



The inverse problem for rough controlled differential equations

Ismaël Bailleul, J. Diehl

► To cite this version:

Ismaël Bailleul, J. Diehl. The inverse problem for rough controlled differential equations. SIAM Journal on Control and Optimization, 2015, 53 (5), pp.2762-2780. 10.1137/140995982 . hal-01225321

HAL Id: hal-01225321

<https://hal.science/hal-01225321>

Submitted on 10 Dec 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THE INVERSE PROBLEM FOR ROUGH CONTROLLED DIFFERENTIAL EQUATIONS*

I. BAILLEUL[†] AND J. DIEHL[‡]

Abstract. We provide a necessary and sufficient condition for a rough control driving a differential equation to be reconstructable, to some order, from observing the resulting controlled evolution. Physical examples and applications in stochastic filtering and statistics demonstrate the practical relevance of our result.

Key words. rough paths theory, rough differential equations, signal reconstruction

AMS subject classifications. 34A55, 93C15

DOI. 10.1137/140995982

1. Introduction. It is a classical topic in control theory to consider differential equations of the form

$$\dot{x}_t = \sum_{i=1}^{\ell} V_i(x_t) u_t^i,$$

with nice enough vector fields $V_i : \mathbb{R}^d \rightarrow \mathbb{R}^d$, say bounded and Lipschitz for the moment, and the controls $u^i : [0, T] \rightarrow \mathbb{R}$, $i = 1, \dots, \ell$, are integrable functions that represent data that can usually be tuned on demand in a real-life problem. It will pay off to consider the controls u^i as the time derivative of absolutely continuous functions X^i , and to reformulate the above dynamics in the form

$$(1) \quad dx_t = V_i(x_t) dX_t^i,$$

with the usual convention for sums. In this work we consider the *inverse problem*, of recovering the control X using only the knowledge of the dynamics generated by the above equation, and assuming that we know the vector fields V_i .

With a view to applications involving noisy signals such as Brownian paths, we are particularly interested in the setting where X is *not* absolutely continuous, but only α -Hölder continuous, for some $\alpha \in (\frac{1}{3}, \frac{1}{2}]$, say. Although (1) seems nonsensical in this case, it is given meaning by the theory of rough paths, as invented by T. Lyons [21], and reformulated and enriched by numerous works since its introduction. We recall the essentials of this theory in section 2 so as to make it accessible to a large audience; it suffices for the moment to emphasize that driving signals in this theory do not consist only of $\mathbb{R}^\ell$ -valued α -Hölder continuous paths X ; rather they are pairs $(X, \mathbb{X})$, where the $\mathbb{R}^\ell \times \mathbb{R}^\ell$ -valued second object $\mathbb{X}$ is to be thought of as the collection of “iterated integrals $\int X^j dX^k$.” The point here is that the latter expression is a priori meaningless if we only know that X is α -Hölder, for $\alpha \leq \frac{1}{2}$, as one cannot

*Received by the editors November 14, 2014; accepted for publication (in revised form) July 2, 2015; published electronically September 2, 2015. This research was partially supported by the program ANR “Retour Post-Doctorants,” under contract ANR 11-PDOC-0025, and by DFG Schwerpunktprogramm 1324.

<http://www.siam.org/journals/sicon/53-5/99598.html>

[†]IRMAR, 263 Avenue du General Leclerc, 35042 Rennes, France (ismael.bailleul@univ-rennes1.fr).

[‡]Institut für Mathematik, MA 7-2, Strasse des 17. Juni 136 10623 Berlin, Germany (diehl@math.tu-berlin.de).

make sense of such integrals without additional structure. So these iterated integrals have to be provided as *additional a priori data*. The enriched control $(X, \mathbb{X})$ is what we call a *rough path*. Lyons' main breakthrough in his fundamental work [21] was to show that (1) can be given sense, and is well-posed, whenever X is understood as a rough path, provided the vector fields V_i are sufficiently regular. Moreover the solution map, which associates to the starting point x_0 and the rough path $(X, \mathbb{X})$ the solution $x_\bullet$, is continuous. This is in stark contrast with the fact that solutions of stochastic differential equations driven by Brownian motion are only measurable functions of their drivers, with no better dependence as a rule. Section 2 offers a short review of rough paths theory sufficient to grasp the core ideas of the theory and for our needs in this work.

From a practical point of view, it may be argued that there are no physical examples of real-life observable paths which are truly α -Hölder continuous paths; rather, all paths can be seen as smooth, albeit with possibly very high fluctuations that make them appear rough on a macroscopic scale. However, the invisible microscopic fluctuations might lead to macroscopic effects when the path (the control) is acting on a dynamical system. A rough path somehow provides a mathematical abstraction of this fact by recording this microscopic scale effects that a highly oscillatory smooth signal may have on a dynamics system in the second order object $\mathbb{X}$.

While rough paths theory could originally be thought of as a theoretical framework for the study of controlled systems, its core notions and tools have proved extremely useful in handling a number of practical important problems. As an example, Lyons and Victoir's cubature method on Wiener space [24] has led to efficient numerical schemes for the simulation of various partial differential equations, such as the HJM or CIR equations, and related quantities of interest in mathematical finance. Let us mention, as another example, that the use of the core concept of signature of a signal in the setting of learning theory is presently being investigated [20, 17], and may well bring deep insights into this subject. In a different direction, in [10], it was shown that the optimal filter in stochastic filtering, which is in general *not* a continuous function on path space, can actually be defined as a continuous functional on the rough path space; see section 6 for details. This result motivated the present work, since it leaves the practitioner with the task of "observing a rough path," if she/he wants to use this continuity result to provide robust approximations for the optimal filter, by feeding her/his approximate rough path into the continuous function that gives back the optimal filter. It should be clear from the above heuristic picture of a rough path that the only way to uncover the "microscopic" second level component of a rough path is to let that rough path act on a dynamical system, a rough differential equation.

After introducing the reader to the essentials of rough paths theory in section 2, in section 3 we give our main result: a necessary and sufficient condition for the reconstruction of a rough path to be possible from the observation of the dynamics generated by a rough differential equation with that rough path as driver. Its proof is given in section 4. Section 5 provides several examples of physical systems that satisfy our assumptions, and section 6 offers applications of our result to problems in filtering and statistics.

2. Rough controlled differential equations. As a first step in rough paths theory, let us consider the controlled ordinary differential equation

$$dx_t = V_i(x_t)dX_t^i$$

driven by a smooth (or bounded variation) $\mathbb{R}^\ell$ -valued control X . It is elementary to iterate the formula

$$V_i(x_t) = V_i(x_s) + \int_s^t \partial_j V_i(x_r) V_k^j(X_r) dX_r^k =: V_i(x_s) + \int_s^t V_k V_i(x_r) dX_r^k,$$

to obtain a kind of Taylor–Euler expansion of the solution path to the above equation, in the form

$$\begin{aligned} x_t = x_s + & \sum_{n=1, \dots, N-1} \sum_{i_1, \dots, i_n=1, \dots, \ell} \left(\int_s^t \int_s^{r_n} \dots \int_s^{r_2} dX_{r_1}^{i_1} \dots dX_{r_n}^{i_n} \right) V_{i_1} \dots V_{i_n}(x_s) \\ & + \sum_{i_1, \dots, i_N=1, \dots, \ell} \int_s^t \int_s^{r_N} \dots \int_s^{r_2} V_{i_1} \dots V_{i_N}(x_{r_1}) dX_{r_1}^{i_1} \dots dX_{r_N}^{i_N}, \end{aligned}$$

provided the vector fields are sufficiently regular for this expression to make sense. If V and its first N derivatives are bounded, the last term is of order $|t - s|^N$. So the numerical scheme

$$(2) \quad \begin{cases} x_0^n := x_0, \\ x_{t_i}^n := x_{t_{i-1}}^n + \sum_{n=1, \dots, N-1} \sum_{i_1, \dots, i_n=1, \dots, \ell} V_{i_1} \dots V_{i_n}(x_{t_{i-1}}^n) \int_s^t \int_s^{r_n} \dots \int_s^{r_2} dX_{r_1}^{i_1} \dots dX_{r_n}^{i_n}, \end{cases}$$

written here for any partition $\tau = (t_i)$ of $[0, T]$, is of order $|\tau|^{N-1}$, where $|\tau|$ denotes the meshsize of the partition. See, for example, Proposition 10.3 in [15] for more details.

