} \otimes Z_{t_i,t_{i+1}} + (Z'_{t_i} \otimes Z'_{t_i})\mathbb{X}_{t_i,t_{i+1}} + (Z_{s,t_i} \otimes Z_{t_i,t_{i+1}})^T + [(Z'_{t_i} \otimes Z'_{t_i})\mathbb{X}_{t_i,t_{i+1}}]^T && (175)\\
&= Z_{s,t_i} \otimes Z_{t_i,t_{i+1}} + (Z_{s,t_i} \otimes Z_{t_i,t_{i+1}})^T + (Z'_{t_i} \otimes Z'_{t_i})(X_{t_i,t_{i+1}} \otimes X_{t_i,t_{i+1}} - [X]_{t_i,t_{i+1}}) && (176)\\
&= Z_{s,t_i} \otimes Z_{t_i,t_{i+1}} + Z_{t_i,t_{i+1}} \otimes Z_{s,t_i} + Z'_{t_i} X_{t_i,t_{i+1}} \otimes Z'_{t_i} X_{t_i,t_{i+1}} - (Y_{t_i} \otimes Y_{t_i})[X]_{t_i,t_{i+1}} && (177)\\
&= Z_{s,t_i} \otimes Z_{t_i,t_{i+1}} + Z_{t_i,t_{i+1}} \otimes Z_{s,t_i} + Z_{t_i,t_{i+1}} \otimes Z_{t_i,t_{i+1}} + R_{t_i,t_{i+1}} - (Y_{t_i} \otimes Y_{t_i})[X]_{t_i,t_{i+1}} && (178)\\
&= Z_{t_{i+1}} \otimes Z_{t_{i+1}} - Z_{t_i} \otimes Z_{t_i} - Z_s \otimes Z_{t_i,t_{i+1}} - Z_{t_i,t_{i+1}} \otimes Z_s + R_{t_i,t_{i+1}} - (Y_{t_i} \otimes Y_{t_i})[X]_{t_i,t_{i+1}} && (179)
\end{aligned}
$$

where $R_{t_i,t_{i+1}} \in C^{3\alpha}$. Summing both sides w.r.t. $[t_i, t_{i+1}] \in \Delta_{s,t}$ and taking $|\Delta_{s,t}| \to 0$ yield

$$
\mathbb{Z}_{s,t} + \mathbb{Z}^T_{s,t} = Z_t \otimes Z_t - Z_s \otimes Z_s - Z_s \otimes Z_{s,t} - Z_{s,t} \otimes Z_s - \int_s^t (Y_u \otimes Y_u)\,\mathrm{d}[X]_u = Z_{s,t} \otimes Z_{s,t} - \int_s^t (Y_u \otimes Y_u)\,\mathrm{d}[X]_u. \quad (180)
$$

Here, since $Y \in C^\alpha$ and $[X] \in C^{2\alpha}$, $\int_s^t (Y_u \otimes Y_u)\,\mathrm{d}[X]_u$ is actually a Young's integral. Therefore,

$$
[Z]_{s,t} = \int_s^t (Y_u \otimes Y_u)\,\mathrm{d}[X]_u. \quad (181)
$$

□

Remark. *Recall that in the context of stochastic integration w.r.t. BM paths, this property is actually saying: if* $Z_t := \int_0^t Y_t\,\mathrm{d}W_t$ *for an admissible integrand* $\{Y_t\}$*, then*

$$
\langle Z, Z \rangle_t = \int_0^t Y_t^2\,\mathrm{d}\,\langle W, W \rangle_t\,. \quad (182)
$$

Depending on an extension of this key property alone, one may define stochastic integration in terms of Lebesgue-Stieljes integration through the Riesz representation theorem.

The following corollary shows that the rough integral with respect to the rough path lifting $(Y, \mathbb{Y})$ of the controlled paths is consistent with the rough integral directly with respect to the controlled path (Y, Y').

