Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 58-17 (2024) 326–331

Revisiting Stochastic Realization Theory using Functional Itô Calculus ⋆

Tanya Veeravalli * and **Maxim Raginsky** *

* *Department of Electrical and Computer Engineering and the Coordinated Science Laboratory, University of Illinois, Urbana, IL 61801 USA (e-mail: veerava2@illinois.edu, maxim@illinois.edu)*

Abstract: This paper considers the problem of constructing finite-dimensional state space realizations for stochastic processes that can be represented as the outputs of a certain type of a causal system driven by a continuous semimartingale input process. The main assumption is that the output process is infinitely differentiable, where the notion of differentiability comes from the functional Itô calculus introduced by Dupire as a causal (nonanticipative) counterpart to Malliavin's stochastic calculus of variations. The proposed approach builds on the ideas of Hijab, who had considered the case of processes driven by a Brownian motion, and makes contact with the realization theory of deterministic systems based on formal power series and Chen–Fliess functional expansions.

Keywords: Stochastic systems; stochastic realization theory; nonlinear control systems.

1. INTRODUCTION

The problem of (strong) stochastic realization can be stated abstractly as follows (Willems and van Schuppen, 1980): Fix a probability space $(\Omega, \mathcal{F}, \boldsymbol{P})$, with two random variables (measurable mappings) $Y_1 : \Omega \to \mathsf{Y}_1$ and $Y_2 : \Omega \to \mathsf{Y}_2$, where $(\mathsf{Y}_i, \mathcal{Y}_i)$, $i = 1, 2$, are given measurable spaces. The objective is to construct a measurable space $(\mathsf{X}, \mathcal{X})$ and a measurable mapping $X : \Omega \to \mathsf{X}$, such that Y_1 and Y_2 are conditionally independent given X. (We say that X *splits* Y_1 and Y_2.) Here, Y_1 and Y_2 are interpreted as external (or manifest) *output variables* associated with a stochastic system, and X is an internal (or latent) *state variable* that "explains" the correlations between Y_1 and Y_2. One can specify various restrictions on X, such as minimality [i.e., if $X' : \Omega \to \mathsf{X}'$ is another random variable that splits Y_1 and Y_2, then there exists a measurable map $f : \mathsf{X}' \to \mathsf{X}$ such that $X = f(X')$]. As detailed by Willems and van Schuppen (1980), many of the salient features of the stochastic realization problem already appear in this stripped-down formulation.

Our interest here is in the dynamical setting, where Y_1 and Y_2 appear, respectively, as the (strict) past and the future of a given stochastic *process* $Y = (Y(t))_{t \in \mathsf{T}}$, $\mathsf{T} \subseteq \mathbb{R}$. That is, $X = (X(t))_{t \in \mathsf{T}}$ is another process defined on the same probability space, such that, for each t, $X(t)$ splits $((X(s), Y(s))_{s \in (-\infty, t) \cap \mathsf{T}}$ and $((X(s), Y(s))_{s \in [t, +\infty) \cap \mathsf{T}}$. In this case, X is a (Markov) state process and Y is the output process of a stochastic system, and we say that the pair (X, Y) is a *state space realization* of Y. Under minimal regularity assumptions on Y and on $(\Omega, \mathcal{F}, \boldsymbol{P})$, there always exists a state space realization with X given by the so-called *prediction process* of Y in the sense of Knight (1975); cf. Taylor and Pavon (1988, 1989) for a related construction. While the resulting state process has many desirable properties (e.g., it is a strong Markov process that takes values in a compact metric space, and is minimal as described in the preceding paragraph), its generality poses considerable obstacles when it comes to applications.

Of particular interest in applications is the case when the state takes values in a *finite-dimensional* vector space; this, along with the Markov property of the state process, allows for efficient computational implementations of various schemes for estimation or control. Hence, an important problem is to determine whether a given process admits a finite-dimensional state space realization and, if so, how one can go about constructing such a realization. When $Y = (Y(t))_{t \in \mathbb{R}}$ is a stationary Gaussian process taking values in $\mathbb{R}^p$, there is an elegant geometric approach to the problem of stochastic realization that makes contact with the realization theory of deterministic linear systems; cf. the comprehensive text by Lindquist and Picci (2015).