In the mid-1990s Lyons [21] understood that one can actually make sense of the controlled differential equation (1), even if X is not of bounded variation, *if one provides a priori the values of sufficiently many iterated integrals* $\int \dots \int dX \dots dX$, and one defines a solution path to the equation as a path for which the above scheme is exact up to a term of size $|t_{i+1} - t_i|^a$ for some constant $a > 1$.

DEFINITION 1. Fix a finite time horizon T , and let $\alpha \in (\frac{1}{3}, \frac{1}{2}]$. An α -Hölder rough path $\mathbf{X} = (X, \mathbb{X})$ consists of a pair $(X, \mathbb{X})$, made up of an α -Hölder continuous path $X : [0, T] \rightarrow \mathbb{R}^\ell$, whose increments are denoted by $X_{ts} := X_t - X_s$, and a two parameter function $\mathbb{X} : \{(t, s); 0 \leq s \leq t \leq T\} \rightarrow (\mathbb{R}^\ell)^{\otimes 2} \simeq \mathbb{R}^{\ell \times \ell}$, such that the inequality

$$|\mathbb{X}_{ts}| \leq C|t - s|^{2\alpha}$$

holds for some positive constant C , and all $0 \leq s \leq t \leq T$, and we have

$$(3) \quad \mathbb{X}_{ts}^{jk} - \mathbb{X}_{tu}^{jk} - \mathbb{X}_{us}^{jk} = X_{us}^j X_{tu}^k$$

for all $0 \leq s \leq u \leq t \leq T$. The rough path $\mathbf{X}$ is said to be weakly geometric if the symmetric part of $\mathbb{X}_{ts}$ is given in terms of X_{ts} by the relation

$$\mathbb{X}_{ts}^{jk} + \mathbb{X}_{ts}^{kj} = X_{ts}^j X_{ts}^k.$$

We define a “norm”¹ on the set of α -Hölder rough path setting

$$\|\mathbf{X}\|_\alpha := \|X\|_\infty + \|X\|_\alpha + \|\mathbb{X}\|_{2\alpha},$$

where $\|\cdot\|_\gamma$ stands for the γ -Hölder norm of a 1- or 2-index map, for any $0 < \gamma < 1$.

¹This defines only a distance, since the space of rough paths is not a vector space.

To make sense of these conditions, think of $\mathbb{X}_{ts}^{jk}$ as $\int_s^t (X_r - X_s)^j dX_r^k$, even though this integral does not make sense in our setting. (Young integration theory [28, 22] can only make sense of such integrals if X is α -Hölder, with $\alpha > \frac{1}{2}$.) When X is smooth its increments have size $(t - s)$, and the increments $\mathbb{X}_{ts}$ have size $(t - s)^2$. Convince yourself that relation (3) comes in that model setting from Chasles' relation $\int_s^u + \int_u^t = \int_s^t$. The symmetry condition satisfied by weak geometric rough paths is satisfied by the rough paths lift of any smooth path. We invite the reader to check that if B stands for a Brownian motion and $\frac{1}{3} \leq \alpha < \frac{1}{2}$, we define a weakly geometric α -Hölder rough path setting

$$\mathbf{B}_{ts} = \left(B_t - B_s, \int_s^t (B_r - B_s) \otimes \circ dB_r \right),$$

where the above integral is a Stratonovich integral. See Chapter 3 in [14] for more background in this direction.

Let denote by $\mathbf{V}$ the collection of some vector fields $(V_1, \dots, V_\ell)$ on $\mathbb{R}^d$.

DEFINITION 2. Fix a finite time horizon T , and let $\mathbf{X}$ be a weak geometric α -Hölder rough path, with $\alpha \in (\frac{1}{3}, \frac{1}{2}]$. A path $x_\bullet : [0, T] \rightarrow \mathbb{R}^d$ is said to solve the rough differential equation

$$(4) \quad dx_t = \mathbf{V}(x_t) \mathbf{X}(dt)$$

if, for any function $f \in \mathcal{C}_b^3$, we have the following Taylor–Euler expansion:

$$(5) \quad f(x_t) = f(x_s) + X_{ts}^i (V_i f)(x_s) + \mathbb{X}_{ts}^{jk} (V_j V_k f)(x_s) + O(|t - s|^{3\alpha}).$$

As a sanity check, one can easily verify that if the vector fields V_i are constant, the solution path to the above rough differential equation, started from x_0 , is given by just $x_t = x_0 + X_{ts}^i V_i$.

It can be proved that if $(X, \mathbb{X})$ is the rough path lift of Brownian motion introduced above, then a solution path to the above rough differential equation gives a version of the solution to the Stratonovich differential equation

$$dx_t = V_i(x_t) \circ dB_t^i.$$

This is what makes rough paths theory so appealing for applications to stochastic calculus. See, e.g., Chapters 9–11 in Friz and Hairer's book [14].

Rather than solving the rough differential equation (4) for each fixed starting point, we can actually construct a flow of maps $(\varphi_{ts})_{0 \leq s \leq t \leq T}$ with the property that the path $(\varphi_{\bullet 0}(x))$ is for each starting point x the solution to (4) started from x . Given a bounded Lipschitz continuous vector field W on $\mathbb{R}^d$ we denote by $\exp(W)$ the time 1 map of W , which associates to any point $x \in \mathbb{R}^d$ the value at time 1 of the solution to the ordinary differential equation $\dot{y} = W(y)$, started from x . Fix $0 \leq s \leq t \leq T$ and denote by $\mathbb{A}_{ts}$ the antisymmetric part of $\mathbb{X}_{ts}$. Setting

$$\mu_{ts} := \exp \left(\sum_{i=1}^{\ell} X_{ts}^i V_i + \sum_{1 \leq j < k \leq \ell} \mathbb{A}_{ts}^{jk} [V_j, V_k] \right),$$

an elementary Taylor–Euler expansion [2, section 3], using the weak geometric character of $\mathbf{X}$, shows that μ_{ts} satisfies the kind of Taylor–Euler expansion we expect from a solution path to the above rough differential equation, as we have

$$\left| f \circ \mu_{ts} - \left(f + X_{ts}^i V_i f + \mathbb{X}_{ts}^{jk} V_j V_k f \right) \right| \leq c(\mathbf{X}, f) |t - s|^{3\alpha}.$$

Even though μ_{ts} is *not* a flow of maps, it turns out it is very close to one, as seen in Theorem 3 below. It was first proved in a different form by Lyons [21, Theorem 4.1.1] and was named “Lyons’ universal limit theorem” by Malliavin. Its present form is a mix of Lyons’ original result, Davie’s approach [11], and the first author’s approach [2] to rough differential equations. A vector field of class $\mathcal{C}_b^2$, with Lipschitz second derivative, is said to be Lip^3 in the sense of Stein.

THEOREM 3 (Lyons’ universal limit theorem). *Let $\mathbf{X}$ be a weak geometric α -Hölder rough path, $\alpha \in (\frac{1}{3}, \frac{1}{2}]$. Let $\mathbf{V} = (V_i)_{i=1,\dots,\ell}$ be a collection of Lip^3 vector fields on $\mathbb{R}^d$.*

- (i) *There exists a unique flow of maps $(\varphi_{ts})_{0 \leq s \leq t \leq T}$ on $\mathbb{R}^d$ such that the inequality*

$$\|\varphi_{ts} - \mu_{ts}\|_{\infty} \leq c(\mathbf{X}, \mathbf{V}) |t - s|^{3\alpha}$$

holds for some positive constant $c(\mathbf{X}, \mathbf{V})$ depending only on $\mathbf{X}$ and $\|\mathbf{V}\|_{\text{Lip}^3}$ for all times $0 \leq s \leq t \leq T$.