Corollary 6. *For any* $(X, \mathbb{X}) \in \Omega^\alpha([0,T], \mathbb{R}^d)$*, let* $(Y, Y') \in D_X^{2\alpha}$ *be a controlled path taking values in* $\mathbb{R}^d \to \mathbb{R}^m$ *as linear mappings, and* $(Z, Z') \in D_Y^{2\alpha}$ *be a controlled path taking values in* $(\mathbb{R}^d \to \mathbb{R}^m) \to \mathbb{R}^k$ *as linear mappings. It can be easily checked by definition that* $(Z, Z' \circ Y') \in D_X^{2\alpha}$*. Then, the following two integrals are always identical*

$$
\int_s^t (Z, Z')_u\,\mathrm{d}(Y, \mathbb{Y})_u = \int_s^t (Z, Z' \circ Y')_u\,\mathrm{d}(Y, Y')_u, \quad (183)
$$

where the LHS is the rough integral with respect to $(Y, \mathbb{Y}) \in \Omega^\alpha$ *and the RHS is the rough integral of a path controlled by* X *with respect to another path controlled by* X.

Proof. Consider the compensated Riemann sum within the definition of the integral on the LHS:

$$Z_{t_i} Y_{t_i,t_{i+1}} + Z'_{t_i} \mathbb{Y}_{t_i,t_{i+1}}, \tag{184}$$

Consider the compensated Riemann sum within the definition of the integral on the RHS:

$$Z_{t_i} Y_{t_i,t_{i+1}} + [(Z'_{t_i} \circ Y'_{t_i}) \otimes Y'_{t_i}] \mathbb{X}_{t_i,t_{i+1}}. \tag{185}$$

Consider their difference and use the powerful representation of $\mathbb{Y}_{t_i,t_{i+1}}$ provided by the sewing lemma to get

$$Z_{t_i} Y_{t_i,t_{i+1}} + Z'_{t_i} \mathbb{Y}_{t_i,t_{i+1}} - Z_{t_i} Y_{t_i,t_{i+1}} - [(Z'_{t_i} \circ Y'_{t_i}) \otimes Y'_{t_i}] \mathbb{X}_{t_i,t_{i+1}} \tag{186}$$

$$= Z'_{t_i} \mathbb{Y}_{t_i,t_{i+1}} - [(Z'_{t_i} \circ Y'_{t_i}) \otimes Y'_{t_i}] \mathbb{X}_{t_i,t_{i+1}} \tag{187}$$

$$= Z'_{t_i} [Y_{t_i,t_i} \otimes Y_{t_i,t_{i+1}} + (Y'_{t_i} \otimes Y'_{t_i}) \mathbb{X}_{t_i,t_{i+1}} + R_{t_i,t_{i+1}}] - [(Z'_{t_i} \circ Y'_{t_i}) \otimes Y'_{t_i}] \mathbb{X}_{t_i,t_{i+1}} \tag{188}$$

$$= Z'_{t_i} R_{t_i,t_{i+1}} \in C^{3\alpha}, \tag{189}$$

where $R_{t_i,t_{i+1}} \in C^{3\alpha}$. Summing both sides w.r.t. $[t_i, t_{i+1}] \in \Delta_{s,t}$ and taking $|\Delta_{s,t}| \to 0$ conclude the proof. □

The following corollary shows the desired associativity of rough integral.

Corollary 7. *For any $(X, \mathbb{X}) \in \Omega^\alpha([0,T], \mathbb{R}^d)$, let $(K, K') \in D_X^{2\alpha}$ be a controlled path taking values in $\mathbb{R}^d \to \mathbb{R}^k$ as linear mappings, so that $(Z, Z') := (\int_0^\cdot (K, K')\, \mathrm{d}(X, \mathbb{X}), K) \in D_X^{2\alpha}$ (cf. Corollary 1). Let $(Y, Y') \in D_X^{2\alpha}$ be another controlled path taking values in $\mathbb{R}^k \to \mathbb{R}^m$ as linear mappings, so that $(Y \circ K, Y'K + Y \circ K') \in D_X^{2\alpha}$ (cf. Example 20). Then,*

$$\int_0^t (Y, Y')_u \, \mathrm{d}(Z, Z')_u = \int_0^t (Y \circ K, Y'K + Y \circ K')_u \, \mathrm{d}(X, \mathbb{X})_u, \tag{190}$$

where both sides take values in $\mathbb{R}^m$.