By contrast, there are relatively few results on nonlinear stochastic realization theory for non-Gaussian processes. For example, Lindquist et al. (1982) use Wiener's homogeneous chaos theory (Stroock, 1987) to construct state space realizations for stationary processes that have a Brownian motion innovation representation; generally, the resulting state processes are infinite-dimensional. To the best of our knowledge, the first steps toward a systematic theory of finite-dimensional nonlinear stochastic realization were taken by Hijab (1983a,b). In particular, by representing the process Y as a smooth causal functional of a Brownian motion (in a sense that will be made precise below), Hijab brought Lie-algebraic techniques from the

⋆ This work was supported in part by the NSF under award CCF-2106358 ("Analysis and Geometry of Neural Dynamical Systems") and in part by the Illinois Institute for Data Science and Dynamical Systems (iDS^2), an NSF HDR TRIPODS institute, under award CCF-1934986.

Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2024.10.190

realization theory of nonlinear deterministic systems (for example, Isidori (1995, Ch. 3)) for the stochastic setting. (See also the use of Lie theory by Mitter (1979), Brockett (1980), Hazewinkel and Marcus (1981), and Sussmann (1981) in the context of finite-dimensional realizations of continuous-time nonlinear filters.)

In this paper, we revisit Hijab's approach and show that it can be extended beyond his original setting of processes driven by a Brownian motion to a much wider class of processes driven by sufficiently regular continuous semimartingales. This expands the scope of his approach to the case of processes that are themselves driven by outputs of other systems, e.g., by diffusion processes governed by Itô stochastic differential equations. Moreover, we show that Hijab's concept of smoothness of a process can be made operationally precise using the machinery of *functional Itô calculus* introduced by Dupire (2009) in the context of mathematical finance and later developed by Cont and Fournié (2010, 2013). In particular, functional Itô calculus allows one to define nonanticipative (causal) derivatives [1] of causal functionals of stochastic processes, and, as this paper will show, Hijab's notion of smoothness amounts to infinite differentiability in this sense.

2. CAUSAL STOCHASTIC SYSTEMS

Our starting point will be the following definition (Willems and van Schuppen, 1980):

Definition 1. A *stochastic system* (in output form) consists of a probability space $(\Omega, \mathcal{F}, \boldsymbol{P})$, a time index set $\mathsf{T} \subseteq \mathbb{R}$, a measurable *output space* $(\mathsf{Y}, \mathcal{Y})$, and a stochastic process $Y : \mathsf{T} \times \Omega \to \mathsf{Y}$.

We will take $\mathsf{T} = [0, T]$ with $T > 0$ fixed and $(\mathsf{Y}, \mathcal{Y}) = (\mathbb{R}, \mathcal{B})$, where $\mathcal{B}$ is the Borel σ-algebra of the open subsets of $\mathbb{R}$. This describes a single-output system which we consider for simplicity (can be extended to $p > 1$ outputs). Now, we can think of a system as a physically realizable operator mapping signals to signals. Then, let Ω be a space of sufficiently regular trajectories defined on $[0, T]$. For reasons that will become clear in the sequel, we will take $\Omega = D([0, T], \mathbb{R}^m)$, the Skorohod space of càdlàg (right-continuous with left limits) paths $w : [0, T] \to \mathbb{R}^m$. The Skorohod space can be equipped with a metric d that makes it a complete separable metric space (Billingsley, 1999), and we let $\mathcal{F}$ be the corresponding Borel σ-algebra.