- (ii) *Given any starting point $x \in \mathbb{R}^d$, there exists a unique solution path $x_{\bullet}$ to the rough differential equation (4); this solution path is actually given, for all $0 \leq t \leq T$, by*

$$x_t = \varphi_{t0}(x_0).$$

- (iii) *The solution path $x_{\bullet} \in (\mathcal{C}([0, T], \mathbb{R}^d), \|\cdot\|_{\infty})$ depends continuously on $\mathbf{X}$.*

The crucial point in the above statement is the continuous dependence of the solution path as a function of the rough signal $\mathbf{X}$, in stark contrast with the fact that solutions of stochastic differential equations are only measurable functionals of the Brownian path. When solving rough differential equations driven by lifted Brownian motion (and thereby obtaining a version of the Stratonovich solution to the corresponding SDE, as mentioned above) the only purely measurable operation that has to be done is in defining the iterated integrals $\int_s^t (B_r - B_s) \otimes \circ dB_r$; once this is done, the machinery for solving the rough differential equation (4) is deterministic and continuous with respect to the Brownian rough path. This continuity result was used, for instance, to give streamlined proofs of deep results in stochastic analysis such as the Stroock–Varadhan support theorem for diffusion processes, or the basics of the Freidlin–Wentzell theory of large deviations for diffusion processes [19].

Parts (ii) and (iii) of this theorem can be proved in several ways. The original approach of Lyons [21] was to recast it under a fixed point problem involving a rough integral, which has to be defined first. (This argument has been streamlined by Gubinelli in [16]; see also the monograph [14, Chapter 4].) Existence and well-posedness results using a second order Milstein-type scheme of the form (2) were introduced in that setting by Davie in [11] and generalized in the work [2] of the first author to deal with rough differential equations driven by weak geometric α -Hölder rough paths, for any $0 < \alpha < 1$, using a geometric approach with roots in the work [27] of Strichartz and the novel tool of approximate flows. Part (1) of the above theorem is from [2, section 3].

We refer the reader to the textbooks [23] of Lyons and Qian, and [15] of Friz and Victoir, for a thorough account of the theory, and to the lecture notes [8, 14, 1] for shorter pedagogical accounts.

3. The inverse problem. We would like to propose the examples of application of rough path theory given above, and many others, as an illustration of one of

Lyons' leitmotivs: *Rough paths are not mathematical abstractions; they appear in Nature.* Starting from this postulate, and keeping in mind that rough paths can be understood as a convenient mathematical setting for describing both the macroscopic and microscopic scales of highly oscillating signals, the aim of this work is to answer an important question that comes with this postulate: *Can one observe and record a rough path?* We shall handle this problem in the model setting of a physical system associated with a rough differential equation, which leads to the following question. *Under what conditions on the driving vector fields can one recover the driving rough path by observing the solution flow to that equation?* It is indeed not always possible to reconstruct the driving signal; as an example, take a rough differential equation with constant vector fields, where the second level of the rough path has no influence on the solution, as made clear after Definition 2.

To give a motivating example, assume that two dynamics are described by two rough differential equations *driven by the same rough path*. Think, for instance, of two multidimensional assets. It may happen that one is interested in one of these assets while one can only observe the other. How should we proceed then if one wants to make a trade only when the unobservable asset is in some given region of its state space? If we could reconstruct approximatively the rough signal from the observation of the first asset dynamics, we could use the continuity statement in Lyons' universal limit theorem, Theorem 3, to plug this approximate signal into the dynamics of the second asset and get an approximate path whose distance to the true second asset path is quantifiable, leading to a trading strategy. Although we shall not develop this example of optimal stopping problem with incomplete information, we shall give other examples where the approximate (or, ideally, exact) reconstruction of a rough signal allows one to compensate for an a priori lack of information.

To be more specific, our problem reads as follows. Given some sufficiently regular vector-field valued 1-form $V = (V_1, \dots, V_\ell)$ on $\mathbb{R}^\ell$, and a weak geometric α -Hölder rough path $\mathbf{X}$ over $\mathbb{R}^\ell$, with $\alpha \in (\frac{1}{3}, \frac{1}{2}]$, defined on the time interval $[0, 1]$ say, denote by $(\varphi_{ts})_{0 \leq s \leq t \leq 1}$ the solution flow [2] to the rough differential equation

$$(6) \quad dx_r = V(x_r) \mathbf{X}(dr)$$

in $\mathbb{R}^d$. (This equation is the correct form that (1) takes when the control h is a rough path $\mathbf{X}$. Not only can it be solved for each fixed initial condition, but it also defines a flow of maps, as we have seen in section 2. The map φ_{ts} associates to x the value at time t of the solution to (6) started at time s from x .) Assume we observe increments of the different solution paths, started from c distinct points x_j , $j = 1, \dots, c$; that is, we have access to the data

$$z_{ts}^{x_1, \dots, x_c} := \left(\varphi_{ts}(x_1), \dots, \varphi_{ts}(x_c) \right).$$

Our goal is to reconstruct the driving signal $\mathbf{X}$ using uniquely this information. As the counterexample of constant vector fields shows, the ability to do so depends on the 1-form V .

We first make precise what we mean by saying that “reconstruction is possible.” Fix $\alpha \in (\frac{1}{3}, \frac{1}{2}]$ for the rest of the paper.

DEFINITION 4. *The 1-form V is said to have the reconstruction property if one can find an integer $c \geq 1$, some points $x_1, \dots, x_c \in \mathbb{R}^d$, a constant $a > 1$, and a function $\mathcal{X} : \mathbb{R}^{cd} \rightarrow T^2(\mathbb{R}^\ell) \cong \mathbb{R}^\ell \oplus (\mathbb{R}^\ell)^{\otimes 2}$, with components $\mathcal{X}^1$ and $\mathcal{X}^2$, such that one can associate to every positive constant M another positive constant C_M such*

that the inequalities

$$(7) \quad \left| \mathcal{X}^1(z_{ts}^{x_1, \dots, x_c}) - X_{ts} \right| \leq C_M |t - s|^a, \quad \left| \mathcal{X}^2(z_{ts}^{x_1, \dots, x_c}) - \mathbb{X}_{ts} \right| \leq C_M |t - s|^a$$

hold for all weak geometric α -Hölder rough paths $\mathbf{X}$ with $\|\mathbf{X}\| \leq M$, for all times $0 \leq s \leq t \leq 1$ sufficiently close.

These inequalities ensure that the $T^2(\mathbb{R}^\ell)$ -valued functional $\mathcal{X}(z_{ts}^{x_1, \dots, x_c})$ is almost-multiplicative [22, Definition 4.2], with associated multiplicative functional $\mathbf{X}$. Hence, by a fundamental result of Lyons [22, Theorem 4.3], one can—in principle—completely reconstruct $\mathbf{X}_{ts}$ from the knowledge of the $\mathcal{X}(z_{ba}^{x_1, \dots, x_c})$, with $s \leq a \leq b \leq t$.

Remark 5. Since $\mathbf{X}$ is weak-geometric, the symmetric part of $\mathbb{X}_{ts}$ is equal to $\frac{1}{2}X_{ts} \otimes X_{ts}$. So the essential information in the rough path $\mathbf{X}$ is given by X_{ts} and the antisymmetric part $\mathbb{A}_{ts}$ of $\mathbb{X}_{ts}$. This pair lives in $\mathbb{R}^{\frac{\ell(\ell+1)}{2}}$. For the reconstruction property to hold, one can then alternatively find a function $\mathcal{R}$ such that

$$\left| \mathcal{R}(z_{ts}^{x_1, \dots, x_c}) - (X_{ts}, \mathbb{A}_{ts}) \right| \leq C |t - s|^a;$$

this is actually what we shall do in the proof of the main theorem below.

Our main result takes the form of a sufficient and necessary condition on the 1-form $\mathbf{V}$ for (6) to have the reconstruction property. Only brackets of the form $[V_j, V_k]$, with $j < k$, appear in the matrix below.