Proof. Consider the compensated Riemann sum within the definition of the integral on the LHS:

$$Y_{t_i} Z_{t_i,t_{i+1}} + (Y'_{t_i} \otimes Z'_{t_i}) \mathbb{X}_{t_i,t_{i+1}}, \tag{191}$$

Consider the compensated Riemann sum within the definition of the integral on the RHS:

$$(Y \circ K)_{t_i} X_{t_i,t_{i+1}} + (Y'K + Y \circ K')_{t_i} \mathbb{X}_{t_i,t_{i+1}}. \tag{192}$$

Consider their difference and use the powerful representation of $Z_{t_i,t_{i+1}}$ provided by the sewing lemma to get

$$Y_{t_i} Z_{t_i,t_{i+1}} + (Y'_{t_i} \otimes Z'_{t_i}) \mathbb{X}_{t_i,t_{i+1}} - (Y \circ K)_{t_i} X_{t_i,t_{i+1}} - (Y'K + Y \circ K')_{t_i} \mathbb{X}_{t_i,t_{i+1}} \tag{193}$$

$$= Y_{t_i} (K_{t_i} X_{t_i,t_{i+1}} + K'_{t_i} \mathbb{X}_{t_i,t_{i+1}} + R_{t_i,t_{i+1}}) - (Y \circ K)_{t_i} X_{t_i,t_{i+1}} - (Y \circ K')_{t_i} \mathbb{X}_{t_i,t_{i+1}} \tag{194}$$

$$= Y_{t_i} R_{t_i,t_{i+1}} \in C^{3\alpha} \tag{195}$$

where $R_{t_i,t_{i+1}} \in C^{3\alpha}$. Summing both sides w.r.t. $[t_i, t_{i+1}] \in \Delta_{0,t}$ and taking $|\Delta_{0,t}| \to 0$ conclude the proof. □

Black-Scholes Model and Derivative Pricing in Rough Environments

Lifting controlled paths back to rough paths is exactly motivated by the pricing problem in rough environments. Here, we consider a rough Black-Scholes (RBS) model for the stock price $\{S_t\}$, where the noise comes from a rough path $(X, \mathbb{X}) \in \Omega^\alpha$:

$$\mathrm{d}S_t = \mu S_t \,\mathrm{d}t + \sigma S_t \,\mathrm{d}(X, \mathbb{X})_t, \;\; S_0 \text{ given.} \tag{196}$$

Assume that the riskless asset (bond) has interest rate r, i.e., the price is $M_t = M_0 e^{rt}$. Firstly, we analytically solve the stock price under the RBS model. If the solution $\{S_t\}$ exists, it must be a path controlled by X. Applying rough Itô's formula for $\log S_t$ yields

$$\mathrm{d}\log S_t = \frac{1}{S_t}\,\mathrm{d}S_t - \frac{1}{2S_t^2}\,\mathrm{d}[S]_t, \tag{197}$$

where $[S]_t$ denotes the rough bracket. By Corollary 5,

$$\mathrm{d}[S]_t = \sigma^2 S_t^2 \,\mathrm{d}[X]_t. \tag{198}$$

Combining both yields

$$\mathrm{d}\log S_t = \mu \,\mathrm{d}t + \sigma \,\mathrm{d}(X, \mathbb{X})_t - \frac{\sigma^2}{2}\,\mathrm{d}[X]_t, \tag{199}$$

which provides the solution

$$S_t = S_0 e^{\mu t - \frac{\sigma^2}{2}[X]_t + \sigma X_t}. \tag{200}$$

Consider a hedging portfolio consisting of γ_t bonds and ξ_t stocks at time t. The stock price S_t can be lifted back to a rough path $(S, \mathbb{S}) \in \Omega^\alpha$, with respect to which rough integration can be carried out. It is required that (γ_t, ξ_t) is a path controlled by the rough path (M_t, S_t), for the further rough integrals to be well-defined. As a result, (γ_t, ξ_t) is a path controlled by (t, X_t). Denote by V_t the value of the hedging portfolio at time t, i.e.,