The next notion we need is that of a causal (nonanticipatory) system. Following Georgiou and Lindquist (2013), by a system we understand a measurable map $F : D([0, T], \mathbb{R}^m) \to D([0, T], \mathbb{R})$ that takes m-dimensional input trajectories w to one-dimensional output trajectories Fw. Any such F determines a family of mappings $F(t, \cdot) : D([0, T], \mathbb{R}^m) \to \mathbb{R}$, $0 \le t \le T$, by $F(t, w) := Fw(t)$. For a causal map, $F(t, \cdot)$ depends only on the restriction of w to $[0, t]$. To make this precise, define for each $t \in [0, T]$ the map $\Pi_t : D([0, T], \mathbb{R}^\bullet) \to D([0, T], \mathbb{R}^\bullet)$ by

$$\Pi_t w(s) := \begin{cases} w(s), & s < t \\ w(t), & t \le s \le T \end{cases}.$$

In other words, Π_t maps a trajectory $w(\cdot)$ to the trajectory $w(\cdot \wedge t)$ stopped at time t.

Definition 2. A measurable map $F : D([0, T], \mathbb{R}^m) \to D([0, T], \mathbb{R})$ is a *causal system* if $\Pi_t \circ F \circ \Pi_t = \Pi_t \circ F$ for all $t \in [0, T]$.

Next, let W be the canonical coordinate process on $(\Omega, \mathcal{F})$, i.e., $W(t, w) = w(t)$, and take $\boldsymbol{P}$ to be a probability measure on $(\Omega, \mathcal{F})$ under which W is a continuous semimartingale (Protter, 2005) with $W(0) = 0$ and with a given quadratic variation process

$$[W](t) = \int_0^t Q(s)\,\mathrm{d}s, \qquad 0 \le t \le T \tag{1}$$

where Q is a càdlàg process taking values in the space $\mathbb{R}_+^{m \times m}$ of $m \times m$ positive semidefinite matrices, such that

$$\boldsymbol{P}\{\det Q(t) > 0 \text{ for all } 0 \le t \le T\} = 1. \tag{2}$$

For example, if $Q(t) = I_m$, the $m \times m$ identity matrix, then $\boldsymbol{P}$ is the probability law of a standard m-dimensional Brownian motion on $[0, T]$.

Remark 1. Due to the continuity assumption on the sample paths of W, the support of $\boldsymbol{P}$ is (a subset of) the space $C([0, T], \mathbb{R}^m)$ of continuous paths $w : [0, T] \to \mathbb{R}^m$, which is a proper subset of $D([0, T], \mathbb{R}^m)$. Nevertheless, we will need the entire Skorohod space in order to construct perturbations of paths.

We now impose the following causal realizability condition on the output process Y: There exists a causal system $F : D([0, T], \mathbb{R}^m) \to D([0, T], \mathbb{R})$, such that

$$Y(t) = F(t, W), \qquad \text{for all } t \in [0, T] \tag{3}$$

—this is just a different way of saying that Y is a progressively measurable process defined on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0,T]}, \boldsymbol{P})$, where $(\mathcal{F}_t)_{t \in [0,T]}$ is the natural filtration induced by W. In view of Remark 1, the representation in Equation (3) is not unique because W has continuous sample paths and we can modify F arbitrarily outside the support of $\boldsymbol{P}$ without affecting the output process Y. All we ask is that Y has a version that admits such a representation. In fact, as we discuss next, we will restrict our attention to a smaller class of processes Y for which the map F in (3) is *smooth* in a certain sense.

3. FUNCTIONAL ITÔ CALCULUS

We will make use of the notions of differentiability of a causal system F introduced by Dupire (2009) and developed further by Cont and Fournié (2010; 2013). We say that F has a *time* (or *horizontal*) *derivative* at (t, w) if the limit

$$\partial_0 F(t, w) := \lim_{h \to 0^+} \frac{F(t + h, \Pi_t w) - F(t, w)}{h} \tag{4}$$

exists, and a *space* (or *vertical*) *derivative* at (t, w) in the direction e_i (the ith element of the standard basis in $\mathbb{R}^m$) if the limit

$$\partial_i F(t, w) := \lim_{h \to 0^+} \frac{F(t, w + h 1_{[t,T]} e_i) - F(t, w)}{h} \tag{5}$$

exists. The derivatives $\partial_i F$, $i = 0, \ldots, m$, are themselves causal systems, and we can thus define higher-order derivatives $\partial_i \partial_j F$ etc., provided they exist. It is important to note that the ∂_i's do not commute in general. We will say

[1] Dupire's definition is closely related to the notion of causal derivative due to Fliess (1983).