THEOREM 6 (reconstruction). *Let $\mathbf{V} = (V_1, \dots, V_\ell)$ be a $\text{Lip}^3(\mathbb{R}^d)$ -valued 1-form on $\mathbb{R}^\ell$. Set*

$$m := \frac{\ell(\ell+1)}{2}.$$

Then (6) has the reconstruction property if and only if there exists an integer c and some points $x_1, \dots, x_c$ in $\mathbb{R}^d$ such that the $(cd \times m)$ matrix

$$\mathbf{M} := \begin{pmatrix} V_1(x_1) & \cdots & V_\ell(x_1) & [V_1, V_2](x_1) & \cdots & [V_{\ell-1}, V_\ell](x_1) \\ \vdots & & \vdots & \vdots & & \vdots \\ V_1(x_c) & \cdots & V_\ell(x_c) & [V_1, V_2](x_c) & \cdots & [V_{\ell-1}, V_\ell](x_c) \end{pmatrix}$$

has rank m . In this case, the exponent a in the definition of the reconstruction property can be chosen to be equal to 3α . We call $\mathbf{M}$ the reconstruction matrix.

The above rank condition holds, for instance, if $\ell = 2$ and $(V_1, V_2, [V_1, V_2])$ forms a free family at some point—for which we need $d \geq 3$. One can actually prove, by classical transversality arguments, that if $x_1, \dots, x_m$ are any given family of $m := \frac{\ell(\ell+1)}{2}$ distinct points in $\mathbb{R}^d$, then the set of tuples $(V_1, \dots, V_\ell)$ of Lip^3 -vector fields on $\mathbb{R}^d$ for which the reconstruction matrix has rank m is dense in $(\text{Lip}^3)^\ell$. This means that one can always reconstruct the rough signal in a “generic” rough differential equation from observing its solution flow at no more than m points. (See the books by Hirsch [18] and Zeidler [29] for a gentle introduction to transversality-type arguments.) This genericity result obviously does not mean that any tuple $(V_1, \dots, V_\ell)$ of vector fields enjoys that property, as shown by the above example with the constant vector fields corresponding to the canonical basis.

Examples. Here are a few illustrative examples where Theorem 6 applies and reconstruction is possible.

1. Note that the above condition on the reconstruction matrix is unrelated to Hörmander's bracket condition, and that there is no need of any kind of ellipticity or hypoellipticity for Theorem 6 to apply. If a $\text{Lip}^3(\mathbb{R}^d)$ -valued 1-form on $\mathbb{R}^\ell$ has the reconstruction property, its trivial extension $\tilde{V} = (\tilde{V}_1, \dots, \tilde{V}_\ell)$ to vector fields on $\mathbb{R}^{d+1} \simeq \mathbb{R}^d \times \mathbb{R}$, with the $\tilde{V}_i = (V_i, 0)$, does not involve a hypoelliptic system while is still has the reconstruction property. As another example of a nonelliptic control system satisfying the assumptions of Theorem 6, consider in $\mathbb{R}^3$, with coordinates (x, y, z) , the following three vector fields:

$$V_1(x, y, z) = yz\partial_x, \quad V_2(x, y, z) = xz\partial_y, \quad V_3(x, y, z) = x\partial_z.$$

Then

$$[V_1, V_2] = z^2(x\partial_x - y\partial_y), \quad [V_1, V_3] = y(z\partial_z - x\partial_x), \quad [V_2, V_3] = -x^2\partial_y.$$

Here $m = 6$, and it is easily checked that taking two observation points (i.e., $c = 2$), such as the points with coordinates $(1, 1, 1)$ and $(1, 2, 3)$, the reconstruction matrix has rank 6.

2. Hypoellipticity (or ellipticity) is also not sufficient for Theorem 6 to hold. Indeed, take $d \geq 2$ and consider

$$X_i = \partial_{x_i} + 2y_i \partial_t, \quad Y_i = \partial_{y_i} - 2x_i \partial_t$$

in $\mathbb{R}^{2d+1}$ with coordinates $x \in \mathbb{R}^d, y \in \mathbb{R}^d$, and $t \in \mathbb{R}$ used in a sub-Riemannian setting to define the Kohn Laplacian $\sum_{i=1}^d (X_i^2 + Y_i^2)$. They satisfy

$$[X_i, X_j] = 0, \quad [Y_i, Y_j] = 0, \quad [X_i, Y_j] = -4\delta_i^j \partial_t,$$

so the reconstruction matrix is always degenerate.

Our method of proof is best illustrated with the example of the rolling ball; see [7] for a thorough treatment and [8, section 2.1] for its introduction in a rough path setting. This equation describes the motion of a ball with unit radius rolled on a table without slipping. The position of the ball at time t is determined by the orthogonal projection $x_t \in \mathbb{R}^2$ of the center of the ball on the table (i.e., the point touching the table, with the latter identified with $\mathbb{R}^2$), and by a (3×3) orthonormal matrix $\Phi_t \in \mathbb{O}(\mathbb{R}^3)$ giving the orientation of the ball. Set

$$A_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} \quad \text{and} \quad A_2 = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

We define right invariant vector fields V_1, V_2 on $\mathbb{O}(\mathbb{R}^3)$ by the formula

$$V_1(M) = A_1 M, \quad V_2(M) = A_2 M$$

for any $M \in \mathbb{O}(\mathbb{R}^3)$. The nonslipping assumption on the motion of the ball relates the evolution of the path $x_\bullet$ to that of $\Phi_\bullet$, when the path $x_\bullet$ is C^1 , as follows:

$$(8) \quad d\Phi_t = V_1(\Phi_t) dx_t^1 + V_2(\Phi_t) dx_t^2.$$

This equation makes perfect sense when $x_\bullet$ is replaced by a rough path $\mathbf{X}$ and the equation is understood in a rough path sense. Set $\mathbf{V} = (V_1, V_2)$. Working with invariant vector fields, the solution flow to the rough differential equation

$$(9) \quad d\varphi_t = \mathbf{V}(\varphi_t) \mathbf{X}(dt)$$

is given by the map

$$\varphi_\bullet : g \in \mathbb{O}(\mathbb{R}^3) \mapsto \Phi_\bullet^0 g,$$

where $\Phi_\bullet^0$ is the solution path to the rough differential equation (9) started from the identity. We know from the work of Strichartz [27] on the Baker–Campbell–Dynkin–Hausdorff formula that the solution to the time-inhomogeneous ordinary differential equation (8) is formally given by the time-1 map of a *time-homogeneous* ordinary differential equation involving a vector field explicitly computable in terms of V_1, V_2 and their brackets, and the iterated integrals of the signal $x_\bullet$, under the form of an infinite series. Truncating this series provides an approximate solution whose accuracy can be quantified precisely under some mild conditions on the driving vector fields. This picture makes perfect sense in the rough path setting of (9) and forms the basis of the flow method put forward in [2]. In the present setting, given a two-dimensional rough path $\mathbf{X}$, with Lévy area process $\mathbb{A}_\bullet$, and given $0 \leq s \leq t \leq 1$, denote by ψ_{ts} the time-1 value of the solution path to the ordinary differential equation

$$dz_u = X_{ts}^1 V_1(z_u) + X_{ts}^2 V_2(z_u) + \mathbb{A}_{ts}[V_1, V_2](z_u), \quad 0 \leq u \leq 1,$$

in $\mathbb{O}(\mathbb{R}^3)$ started from the identity; that is,

$$\psi_{ts} = \exp \left(X_{ts}^1 A_1 + X_{ts}^2 A_2 + \mathbb{A}_{ts}[A_2, A_1] \right).$$

Write φ_{ts} for $\varphi_t \varphi_s^{-1}$. Then it follows from the results in [2, section 3.2] that there exists some positive constant c_1 such that the inequality

$$(10) \quad \|\varphi_{ts} - \psi_{ts}\|_\infty \leq c_1 |t - s|^{3\alpha}$$

holds for all $0 \leq s \leq t \leq 1$. Since the vectors $A_1, A_2, [A_2, A_1]$ form a basis of the vector space $\mathcal{A}_3$ of antisymmetric (3×3) matrices, and the exponential map is a local diffeomorphism between a neighborhood of 0 in $\mathcal{A}_3$ and a neighborhood of the identity in $\mathbb{O}(\mathbb{R}^3)$, we get back the coefficient X_{ts}^1, X_{ts}^2 , and $\mathbb{A}_{ts}$ from the knowledge of φ_{ts} and relation (10), up to an accuracy of order $|t - s|^{3\alpha}$. This shows that one can reconstruct $\mathbf{X}$ from $\varphi_\bullet$, in the sense of Definition 4, as the diagonal terms of $\mathbb{X}_{ts}$ are given in terms of X_{ts} (see Remark 5). One could argue that perfect knowledge of φ_{ts} may seem unrealistic from a practical point of view. Note that the above proof makes it clear that it is sufficient to know φ_{ts} up to an accuracy of order $|t - s|^{3\alpha}$ to get the reconstruction result.