$$V_t := \gamma_t M_t + \xi_t S_t. \tag{201}$$

The self-financing condition requires that

$$\begin{aligned} \mathrm{d}V_t &= \gamma_t \,\mathrm{d}M_t + \xi_t \,\mathrm{d}(S, \mathbb{S})_t \qquad (202)\\ &= \gamma_t r e^{rt}\,\mathrm{d}t + \xi_t \mu S_t \,\mathrm{d}t + \xi_t \sigma S_t \,\mathrm{d}(X, \mathbb{X})_t, \qquad (203) \end{aligned}$$

where we have applied the associativity (cf. Corollary 7). Plug back $\gamma_t = e^{-rt}(V_t - \xi_t S_t)$ to get

$$\mathrm{d}V_t = rV_t \,\mathrm{d}t + (\mu - r)\xi_t S_t \,\mathrm{d}t + \xi_t \sigma S_t \,\mathrm{d}(X, \mathbb{X})_t. \tag{204}$$

Following the idea of risk-neutral pricing, the next step is to identify a rough path $(Y, \mathbb{Y}) \in \Omega^\alpha$ such that $\sigma \xi_t S_t \,\mathrm{d}Y_t = (\mu - r)\xi_t S_t \,\mathrm{d}t + \xi_t \sigma S_t \,\mathrm{d}(X, \mathbb{X})_t$. Therefore, we define

$$Y_t := \frac{\mu - r}{\sigma} t + X_t. \tag{205}$$

Now we lift the controlled path Y back to a rough path. Followed by the definition of $\mathbb{Y}_{s,t}$,

$$\mathbb{Y}_{s,t} = \int_s^t (Y_{s,u}, Y'_u)\,\mathrm{d}(Y, Y')_u. \tag{206}$$

Since $Y' \equiv 1$, we calculate this integral by checking the compensated Riemann sum

$$\begin{aligned}
& Y_{s,t_i} Y_{t_i,t_{i+1}} + Y'_{t_i} Y'_{t_i} \mathbb{X}_{t_i,t_{i+1}} && (207)\\
&= \left[\frac{\mu - r}{\sigma}(t_i - s) + X_{s,t_i}\right]\left[\frac{\mu - r}{\sigma}(t_{i+1} - t_i) + X_{t_i,t_{i+1}}\right] + \mathbb{X}_{t_i,t_{i+1}} && (208)\\
&= \left(\frac{\mu - r}{\sigma}\right)^2 (t_i - s)(t_{i+1} - t_i) + \frac{\mu - r}{\sigma}(t_i - s)X_{t_i,t_{i+1}} + \frac{\mu - r}{\sigma} X_{s,t_i}(t_{i+1} - t_i) + X_{s,t_i} X_{t_i,t_{i+1}} + \mathbb{X}_{t_i,t_{i+1}} && (209)\\
&= \left(\frac{\mu - r}{\sigma}\right)^2 (t_i - s)(t_{i+1} - t_i) + \frac{\mu - r}{\sigma}(t_i - s)X_{t_i,t_{i+1}} + \frac{\mu - r}{\sigma} X_{s,t_i}(t_{i+1} - t_i) + (\mathbb{X}_{s,t_{i+1}} - \mathbb{X}_{s,t_i}), && (210)
\end{aligned}$$

where we apply Chen's identity. Summing both sides w.r.t. $[t_i, t_{i+1}] \in \Delta_{s,t}$ and taking $|\Delta_{s,t}| \to 0$ yield

$$\begin{aligned}
\mathbb{Y}_{s,t} &= \mathbb{X}_{s,t} + \left(\frac{\mu - r}{\sigma}\right)^2 \int_s^t (u - s)\,\mathrm{d}u + \frac{\mu - r}{\sigma}\int_s^t (u - s)\,\mathrm{d}(X, \mathbb{X})_u + \frac{\mu - r}{\sigma}\int_s^t X_{s,u}\,\mathrm{d}u && (211)\\
&= \mathbb{X}_{s,t} + \left(\frac{\mu - r}{\sigma}\right)^2 \frac{(t - s)^2}{2} + \frac{\mu - r}{\sigma}\int_s^t (u - s)\,\mathrm{d}(X, \mathbb{X})_u + \frac{\mu - r}{\sigma}\int_s^t X_{s,u}\,\mathrm{d}u. && (212)
\end{aligned}$$