F is a *smooth causal system* if it is continuous and has continuous derivatives of all orders in the sense of Dupire.

The above definitions form the basis of *functional Itô calculus*, which deals with causal functionals of sufficiently regular stochastic processes. Consider, in particular, a continuous semimartingale W satisfying the conditions (1) and (2). The 'classical' version of Itô's lemma says that, for a $C^{1,2}$ function $f : [0,T] \times \mathbb{R}^m \to \mathbb{R}$, the following holds $\boldsymbol{P}$-a.s. for every $t \in [0,T]$:

$$\begin{aligned} & f(t, W(t)) - f(0, W(0)) \\ &= \int_0^t \partial_0 f(s, W(s))\,\mathrm{d}s + \sum_{i=1}^m \int_0^t \partial_i f(s, W(s))\,\mathrm{d}W_i(s) \\ &\quad + \frac{1}{2} \sum_{i,j=1}^m \int_0^t \partial_i \partial_j f(s, W(s)) Q_{ji}(s)\,\mathrm{d}s, \end{aligned}$$

where $\partial_0 f(s,x) := \frac{\partial f}{\partial s}(s,x)$,

$$\partial_i f(s,x) := \frac{\partial f}{\partial x_i}(s,x), \qquad i = 1, \ldots, m$$

etc., and it is an Itô stochastic integral. We can also rewrite it in Stratonovich form:

$$\begin{aligned} & f(t, W(t)) - f(0, W(0)) \\ &= \int_0^t \partial_0 f(s, W(s))\,\mathrm{d}s + \sum_{i=1}^m \int_0^t \partial_i f(s, W(s)) \circ \mathrm{d}W_i(s). \end{aligned} \tag{6}$$

The functional formulation extends this to causal functionals of W:

Theorem 1. (Dupire, 2009; Cont and Fournié, 2010) Let F be continuous causal system with continuous first- and second-order derivatives $\partial_0 F, \ldots, \partial_m F$ and $\partial_i \partial_j F$, $i, j = 1, \ldots, m$. Let W be a continuous semimartingale satisfying the conditions (1) and (2). Then, for any $t \in [0,T]$, the following holds $\boldsymbol{P}$-a.s.:

$$\begin{aligned} & F(t, W) - F(0, W) \\ &= \int_0^t \partial_0 F(s, W)\,\mathrm{d}s + \sum_{i=1}^m \int_0^t \partial_i F(s, W)\,\mathrm{d}W_i(s) \\ &\quad + \frac{1}{2} \sum_{i,j=1}^m \int_0^t \partial_i \partial_j F(s, W) Q_{ji}(s)\,\mathrm{d}s. \end{aligned} \tag{7}$$

Remark 2. Since F and its derivatives are causal systems, the integrands in (7) can be equivalently written as $\partial_i F(s, \Pi_s W)$, $\partial_i \partial_j F(s, \Pi_s W)$, etc.

Remark 3. Eq. (7) can be written in Stratonovich form as

$$\begin{aligned} & F(t, W) - F(0, W) \\ &= \int_0^t \partial_0 F(s, W)\,\mathrm{d}s + \sum_{i=1}^m \int_0^t \partial_i F(s, W) \circ \mathrm{d}W_i(s). \end{aligned}$$

4. DUPIRE-DIFFERENTIABLE CAUSAL STOCHASTIC SYSTEMS

We now turn our attention back to stochastic systems introduced in Section 2 and to real-valued processes $Y = (Y(t))_{t \in [0,T]}$ satisfying condition (3) for some causal system F. At this point, we impose the additional requirement that F is continuous and has continuous first- and second-order Dupire derivatives $\partial_0 F, \ldots, \partial_m F$ and $\partial_i \partial_j F$ for $1 \le i, j \le m$. Here, some care must be exercised in light of the non-uniqueness issue mentioned at the end of Section 2: While the process Y does not depend on the behavior of F outside the support of $\boldsymbol{P}$, the Dupire derivatives of F do depend on it (indeed, the definition of the vertical derivative involves càdlàg perturbations of input trajectories). Fortunately, our nondegeneracy assumption on the quadratic variation process of W allows us to define the Dupire derivatives of Y *intrinsically* (Cont and Fournié, 2013), as follows.