4. Proof of the reconstruction theorem. Our proof is based on the basic approximation method put forward in [2] to construct the solution flow to a rough differential equation, and used independently later in some works [6] and [25] of Lyons and coauthors. As in the rolling ball example, it rests on the fact that one can obtain a good approximation of the solution flow φ_{ts} to the rough differential equation (6)

by looking at the time-1 map of an auxiliary time-homogeneous ordinary differential equation constructed from the vector fields V_i , their brackets, and $\mathbf{X}_{ts}$. More specifically, let ψ_{ts} stand for the time-1 map of the ordinary differential equation

$$(11) \quad dz_u = \sum_{i=1}^{\ell} X_{ts}^i V_i(z_u) + \sum_{1 \leq j < k \leq \ell} \mathbb{A}_{ts}^{jk} [V_j, V_k](z_u), \quad 0 \leq u \leq 1,$$

that associates to any $x \in \mathbb{R}^d$ the value at time 1 of the solution to the above equation started from x . Then, by Theorem 3, there exists a positive constant c_1 such that one has

$$(12) \quad \|\varphi_{ts} - \psi_{ts}\|_{\infty} \leq c_1 |t - s|^{3\alpha}$$

for all $0 \leq s \leq t \leq 1$. The constant c_1 depends only on $\|V\|_{\text{Lip}^3}$ and any upper bound M on the rough path norm of $\mathbf{X}$. We write formally

$$\psi_{ts} = \exp \left(X_{ts}^i V_i + \mathbb{A}_{ts}^{j < k} [V_j, V_k] \right)$$

and set $m = \frac{\ell(\ell+1)}{2}$. Working with s and t close to each other, we expect the coefficients of $\mathbf{X}_{ts}$ appearing in (11) to lie in any a priori given closed ball $B_{\epsilon}(0)$ of $\mathbb{R}^m$ with center 0 and radius ϵ . The simplest idea to get them back from the knowledge of φ_{ts} is then to try and minimize over $B_{\epsilon}(0)$ the quantity

$$(13) \quad \left\| \varphi_{ts} - \exp \left(A^i V_i + B^{j < k} [V_j, V_k] \right) \right\|_{\infty}.$$

Remark 7. Note that the “approximation scheme” ψ_{ts} is equal to φ_{ts} itself in the very special case where $d = 1$ and $\ell = 1$ (actually only $\ell = 1$ is necessary), by the well-known Doss–Sussmann representation. So if in that case there is a point $y \in \mathbb{R}$ with $V_1(y) \neq 0$, the map $\mathfrak{f} : a \in \mathbb{R} \mapsto \exp(aV_1)(y)$ is a local diffeomorphism between a neighborhood of 0 in $\mathbb{R}$ and a neighborhood $\mathcal{V}$ of y in $\mathbb{R}$. One thus has $X_{ts} = \mathfrak{f}^{-1}(\varphi_{ts}(y))$ for s and t close enough for $\varphi_{ts}(y)$ to be in $\mathcal{V}$. The reconstruction is perfect in that case. (Note that a one-dimensional rough path does not have an “area”.)

Proof of Theorem 6.

Sufficiency. Assume for the moment that $d \geq m$, and suppose that at some point $y \in \mathbb{R}^d$ the vectors

$$\left(V_i(y), [V_j, V_k](y); 1 \leq i \leq \ell, 1 \leq j < k \leq \ell \right)$$

are independent. Define a map Ψ_y from $\mathbb{R}^m$ to $\mathbb{R}^d$ setting

$$\Psi_y(A, B) = \exp \left(A^i V_i + B^{j < k} [V_j, V_k] \right)(y).$$

By Lemma 12 there exist two explicit positive constants ϵ_1, ϵ_2 , depending only on the Lip^3 -norm of $V_1, \dots, V_{\ell}$, such that for any two points $\mathbf{a}, \mathbf{a}'$ in the ball $B_{\epsilon_1}(0)$ of $\mathbb{R}^m$, we have

$$(14) \quad \|\Psi_y(\mathbf{a}) - \Psi_y(\mathbf{a}')\| \geq \epsilon_2 \|\mathbf{a} - \mathbf{a}'\|.$$

We claim that any minimizer (A, B) in $B_{\epsilon_1}(0)$ of the expression

$$(15) \quad \left| \varphi_{ts}(y) - \exp \left(A^i V_i + B^{j < k} [V_j, V_k] \right) (y) \right|$$

satisfies the identity²

$$(16) \quad (A, B) - (X_{ts}, \mathbb{A}_{ts}) = O(|t - s|^{3\alpha})$$

for $t - s$ small enough, with a constant in the $O(\cdot)$ term independent of the minimizer. Assume, by contradiction, the existence for every $M > 0$ and $\delta > 0$ of times (s, t) with $0 \leq t - s \leq \delta$, and some minimizer (A_{ts}, B_{ts}) in $B_{\epsilon_1}(0)$ such that

$$\left| (A_{ts}, B_{ts}) - (X_{ts}, \mathbb{A}_{ts}) \right| \geq M|t - s|^{3\alpha}.$$

Then the inequality

$$\left| \Psi_y(A_{ts}, B_{ts}) - \Psi_y(X_{ts}, \mathbb{A}_{ts}) \right| \geq \epsilon_2 M |t - s|^{3\alpha}$$

would follow from (14), giving, for a choice of $M = \frac{2c_1}{\epsilon_2}$, the conclusion

$$\left| \Psi_y(A_{ts}, B_{ts}) - \varphi_{ts}(y) \right| > c_1 |t - s|^{3\alpha},$$

contradicting identity (12), where $(X_{ts}, \mathbb{A}_{ts})$ belongs to $B_{\epsilon_1}(0)$ for δ small enough, and the fact that (A_{ts}, B_{ts}) is a minimizer. This proves Theorem 6 in the special case where $d \geq m$ and where for some $y \in \mathbb{R}^d$ the family $(V^i(y), [V_j, V_k](y); 1 \leq i \leq \ell, 1 \leq j < k \leq \ell)$ is free.

To handle the general case, identify $(\mathbb{R}^d)^c$ and $\mathbb{R}^{cd}$, and denote by $z = (y_1, \dots, y_c)$ a generic element of $\mathbb{R}^{cd}$, with $y_i \in \mathbb{R}^d$. Introduce the vector fields W_i on $\mathbb{R}^{cd}$, given by the formula

$$W_i(z) = \begin{pmatrix} V_i(y_1) \\ \vdots \\ V_i(y_c) \end{pmatrix}.$$

These vector fields satisfy, under the assumptions of Theorem 6, the restricted assumptions under which we have proved Theorem 6 above. So this special case applies and implies the general case. The above proof shows in particular that

$$\left| (A_{ts}, B_{ts}) - (X_{ts}, \mathbb{A}_{ts}) \right| \leq \frac{2c_1}{\epsilon_2} |t - s|^{3\alpha}.$$

Necessity. Define the linear space

$$Z := \text{span} \{ V_1(y), \dots, V_\ell(y), [V_1, V_2](y), \dots, [V_{\ell-1}, V_\ell](y); y \in \mathbb{R}^d \}$$

and assume that $Z \neq \mathbb{R}^m$. Pick a nonzero vector v in the orthogonal complement of Z . Then for every $\mathbf{a} \in \mathbb{R}^m$ and every $y \in \mathbb{R}^d$,

$$(17) \quad \Psi_y(\mathbf{a}) = \Psi_y(\mathbf{a} + v).$$

²This is a slight abuse of notation, since $\mathbb{A}$ is a matrix. By $(X_{ts}, \mathbb{A}_{ts})$ we mean here the vector $(X_{ts}^i, \mathbb{A}_{ts}^{j,k} : i = 1, \dots, \ell, 1 \leq j < k \leq \ell)$.