When the rough path $(X, \mathbb{X})$ is random and follows measure $\mathbb{P}$ on the path space, one may find a measure $\mathbb{Q}$ on the path space such that the law of Y under $\mathbb{Q}$ matches with the law of X under $\mathbb{P}$. This change of measure is explicitly provided by the Cameron-Martin formula when the measures are Gaussian, but is otherwise not guaranteed to exist. The introduction of rough integral enables us to discuss finance in rough environments, and proposes a construction of the risk-neutral measure $\mathbb{Q}$ for Gaussian rough paths (e.g., fBM).

Remark. *Note that (γ_t, ξ_t) is required to be a controlled path for the rough integral to be well-defined, while most controlled-type portfolios are actually Markovian, which only form a small class of admissible hedging strategies. If one hopes to enlarge the set of admissible hedging strategies, e.g., allowing one jump, then the only Gaussian rough path $(X, \mathbb{X})$ such that $\mathbb{Q}$ defined above remains risk-neutral is a deterministic time-change of BM.*

In other words, the setting above is a nice financial framework discussing arbitrage-free pricing that restricts to Markovian strategies. However, it is hard to be extended to a general framework of arbitrage-free pricing, under which adapted strategies are allowed.

Applications of Signatures

Signature Conditional Wasserstein Generative Adversarial Network (SigCWGAN)

We assume that the readers have relevant background in: (i). The problem of generative modeling (ii). The design and mechanism of GAN (iii). How the introduction of Wasserstein distance stabilizes the training of GAN.

Consider a given discrete-time stochastic process $\{X_t\}$ on the time horizon $\{0, 1, ..., T\}$ in $\mathbb{R}^d$, where $X_0 = 0$. For a fixed time t, denote by $X_{\text{past},t} := (X_{t-\overline{p}+1}, ..., X_t)$ the $\overline{p}$-step past and $X_{\text{future},t} := (X_{t+1}, ..., X_{t+\overline{q}})$ the $\overline{q}$-step future. The problem of conditional generative modeling hopes to learn the law of $X_{\text{future},t}|_{X_{\text{past},t}}$. Note that, if the dynamics of $\{X_t\}$ only admits a $\overline{p}$-step path dependence into the past, learning $X_{t+1}|_{X_{\text{past},t}}$ (i.e., setting $\overline{q} = 1$) is sufficient for recovering the law of the whole process $\{X_t\}$.

A natural method for this conditional generative modeling problem is called CWGAN. Adopting the idea of GAN, one transfers randomness instead of creating new randomness, i.e., putting up a parameterized generator

$$G : \Theta^{(g)} \times \mathcal{X} \times \mathcal{Z} \to \mathcal{Y}, \tag{213}$$

where $\Theta^{(g)}$ denotes the parameter space, $\mathcal{X} := \mathbb{R}^{d\times\overline{p}}$ denotes the space where the past paths live, $\mathcal{Y} := \mathbb{R}^{d\times\overline{q}}$ denotes the space where future paths live, and $\mathcal{Z} := \mathbb{R}^{d_z}$ denotes the latent space where the standard Gaussian random vector (the source of randomness) takes values. Denote μ_Z as the standard Gaussian measure on $\mathcal{Z}$. Tracking the source of randomness, a given set of parameters $\theta^{(g)} \in \Theta^{(g)}$ and a given past path $x_{\text{past},t} \in \mathcal{X}$, $G(\theta^{(g)}, x_{\text{past},t}) : \mathcal{Z} \to \mathcal{Y}$ induce a pushforward

$$\nu(\theta^{(g)}, x_{\text{past},t}) := [G(\theta^{(g)}, x_{\text{past},t})]_{\#}\mu_Z \in \mathscr{P}(\mathcal{Y}). \tag{214}$$

Since the output of the generator would be the artificially generated future path, one hopes to find the optimal $\theta^{(g)}$ such that $\nu(\theta^{(g)}, x_{\text{past},t})$ is as indistinguishable as possible from the following measure:

$$\mu(x_{\text{past},t}) := \text{Law}(X_{\text{future},t}|_{X_{\text{past},t}=x_{\text{past},t}}). \tag{215}$$