Let Y be a continuous progressively measurable process on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \boldsymbol{P})$; cf. the discussion at the end of Section 2. We say that Y is *Dupire-differentiable* if there exist progressively measurable processes $Z_0, \ldots, Z_m$ on the same probability space, such that

$$\int_0^t |Z_0(s)|\,\mathrm{d}s + \sum_{i,j=1}^m \int_0^t Z_i(s) Z_j(s) Q_{ij}(s)\,\mathrm{d}s < \infty$$

and

$$Y(t) = Y(0) + \int_0^t Z_0(s)\,\mathrm{d}s + \sum_{i=1}^m \int_0^t Z_i(s) \circ \mathrm{d}W_i(s) \tag{8}$$

for all $t \in [0,T]$ $\boldsymbol{P}$-a.s. We need to show that if $Y \equiv 0$ $\boldsymbol{P}$-a.s., then all the $Z_i \equiv 0$ $\boldsymbol{P}$-a.s. as well. By Itô's lemma,

$$0 = |Y(T)|^2 = \sum_{i,j=1}^m \int_0^T Z_i(t) Z_j(t) Q_{ij}(t)\,\mathrm{d}t.$$

Since $Q(\cdot)$ is a.s. positive definite, it follows that $Z_i = 0$ for all $i = 1, \ldots, m$ a.s., and therefore $Z_0 = 0$ a.s. as well. Hence, if Y is Dupire-differentiable, then the processes $Z_0, \ldots, Z_m$ are a.s. uniquely determined. Moreover, if there exists a causal map F satisfying the Dupire differentiability conditions listed earlier and such that $Y = F(W)$, then it follows from the functional Itô's lemma (Theorem 1) that Y is Dupire-differentiable and

$$\boldsymbol{P}\{Z_i(t) = \partial_i F(t, W), i = 0, \ldots, m \text{ for all } 0 \le t \le T\} = 1.$$

In addition, the above argument shows that, if Y has an alternative representation as $\tilde{F}(W)$ with $\tilde{F} \neq F$, then the Dupire derivatives of F and $\tilde{F}$ computed along the paths of W are almost surely equal. Taking this into account, we can introduce the linear mappings $S_0, \ldots, S_m$ that take any Dupire-differentiable process Y to the respective processes $Z_0, \ldots, Z_m$ in (8).

Definition 3. We say that Y is *continuously Dupire-differentiable* if $S_0 Y, \ldots, S_m Y$ are continuous. For $k \ge 1$, we say that Y is $(k+1)$-*times continuously Dupire-differentiable* if it is Dupire-differentiable and $S_0 Y, \ldots, S_m Y$ are k-times continuously Dupire-differentiable. Finally, if Y is k-times continuously Dupire-differentiable for all $k \ge 0$, then we say that it is *Dupire-smooth.*

We introduce the following notation and definitions for later use: Let M denote the set of all finite tuples, or words, $\boldsymbol{i} = (i_1, \ldots, i_k)$ with $i_j \in \{0, \ldots, m\}$ for $k \ge 0$ ($k = 0$ corresponds to the empty word $\diamond$). For each $\boldsymbol{i} = (i_1, \ldots, i_k) \in \mathsf{M}$ we define the iterated operators