Hence the null rough path $(X, \mathbb{X}) := (0, 0)$ and the weak geometric α -rough path

$$\bar{X}_{ts} := \begin{pmatrix} v^1 \\ \vdots \\ v^\ell \end{pmatrix} (t-s),$$

$$\bar{\mathbb{X}}_{ts} := \frac{1}{2}(v^1, \dots, v^\ell) \otimes (v^1, \dots, v^\ell)(t-s)^2 + \begin{pmatrix} 0 & v^{\ell+1} & \dots & v^{2\ell-1} \\ -v^{\ell+1} & 0 & \dots & \vdots \\ \vdots & \vdots & \dots & v^m \\ -v^{2\ell-1} & \dots & -v^m & 0 \end{pmatrix} (t-s)$$

have the same effect on the rough differential equation, since $(\bar{X}_{ts}, \bar{\mathbb{A}}_{ts}) = (t-s)v$ (with the same abuse of notation as in the proof of sufficiency). Reconstruction is then not possible in this case. $\square$

Remarks 8. 1. The above proof shows that any minimizer to problem (15) satisfies identity (16) if $0 \leq t-s \leq \delta$, provided $\delta > 0$ is chosen such that $(X_{ts}, \mathbb{X}_{ts}) \in \mathcal{U}$. The results of [2, sections 2–3] show that δ is of order $(1 + \|\mathbf{X}\|)^{-3}$; this quantity is a priori unknown since $\mathbf{X}$ itself is unknown. In practice, one should work with a sufficiently small a priori given δ and refine it if necessary.

2. Note that the use of ordinary differential equations as a tool makes the above method perfectly suited for dealing with rough differential equations (6) with values in manifolds. This would not have been the case if we had replaced the exponential map used to define ψ_t by a Taylor polynomial (as we shall do below, in the second proof of Theorem 6), which does not have any intrinsic meaning on a manifold. Denote by Vf the derivative of a function f in the direction of a vector field V . In the manifold setting, the reconstructability condition takes the following form. There exist a $\mathcal{C}^3$ function f on the manifold and some points $x_1, \dots, x_c$ such that the following matrix has rank m :

$$\begin{pmatrix} (V_1 f)(x_1) & \dots & (V_\ell f)(x_1) & ([V_1, V_2]f)(x_1) & \dots & ([V_{\ell-1}, V_\ell]f)(x_1) \\ \vdots & & \vdots & \vdots & & \vdots \\ (V_1 f)(x_c) & \dots & (V_\ell f)(x_c) & ([V_1, V_2]f)(x_c) & \dots & ([V_{\ell-1}, V_\ell]f)(x_c) \end{pmatrix}.$$

Keeping in mind the rolling ball example, it suffices in that case to observe the flow at only one point and to take as function f the logarithm map, from $\mathbb{O}(\mathbb{R}^3)$ to the linear space of antisymmetric (3×3) matrices.

3. We only deal in the reconstruction theorem with dynamics driven by weak geometric α -Hölder rough paths for $\alpha > \frac{1}{3}$; this is the most interesting regime for applications; see section 6. It is possible to extend the sufficiency part of the reconstruction theorem to dynamics driven by a weak geometric α -Hölder rough path, for any regularity index α , using the same approach as above building on the work [2], that covers dynamics driven by any weak geometric Hölder rough paths. The criterion becomes awkward to write in the general case though, due to the Jacobi relations between the different brackets $[V_i, [V_j, V_k]]$ and higher iterated brackets. An optimal necessary and sufficient condition for reconstruction requires the choice and use of a Hall basis.

5. Examples. We present some examples of dynamical systems where the conditions of Theorem 6 hold. The examples are physical, in the sense that they model dynamics that can be realized as concrete machines.

5.1. Rolling ball and its implementations. The rolling ball was already considered at the end of section 3.

Another incarnation is available in the field of quantum control. In fact the problem studied in [5] is of the same form (see [5, eqn. (13)]). We should point out a caveat in this case: due to quantum effects the dynamics of the system are only well-described by the differential equation of the rolling ball for forcing signals with derivative of small amplitude [4]. This is of course not the case for a general rough path.

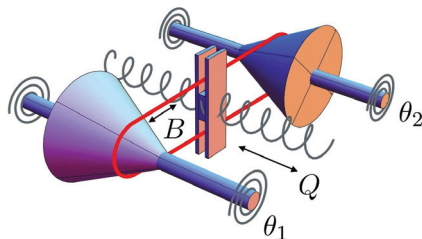


FIG. 1. An implementation of the continuously variable transmission. The image is from [26], with kind permission of the authors.

5.2. Continuously variable transmission. The uncontrolled dynamics of the system illustrated in Figure 1 is studied in [26]. The controlled setting is as follows. The statespace is given by θ_1, θ_2 , the rotation angles of the two cones, B , the vertical position of the belt, and Q , the translation of the belt, constrained to be in an interval, say $(0, 1)$. The controller acts directly on B and Q , and so for $Q(t) \in (0, 1)$ the corresponding equations are given by

$$\begin{aligned} d\theta_1(t) &= \frac{1}{Q(t)} dB(t), \\ d\theta_2(t) &= \frac{1}{1-Q(t)} dB(t), \\ dB(t) &= dX^1(t), \\ dQ(t) &= dX^2(t). \end{aligned}$$

With $x := (\theta_1, \theta_2, B, Q)$ this reads as

$$dx_t = V_i(x_t) dX_t^i,$$

where

$$V_1(x) = \begin{pmatrix} \frac{1}{q} \\ \frac{1}{1-q} \\ 1 \\ 0 \end{pmatrix}, \quad V_2(x) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix},$$

We calculate

$$[V_1, V_2](x) = \begin{pmatrix} \frac{1}{q^2} \\ -\frac{1}{(1-q)^2} \\ 0 \\ 0 \end{pmatrix}.$$

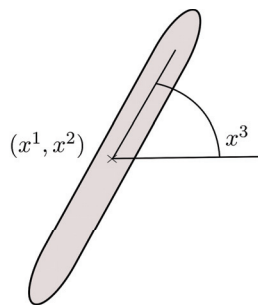


FIG. 2. The unicycle.

The condition of Theorem 6 is hence satisfied with $c = 1$ and any point $x = (\theta_1, \theta_2, q, b)$ with $q \in (0, 1)$.

5.3. Unicycle. Consider a unicycle on the sphere with position (x^1, x^2) and orientation x^3 sketched in Figure 2. The controller acts by changing the orientation x^3 and by moving the unicycle forward or backward along the current orientation. The dynamics hence reads as (see Example 4.3.2 in [3])

$$dx_t = V_1(x_t)dX_t^1 + V_2(x_t)dX_t^2,$$

where

$$V_1(x) := \begin{pmatrix} \cos(x^3) \\ \sin(x^3) \\ 0 \end{pmatrix}, \quad V_2(x) := \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Then

$$[V_1, V_2](x) = \begin{pmatrix} \sin(x^3) \\ -\cos(x^3) \\ 0 \end{pmatrix}.$$

The condition of Theorem 6 is hence satisfied with $c = 1$ and any point.