As pointed out by the philosophy of WGAN, one shall naturally select the Wasserstein-1 distance as a discrepancy measure between two probability distributions, which leads to the Kantorovich-Rubinstein (KR) representation:

$$W_1(\mu(x_{\text{past},t}), \nu(\theta^{(g)}, x_{\text{past},t})) = \sup_{\|f\|_L\leq 1} \mathbb{E}_{\mu(x_{\text{past},t})} f(X_{\text{future},t}) - \mathbb{E}_{\nu(\theta^{(g)},x_{\text{past},t})} f(X_{\text{future},t}), \tag{216}$$

where the measure under the expectation by default indicates the law of $X_{\text{future},t}$. Note that this W_1 distance is a function of $x_{\text{past},t}$, and taking expectation w.r.t. $X_{\text{past},t}$ provides the loss of CWGAN. Therefore, the CWGAN algorithm solves the following optimization problem:

$$\inf_{\theta^{(g)}} \mathbb{E}_{X_{\text{past},t}\sim\eta} W_1(\mu(X_{\text{past},t}), \nu(\theta^{(g)}, X_{\text{past},t})). \tag{217}$$

Within the KR representation, f is interpreted as the discriminator, which is typically maintained as a neural network with parameters $\theta^{(d)}$, i.e., $f : \Theta^{(d)} \times \mathcal{Y} \to \mathbb{R}$. To impose the condition $\|f\|_L \leq 1$, one needs to carry out parameter clippings on $\theta^{(d)}$, i.e., restricting the values $\theta^{(d)}$ can take to the parameter space $\Theta^{(d)}$. This turns the optimization problem of CWGAN into a min-max problem w.r.t. $\theta^{(g)} \in \Theta^{(g)}$ and $\theta^{(d)} \in \Theta^{(d)}$. Due to practical infeasibility of solving a maximizer $\theta^{(d),*}$ for each given $\theta^{(g)}$, this optimization is always not exactly solved, but is instead carried out in an alternating way among the generator and the discriminator under a certain relative frequency of updates. This poses a huge challenge for the training stability/convergence. In most cases, it turns out that the training is sensitive to the selection of hyperparameters (adjusting the learning speed of the generator vs. the discriminator). For the completeness of the discussion, we present the CWGAN optimization problem with a generator and a discriminator:

$$\inf_{\theta^{(g)} \in \Theta^{(g)}} \sup_{\theta^{(d)} \in \Theta^{(d)}} \mathbb{E}_{X_{\text{past},t} \sim \eta}[\mathbb{E}_{\mu(X_{\text{past},t})} f(\theta^{(d)}, X_{\text{future},t}) - \mathbb{E}_{\nu(\theta^{(g)}, X_{\text{past},t})} f(\theta^{(d)}, X_{\text{future},t})]. \tag{218}$$

The introduction of signatures into CWGAN serves as a feature extraction method for paths. In addition, due to the universal approximation property of signature, one may restrict the discriminator to the linear functional class, which makes it possible to exactly solve out the optimal discriminator for each set of generator parameters. To begin with, we denote μ, ν as two compactly (w.r.t. the rough path topology) supported probability measures on the path space Ω_0. The path space Ω_0 contains time-augmented paths $\{(t, X_t)\}$ with $X_0 = 0$. Clearly, the time augmentation eliminates the time reparameterization invariance and tree-like structures, whereas the fixed initial point eliminates translation invariance. As a result, on this specific path space, the signature of a path $S(Y)$, $Y \in \Omega_0$ uniquely identifies the path itself. For simplicity of notations, we still denote the time-augmented paths within Ω_0 as X.