$$S_{\boldsymbol{i}} := S_{i_k} S_{i_{k-1}} \ldots S_{i_1} \tag{9}$$

with $S_\diamond := \mathrm{id}$, as well as the iterated Stratonovich integrals

$$\int_0^t \circ\, \mathrm{d}W_{\boldsymbol{i}} := \int_{\Delta^k[0,t]} \circ\, \mathrm{d}W_{i_k}(t_k) \circ \cdots \circ \mathrm{d}W_{i_2}(t_2) \circ \mathrm{d}W_{i_1}(t_1) \quad (10)$$

with $\mathrm{d}W_0(t) := \mathrm{d}t$ and $\int_0^t \circ\, \mathrm{d}W_\diamond := 1$, where the integration is over the k-dimensional simplex

$$\Delta^k[0,t] := \big\{(t_1,\dots,t_k) \in [0,t]^k : t_k \le \cdots \le t_2 \le t_1\big\}.$$

Finally, we define the linear operator c that takes the process Y to its initial value $Y(0)$, which is a deterministic constant since $W(0) = 0$.

4.1 Examples

Example 1. For the 'memoryless' system $Y(t) = f(t, W(t))$ with f of class $C^{1,2}$, we simply recover the Itô–Stratonovich formula (6): $S_i Y(t) = \partial_i f(t, W(t))$ for $i = 0,\dots,m$.

Example 2. Consider the process Y obtained by passing W through a linear filter:

$$Y(t) = \sum_{i=1}^m \int_0^t h_i(t-s)\, \mathrm{d}W_i(s), \qquad 0 \le t \le T$$

where the h_i's are smooth (analytic or C^∞) functions $[0,T] \to \mathbb{R}$. Then, for $k \ge 0$,

$$S_0^k Y(t) = \sum_{i=1}^m \int_0^t \frac{\partial^k h_i}{\partial t^k}(t-s)\, \mathrm{d}W_i(s)$$

and $S_i S_0^k Y(t) = h_i^{(k)}(0)$ for $i = 1,\dots,m$ corresponding to the words $(0,\dots,0)$ and $(0,\dots,0,i)$, $i \in \{1,\dots,m\}$ of lengths k and $k+1$, respectively. All other $S_{\boldsymbol{i}} Y$ are equal to zero.

Example 3. We consider a particular instance of a finite-dimensional state space realization. Let the following be given: a nonrandom point $x_0 \in \mathbb{R}^n$, $m+1$ smooth vector fields $g_0,\dots,g_m : \mathbb{R}^n \to \mathbb{R}^n$, and a smooth function $h : \mathbb{R}^n \to \mathbb{R}$. We assume that the solution of the Stratonovich integral equation

$$X(t) = x_0 + \int_0^t g_0(X(s))\, \mathrm{d}s + \sum_{i=1}^m \int_0^t g_i(X(s)) \circ \mathrm{d}W_i(s)$$

exists for all $t \in [0,T]$, and take $Y(t) = h(X(t))$. Then it follows from the Itô–Stratonovich formula that the processes $S_i Y(t)$ are given by the Lie derivatives of h along the vector fields g_i:

$$S_i Y(t) = L_{g_i} h(X(t)) := \sum_{j=1}^n \frac{\partial h}{\partial x_j}(X(t)) g_{i,j}(X(t)).$$

Since $g_0,\dots,g_m$ and h are smooth, the process Y is Dupire-smooth, and the action of $S_{\boldsymbol{i}}$ for $\boldsymbol{i} = (i_1,\dots,i_k) \in \mathsf{M}$ corresponds to taking iterated Lie derivatives: $S_{\boldsymbol{i}} Y(t) = L_{g_{i_k}} L_{g_{i_{k-1}}} \dots L_{g_{i_1}} h(X(t))$. The pair (X,Y) is a state space realization of Y.