6. Applications. We describe in this section two applications of the reconstruction Theorem 6.

6.1. Filtering and maximum likelihood estimator. We give a brief overview on two recent results in the area of stochastic filtering and maximum likelihood estimation which both emphasize the need in different practical situations to measure signals in a rough path sense. The point of the present work is that if these real life signals can be used as input to an *additional* physical system that is modeled by a rough differential equation satisfying the assumptions of our reconstruction theorem, then one can indeed have a good approximation of the rough signal, which suffices for practical purposes.

(a) Filtering. Consider the following multidimensional two-component stochastic differential equation:

$$\begin{aligned}x_t &= x_0 + \int_0^t V(x_r, y_r) dr + \int_0^t V_j(x_r, y_r) \circ dB_r^j \\&\quad + \int_0^t V'_k(x_r, y_r) \circ dy_r^k, \\y_t &= \int_0^t h(x_r, y_r) dr + W_t,\end{aligned}$$

where the path $y_\bullet$ is observed and the path $x_\bullet$ is unobserved. The letters W and B here stand for independent Brownian motions. The stochastic filtering problem consists in calculating the best guess (in L^2 sense) of the signal given the observation

$$\pi_t = \mathbb{E}[f(x_t) \mid \sigma(y_s : s \leq t)].$$

Here f is some smooth nice enough function. (More generally one is interested in the *distribution* of x_t given the past of $y_\bullet$.) From a practitioner's point of view it is highly desirable that the estimation procedure be continuous in the observation path [9]. If the vector fields V'_k are null (the uncorrelated case) or if $y_\bullet$ is one-dimensional, it has been shown that this in fact the case: π_t is continuous with respect to the path $y_{[0,t]}$, where distance is measured in supremum norm (see [9, 12]). Counterexamples show that in the general case this is *not* true. As recalled above in section 2, Stratonovich integrals with continuous semimartingale integrand and the corresponding rough paths integrals against the rough path lift of the integrand coincide whenever both make sense. So one can recast the above equation for the path $x_\bullet$ as a rough differential equation involving the rough path lift $\mathbf{y}_\bullet$ of the observed path $y_\bullet$.

THEOREM 9 (see [10, Theorems 6 and 7]). *Under appropriate assumptions on the vector fields,³ there exists a continuous deterministic function $\mathfrak{f}$ on the rough path space such that*

$$\pi_t = \mathfrak{f}(\mathbf{y}_{[0,t]}),$$

where $\mathbf{y}_\bullet$ is the Stratonovich lift of $y_\bullet$ to a rough path.

Note that even if we had an observation model

$$dy = h(x)dt + \sigma(y)dW,$$

with σ satisfying the assumptions of Theorem 6 in the ideal case where $c = 1$ (for which we need $d_Y \geq d_W(d_W + 1)/2$), observing just Y is *still not enough*. This comes from the fact that the (random!) drift term introduces an error of order $|t - s|^1$. It is therefore indeed necessary to have an additional “measuring device” which is fed the observation y and is modeled as a rough differential equation satisfying the assumptions of Theorem 6.

(b) Statistics. Consider now the problem of estimating the parameter $A \in L(\mathbb{R}^d)$ in the stochastic differential equation

$$dy_t = A h(y_t) dt + \Sigma(y_t) dB_t$$

³This refers to boundedness and sufficient smoothness; no bracket assumption as in Theorem 6 is needed for this result.

driven by a Brownian motion B or other Gaussian process, by observing $y_\bullet$ on some fixed time interval. Under appropriate conditions, the measures on path space for different A are mutually absolutely continuous, so one can use the method of maximum likelihood estimation, leading to an estimator $\hat{A}_T$ for A . Denote by $\mathbf{y}_\bullet$ the Stratonovich lift of $y_\bullet$ into a rough path.

THEOREM 10 (see [13, Theorem 2]). *Under appropriate assumptions on the functions h and Σ , there exists a subset $\mathcal{D}$ of the space of weak geometric α -Hölder rough paths and a continuous deterministic function $\mathcal{A}_T : \mathcal{D} \rightarrow L(\mathbb{R}^d)$ such that $\mathbf{y}_\bullet$ belongs almost surely to $\mathcal{D}$ and $\mathcal{A}_T(\mathbf{y}_\bullet)$ is almost surely equal to $\hat{A}_T$.*

Again, there exist counterexamples that show that taking only the *path* of the observation as an input does *not* yield a robust estimation procedure.

In both examples the practitioner is hence left with the task of actually *recording* the rough path $\mathbf{Y}$ in order to implement on a practical basis the above theoretical results. The reconstruction theorem provides conditions under which this can be done. This requires from the practitioner to observe another system where the signal of interest serves as an input.

7. Appendix. We gather in this appendix a few elementary lemmas that were used in the proof of the reconstruction theorem. We start with a basic result that is used in the two lemmas below.

LEMMA 11. *Let $W : \mathbb{R}^n \times \mathbb{R}^N \rightarrow \mathbb{R}^N$ be a function of class C^2 such that the vector field $W(a, \cdot)$ is Lipschitz continuous for any fixed $a \in \mathbb{R}^n$. Let $\phi_t(a, \cdot)$ be the flow at time t started at time 0 to the ordinary differential equation*

$$dx_t = W(a, x_t)dt.$$

Then

$$c_0 := \sup_{|a| \leq 1, x \in \mathbb{R}^\ell} |D_a^2 \phi_t(a, x)| < \infty.$$

Proof. Note that ϕ_t is the projection on the last ℓ coordinates of the flow ψ_t to the enlarged equation

$$\begin{aligned} da_t &= 0, \\ dx_t &= W(a_t, x_t)dt. \end{aligned}$$

Hence, for $k = 1, \dots, N$, $i = 1, \dots, n$,

$$\begin{aligned} d\partial_{a_i} \phi_t^k &= \partial_{y_h} W^k(a_t, x_t) \partial_{a_i} \phi_t^h dt + \partial_{a_r} W^k(a_t, x_t) dt, \\ d\partial_{a_j, a_i} \phi_t^k &= \partial_{y_g, y_h} W^k(a_t, x_t) \partial_{a_j} \phi_t^g \partial_{a_i} \phi_t^h dt + \partial_{a_s, y_h} W^k(a_t, x_t) \partial_{a_i} \phi_t^h dt \\ &\quad + \partial_{y_h} W^k(a_t, x_t) \partial_{a_j, a_i} \phi_t^h dt + \partial_{y_g, a_r} W^k(a_t, x_t) \partial_{a_j} \phi_t^g dt + \partial_{a_s, a_r} W^k(a_t, x_t) dt; \end{aligned}$$

see, for example, [15, Chapter 4]. An application of Grönwall's lemma now gives the desired result. $\square$

LEMMA 12 (nondegeneracy of the flow approximation). *Let $N \geq n$ and $W_1, \dots, W_n \in \text{Lip}^2(\mathbb{R}^N)$ be vector fields that are linearly independent at some point $y \in \mathbb{R}^N$. Then the function*

$$\begin{aligned} \Psi_y : \mathbb{R}^n &\rightarrow \mathbb{R}^N, \\ \mathbf{a} = (a_1, \dots, a_n) &\mapsto \exp(a_1 W_1 + \dots + a_n W_n)(y) \end{aligned}$$

is C^2 and there is a neighborhood $\mathcal{U}$ of 0 in $\mathbb{R}^n$ and a positive constant ϵ_1 such that for $\mathbf{a}, \mathbf{a}' \in \mathcal{U}$ we have

$$|\Psi_y(\mathbf{a}) - \Psi_y(\mathbf{a}')| \geq \epsilon_1 |\mathbf{a} - \mathbf{a}'|.$$

Proof. By definition

$$\exp(a_1 W_1 + \cdots + a_n W_n)(y) = Y_1^{a,y},$$

where

$$\begin{aligned} dY_t^{a,y} &= a_i W_i(Y_t^{a,y}) dt, \\ Y_0^{a,y} &= y. \end{aligned}$$