The key quantity of our interest is the signature Wasserstein-1 metric defined as follows:

$$\text{Sig-}W_1(\mu, \nu) := \sup_{\|\ell\|_L \leq 1, \ell \text{ is linear}} \mathbb{E}_{X \sim \mu} \ell(S(X)) - \mathbb{E}_{Y \sim \nu} \ell(S(Y)). \tag{219}$$

Remark. *The natural choice to measure the discrepancy between two path measures μ and ν is*

$$W_1(\mu, \nu) = \sup_{\|f\|_L \leq 1} \mathbb{E}_{X \sim \mu} f(X) - \mathbb{E}_{Y \sim \nu} f(Y). \tag{220}$$

By the universal approximation property of signature, for any function $\|f\|_L \leq 1$, any compact $K \subset\subset \Omega_0$ and any $\varepsilon > 0$, there exists a linear ℓ such that $\sup_{X \in K} |\ell(S(X)) - f(X)| < \varepsilon$, which directly motivates the Sig-W_1 *metric. However, it is only guaranteed that $|\ell(S(X)) - \ell(S(Y))| \leq \|S(X) - S(Y)\|$, so $\ell \circ S$ is even not necessarily Lipschitz (see Example 14). Consequently, the* Sig-W_1 *metric is heuristically motivated by the universal approximation and the KR representation, and shall not be expected to be identical to $W_1(\mu, \nu)$.*

Although numerically imposing constraints on the Lipchitz constant is hard, thanks to the linearity of ℓ, its Lipschitz constant can be calculated analytically. We assume that the signature $S(X)$ has a finite ℓ_p norm for some $p > 1$. A simple application of Holder's inequality yields

$$\|\ell\|_L := \sup_{x \neq y} \frac{|\ell(x - y)|}{\|x - y\|_p} = \|\ell\|_q, \tag{221}$$

where q is the Holder conjugate of p, i.e., $\frac{1}{p} + \frac{1}{q} = 1$. This implies

$$\text{Sig-}W_1(\mu, \nu) = \sup_{\|\ell\|_q \le 1, \ell \text{ is linear}} \ell(\mathbb{E}_{X\sim\mu} S(X) - \mathbb{E}_{Y\sim\nu} S(Y)) = \|\mathbb{E}_{X\sim\mu} S(X) - \mathbb{E}_{Y\sim\nu} S(Y)\|_p, \tag{222}$$

which is nothing but the ℓ_p norm of the difference in the expected signature.

Remark. *By taking advantage of the linearity brought by signatures, it is no longer necessary to maintain a discriminator network, which reduces the adversarial training scheme back to the normal training scheme.*

In the following context, we assume $p = 2$, truncate the signatures of the past paths up to degree M_1, and truncate the signatures of the future paths up to degree M_2. This provides the following optimization problem of Sig-CWGAN:

$$\inf_{\theta^{(g)}} \mathbb{E}_{X_{\text{past},t}\sim\eta} \|\mathbb{E}_{\mu(X_{\text{past},t})} S_{M_2}(X_{\text{future},t}) - \mathbb{E}_{\nu(\theta^{(g)}, X_{\text{past},t})} S_{M_2}(X_{\text{future},t})\|_2. \tag{223}$$

The last problem to address is the numerical computation of the conditional expectation $\mathbb{E}_{\mu(X_{\text{past},t})} S_{M_2}(X_{\text{future},t})$. Due to a limited amount of training data, this conditional expectation is hard to be well approximated directly from the dataset (this problem does not exist for CWGAN, think about why). One possible remedy is to use the characterization of conditional expectation as the best approximator under MSE, i.e.

$$\mathbb{E}_{\mu(X_{\text{past},t})} S_{M_2}(X_{\text{future},t}) = h^*(X_{\text{past},t}), \;\; h^* = \arg\inf_h \mathbb{E}_{X_{\text{past},t}\sim\eta, X_{\text{future},t}\sim\mu(X_{\text{past},t})} \|S_{M_2}(X_{\text{future},t}) - h(X_{\text{past},t})\|_2^2. \tag{224}$$

However, this requires the training of another neural network and causes additional bias, which is not favorable. The wisdom is again to apply the universal approximation theorem. On assuming that $\mathbb{E}_{\mu(x_{\text{past},t})} S_{M_2}(X_{\text{future},t})$ is a continuous function in $x_{\text{past},t}$, it can be arbitrarily well approximated by a linear function of $S(x_{\text{past},t})$, which results in the calculation of the conditional expectation using linear regression:

$$h^* \approx \arg \inf_{\ell \text{ linear}} \mathbb{E}_{X_{\text{past},t}\sim\eta, X_{\text{future},t}\sim\mu(X_{\text{past},t})} \|S_{M_2}(X_{\text{future},t}) - \ell(S_{M_1}(X_{\text{past},t}))\|_2^2. \tag{225}$$