4.2 Hijab's formulation as a special case

In the work of Hijab (1983a,b), the driving process W is a one-dimensional standard Brownian motion, and the process Y is *Itô-differentiable* if there exist two progressively measurable processes $\tilde{Z}_0, \tilde{Z}_1$, such that

$$Y(t) = Y(0) + \int_0^t \tilde{Z}_0(s)\, \mathrm{d}s + \int_0^t \tilde{Z}_1(s)\, \mathrm{d}W(s)$$

for all $t \in [0,T]$ a.s. , where the integral is Itô. These processes are a.s. uniquely determined by Y, which can be proved using the same argument as the one used to prove the a.s. uniqueness of $Z_0,\dots,Z_m$. Hijab then defines the linear operators A and B that send $Y \mapsto \tilde{Z}_0$ and $Y \mapsto \tilde{Z}_1$, respectively. While he does not give any operational characterization of A, B, it follows readily from the relation between the Itô and the Stratonovich integrals that

$$A = S_0 + \frac{1}{2} S_1^2, \qquad B = S_1$$

(Hijab's X_0 and X_1 operators correspond to our definitions of S_0, S_1). Thus, we see that Hijab's notion of Itô-differentiability and our notion of Dupire-differentiability coincide. In retrospect, it is easy to see why Hijab did not relate his construction to any explicit definition of a causal derivative: In order to properly define them, we need to consider càdlàg perturbations of the sample paths of W, which in turn requires the use of the Skorohod space $D([0,T],\mathbb{R})$. By contrast, Hijab takes $C([0,T],\mathbb{R})$ as the sample space.

5. REALIZATION THEORY FOR DUPIRE-SMOOTH PROCESSES

Example 3 from the preceding section provides a blueprint for a stochastic realization theory for Dupire-smooth processes that closely parallels the realization theory for deterministic systems based on Chen–Fliess functional expansions (Isidori, 1995, Ch. 3). In this section, we outline this approach, building on the ideas of Hijab; our treatment here is primarily formal, and we leave the detailed analysis of convergence, truncation errors, etc. for future work.

Remark 4. Chen–Fliess representations of stochastic processes have been considered by Sussmann (1988) for smooth functions of Itô diffusion processes, by Litterer and Oberhauser (2014) for Dupire-differentiable functionals of diffusion processes, and by Dupire and Tissot-Daguette (2022) in the general setting of functional Itô calculus. However, none of these works are concerned with the questions of realization.

Let Y be a Dupire-smooth process. Then, using the definitions of $S_\diamond, S_0, \dots, S_m$ and c, we can rewrite (8) in the following way:

$$Y(t) = c(S_\diamond Y) + \sum_{i_1=0}^m \int_0^t S_{i_1} Y(t_1) \circ \mathrm{d}W_{i_1}(t_1) \qquad (11)$$

(recall that $\mathrm{d}W_0 = \mathrm{d}t$). Since $S_{i_1} Y$ is continuously Dupire-differentiable, we have

$$S_{i_1} Y(t_1) = c(S_{i_1} Y) + \sum_{i_2=0}^m \int_0^{t_1} S_{(i_1,i_2)} Y(t_2) \circ \mathrm{d}W_{i_2}(t_2), \qquad (12)$$

where $S_{(i_1,i_2)} = S_{i_2} S_{i_1}$, cf. (9). Substituting (12) into (11) gives

$$Y(t) = c(S_\diamond Y) + \sum_{i_1=0}^m c(S_{i_1} Y) \int_0^t \circ\, \mathrm{d}W_{i_1} + \sum_{i_1,i_2=0}^m \int_0^t \int_0^{t_1} S_{(i_1,i_2)} Y(t_2) \circ \mathrm{d}W_{i_2}(t_2) \circ \mathrm{d}W_{i_1}(t_1).$$

Continuing inductively, we obtain the following formal infinite series expansion of the Chen–Fliess type:

$$Y(t) = \sum_{\boldsymbol{i} \in \mathsf{M}} c(S_{\boldsymbol{i}} Y) \int_0^t \circ \, \mathrm{d}W_{\boldsymbol{i}}, \qquad 0 \le t \le T. \qquad (13)$$

Observe that the coefficients $c(S_{\boldsymbol{i}}Y)$ are deterministic constants (iterated Dupire derivatives of Y at 0), and all the randomness has been pushed into the iterated Stratonovich integrals of W.