Then

$$\begin{aligned} |Y_t^{a,y} - y| &= \left| \int_0^t a_i W_i(Y_r^{a,y}) dr \right| \leq t |a_i| \|W_i\|_\infty, \\ |Y_1^{a,y} - y - a_i W_i(y)| &= \left| \int_0^1 a_i W_i(Y_r^{a,y}) - W_i(y) dr \right| \\ &\leq |a_i| \int_0^1 \|W_i\|_{\text{Lip}^1} |Y_r^{a,y} - y| dr \\ &\leq |a_i| \|W_i\|_{\text{Lip}^1} |a_j| \|W_j\|_\infty \\ &\leq |a|^2 \|W\|_{\text{Lip}^1}^2. \end{aligned}$$

Hence

$$D|_{\mathbf{a}=\mathbf{0}} \Psi_y = (W_1 \cdots W_n) \in \mathbb{R}^{N \times n}$$

has rank n by assumption. Now

(18)

$$\Psi_y(\mathbf{a}) - \Psi_y(\mathbf{a}') = D|_{\mathbf{a}=\mathbf{0}} \Psi_y(\mathbf{a} - \mathbf{a}') + (D|_{\mathbf{a}=\mathbf{a}'} \Psi_y - D|_{\mathbf{a}=\mathbf{0}} \Psi_y)(\mathbf{a} - \mathbf{a}') + O(|\mathbf{a} - \mathbf{a}'|^2),$$

where the last term is of the form

$$|\mathbf{a} - \bar{\mathbf{a}}|^2 \max_{z \in U} |D^2|_{\mathbf{a}=\mathbf{z}} \Psi_y|$$

for $\mathbf{a}, \mathbf{a}' \in U$. Now, since $D|_{\mathbf{a}=\mathbf{0}} \Phi_y$ has rank n , we have

$$|D|_{\mathbf{a}=\mathbf{0}} \Psi_y(\mathbf{a} - \bar{\mathbf{a}})| \geq 2\epsilon_1 |\mathbf{a} - \bar{\mathbf{a}}|,$$

with $\epsilon_1 := \|(D|_{\mathbf{a}=\mathbf{0}} \Psi_y^T D|_{\mathbf{a}=\mathbf{0}} \Psi_y)^{-1} D|_{\mathbf{a}=\mathbf{0}} \Psi_y^T\|^{-1}$ positive. Note that by Lemma 11, $\|D^2 \Phi_y\| \leq c_0$ in a ball of radius one around zero. Then choosing $\mathcal{U} := B_{\epsilon_2}(0)$, with

$$\epsilon_2 = \frac{1}{2c_0} \wedge 1,$$

the second and third terms in (18) are dominated by $\frac{\epsilon_1}{2} |\mathbf{a} - \mathbf{a}'|$, which yields the desired result. $\square$

Acknowledgments. The first author thanks the U.B.O. for their hospitality. The second author would like to thank Paul Gassiat and Nicolas Perkowski for helpful remarks.

REFERENCES

- [1] I. BAILLEUL, *A Flow-Based Approach to Rough Differential Equations*, preprint, arXiv:1404.0890, 2014.
- [2] I. BAILLEUL, *Flows driven by rough paths*, Rev. Mat. Iberoamericana, 31 (2015).
- [3] A. M. BLOCH, *Nonholonomic Mechanics and Control*, Interdiscip. Appl. Math. 24, Springer, 2003.
- [4] U. BOSCAIN, *Personal communication*.
- [5] U. BOSCAIN, T. CHAMBRION, AND G. CHARLOT, *Nonisotropic 3-Level Quantum Systems: Complete Solutions for Minimum Time and Minimum Energy*, preprint, arXiv:quant-ph/0409022, 2004.
- [6] Y. BOUTAIB, L. GYURKO, T. LYONS, AND D. YANG, *Dimension-Free Euler Estimates of Rough Differential Equations*, preprint, arXiv:1307.4708, 2013.
- [7] R. W. BROCKETT AND L. DAI, *Non-holonomic kinematics and the role of elliptic functions in constructive controllability*, in Nonholonomic Motion Planning, Springer US, 1993, pp. 1–21.
- [8] M. CARUANA, T. LÉVY, AND T. LYONS, *Differential Equations Driven by Rough Paths*, Lecture Notes in Math. 1908, Springer, Berlin, 2007.
- [9] J. M. C. CLARK, *The design of robust approximations to the stochastic differential equations of nonlinear filtering*, in Communication Systems and Random Process Theory, NATO Advanced Study Inst. Ser. E Appl. Sci. 25, Sijthoff & Noordhoff, Alphen aan den Rijn, 1978, pp. 721–734.
- [10] D. CRISAN, J. DIEHL, AND H. OBERHAUSER, *Robust filtering: Correlated noise and multidimensional observation*, Ann. Appl. Probab., 23 (2013), pp. 2139–2160.
- [11] A. M. DAVIE, *Differential equations driven by rough paths: An approach via discrete approximation*, Appl. Math. Res. Express. AMRX, 2007, abm009.
- [12] M. H. A. DAVIS, *Pathwise nonlinear filtering with correlated noise*, The Oxford Handbook of Nonlinear Filtering, Oxford University Press, Oxford, 2011, pp. 403–424.
- [13] J. DIEHL, P. FRIZ., AND H. MAI, *Pathwise Stability of Likelihood Estimators for Diffusions via Rough Paths*, preprint, arXiv:1311.1061, 2013.
- [14] P. FRIZ AND M. HAIRER, *A Course on Rough Paths*, Springer, Cham, 2014.
- [15] P. K. FRIZ AND N. B. VICTOIR, *Multidimensional Stochastic Processes as Rough Paths: Theory and Applications*, Cambridge Stud. Adv. Math. 120, Cambridge University Press, Cambridge, UK, 2010.
- [16] M. GUBINELLI, *Controlling rough paths*, J. Funct. Anal., 216.1 (2004), pp. 86–140.
- [17] L. GYURKO, T. LYONS, M. KONTKOWSKI, AND J. FIELD, *Extracting Information from the Signature of a Financial Data Stream*, preprint, arXiv:1307.7244, 2013.
- [18] M. HIRSCH, *Differential Topology*, Springer, 1976.
- [19] M. LEDOUX, Z. QIAN, AND T. ZHANG, *Large deviations and support theorem for diffusion processes via rough paths*, Stochastic Process. Appl., 102 (2002), pp. 265–283.
- [20] D. LEVIN, T. LYONS, AND H. NI, *Learning from the Past, Predicting the Statistics for the Future, Learning an Evolving System*, preprint, arXiv:1309.0260, 2013.
- [21] T. LYONS, *Differential equations driven by rough signals*, Rev. Mat. Iberoamericana, 14 (1998), pp. 215–310.
- [22] T. LYONS, M. CARUANA, AND T. LÉVY, *Differential equations driven by rough paths*, Saint Flour Lecture Notes, Lecture Notes in Math. 1908, Springer, Berlin, 2007.
- [23] T. LYONS AND Z. QIAN, *System Control and Rough Paths*, Oxford University Press, Oxford, 2002.
- [24] T. LYONS AND N. VICTOIR, *Cubature on Wiener space. Stochastic analysis with applications to mathematical finance*, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci., 460 (2004), pp. 169–198.
- [25] T. LYONS AND D. YANG, *Rough Differential Equation in Banach Space Driven by Weak Geometric p -Rough Path*, preprint, arXiv:1402.2900, 2014.
- [26] K. MODIN AND O. VERDIER, *Integrability of nonholonomically coupled oscillators*, Discrete Contin. Dyn. Syst., 34 (2014), pp. 1121–1130.

- [27] R. STRICHARTZ, *The Campbell-Baker-Hausdorff-Dynkin formula and solutions of differential equations*, J. Funct. Anal., 72 (1987), pp. 320–345.
- [28] L. C. YOUNG, *An inequality of the Hölder type, connected with Stieltjes integration*, Acta Math., 67 (1936), pp. 251–282.
- [29] E. ZEIDLER, *Nonlinear Functional Analysis and Its Applications: Part 4: Applications to Mathematical Physics*, Springer, 1997.