With the expectation approximated by the empirical average w.r.t. the training data, this minimization problem is solving for an optimal linear mapping ℓ that minimizes the MSE. Hence, it is sufficient to solve the linear regression problem under the model:

$$S_{M_2}(X_{\text{future},t}) = \ell(S_{M_1}(X_{\text{past},t})) + \varepsilon, \tag{226}$$

where ε is an exogenous random error (we omit standard assumptions for OLS).

Remark. *Once more due to linearity, such h^* can be solved once and for all, prior to the training of Sig-CWGAN as a pre-processing step.*

This concludes the Sig-CWGAN training procedure, with the pseudocode provided in Algorithm 1. As emphasized in the previous context, this method does not maintain a discriminator network, which greatly simplifies the training and fine-tuning.

Last but not least, we briefly comment on the numerical performance of Sig-CWGAN. As a generative model, we evaluate its performance according to: (i). The error in the flow of marginal distributions of the process generated

Algorithm 1 Signature-based Conditional Wasserstein Generative Adversarial Network (Sig-CWGAN)

Input: A generator network G with parameter $\theta^{(g)}$
1: Initialize $\theta^{(g)}$
2: Compute truncated signatures $S_{M_1}(X_{\text{past},t})$, $S_{M_2}(X_{\text{future},t})$ for all training samples
3: Compute h^* as an approximation of $x_{\text{past},t} \mapsto \mathbb{E}_{\mu(x_{\text{past},t})} S_{M_2}(X_{\text{future},t})$ through linear regression (226)
4: **repeat**
5: Sample a batch of training data
6: Approximate $\mathbb{E}_{\nu(\theta^{(g)},X_{\text{past},t})} S_{M_2}(X_{\text{future},t})$ through Monte Carlo and the forward propagation of the generator network on the batch
7: Calculate the loss $\mathbb{E}_{X_{\text{past},t}\sim\eta} \|h^*(X_{\text{past},t}) - \mathbb{E}_{\nu(\theta^{(g)},X_{\text{past},t})} S_{M_2}(X_{\text{future},t})\|_2$ through Monte Carlo on the batch
8: Update parameters $\theta^{(g)}$
9: **until** Enough training epochs are carried out

Output: A trained generator network for conditional generative modeling

(mimicking quality) (ii). The error in the auto-correlation function (temporal dependency) (iii). The error in the correlation across different features (feature dependency) (iv). The difference in the prediction capability of machine learning models that are trained based on real and synthetic data respectively (usefulness of synthetic data).

Numerical experiments for the VAR(1) (one-step vector autoregressive) model show that, Sig-CWGAN does much better in all aspects than other SOTA GAN methods. On real data from stock markets (SPX and DJI time series), Sig-CWGAN does the best in capturing temporal and feature dependencies, but not in other aspects. On real data for Bitcoin-USD, Sig-CWGAN does better in mimicking the flow of marginal distributions.

Remark. *Sig-CWGAN is a useful extension of GAN to generate time-series data. However, it does not address the issue of financial data scarcity. For real stock price data that only has a single observation of the trajectory, one has to assume that the time series is strongly stationary (distribution remains the same for all t), in order to establish multiple windows of lengths $\bar{p} + \bar{q}$ to construct the training data for Sig-CWGAN.*

In my opinion, novel generative modeling needs to be developed under prior knowledge on the supporting manifold of the measure (depending on the type of data generated, images? price?). Can Sig-CWGAN be generalized in this direction (e.g., generating videos as a time series of images?) What about combining it with diffusion map embeddings (capture intrinsic manifold structure of the data)? Currently, Sora (developed by OpenAI) generates video that does not obey physics laws, would using signatures mitigate that problem? Can we further impose PDE restrictions that characterizes the physical laws in the real world?