We can now state the stochastic realization problem in the following way: Given a Dupire-smooth process Y, find an integer n, a point $x_0 \in \mathbb{R}^n$, $m+1$ smooth vector fields $g_0, \dots, g_m$ on $\mathbb{R}^n$, and a smooth function $h : \mathbb{R}^n \to \mathbb{R}$ defined on a neighborhood of x_0, such that, for every $\boldsymbol{i} = (i_1, \dots, i_k) \in \mathsf{M}$,

$$c(S_{\boldsymbol{i}}Y) = L_{g_{i_k}} L_{g_{i_{k-1}}} \dots L_{g_{i_1}} h(x_0). \qquad (14)$$

At this point, we can make use of the theory of formal power series, exactly as in the setting of deterministic realization theory (Fliess, 1981). Let $\mathcal{Z} = \{z_0, \dots, z_m\}$ be a set of $m+1$ noncommuting indeterminates. With each word $\boldsymbol{i} = (i_1, \dots, i_k) \in \mathsf{M}$, we associate the formal monomial $z_{\boldsymbol{i}} := z_{i_1} \dots z_{i_k}$; the empty word $\diamond$ will be associated with the constant term $z_\diamond = 1$. A *formal power series* in $\mathcal{Z}$ with real coefficients is an expression of the form $R = \sum_{\boldsymbol{i} \in \mathsf{M}} R(\boldsymbol{i}) z_{\boldsymbol{i}}$, where $R(\boldsymbol{i})$ take real values. The set of all such formal power series, denoted by $\mathbb{R}\langle\langle \mathcal{Z} \rangle\rangle$, is a noncommutative associative $\mathbb{R}$-algebra, with $\alpha R + \beta S$ and RS defined for $R, S \in \mathbb{R}\langle\langle \mathcal{Z} \rangle\rangle$ and $\alpha, \beta \in \mathbb{R}$ by $(\alpha R + \beta S)(\boldsymbol{i}) := \alpha R(\boldsymbol{i}) + \beta S(\boldsymbol{i})$ and

$$RS(\boldsymbol{i}) := \sum_{\substack{\boldsymbol{i}', \boldsymbol{i}'' \in \mathsf{M} \\ \boldsymbol{i} = \boldsymbol{i}'\boldsymbol{i}''}} R(\boldsymbol{i}') S(\boldsymbol{i}''),$$

where $\boldsymbol{i}'\boldsymbol{i}''$ denotes the concatenation of $\boldsymbol{i}' = (i'_1, \dots, i'_{k'})$ and $\boldsymbol{i}'' = (i''_1, \dots, i''_{k''})$: $\boldsymbol{i}'\boldsymbol{i}'' = (i'_1, \dots, i'_{k'}, i''_1, \dots, i''_{k''})$. A *formal polynomial* is an element $P \in \mathbb{R}\langle\langle \mathcal{Z} \rangle\rangle$, for which $P(\boldsymbol{i}) = 0$ for all but finitely many $\boldsymbol{i} \in \mathsf{M}$. The space of all formal polynomials (also an algebra) will be denoted by $\mathbb{R}\langle \mathcal{Z} \rangle$. Moreover, it can be endowed with the structure of a *Lie algebra* with the Lie bracket of two polynomials $P, Q \in \mathbb{R}\langle \mathcal{Z} \rangle$ given by $[P, Q] := PQ - QP$. We will denote by $\mathcal{L}(\mathcal{Z})$ the smallest subspace of $\mathbb{R}\langle \mathcal{Z} \rangle$ that contains the monomials $z_0, \dots, z_m$ and is closed under Lie bracketing with $z_0, \dots, z_m$. The elements of $\mathcal{L}(\mathcal{Z})$ are called (formal) *Lie polynomials.*

Now, following Hijab (1983a), we let $\mathcal{V}_Y$ denote the smallest vector space of processes containing Y and closed under all $S_{\boldsymbol{i}}$. There is a natural $\mathbb{R}$-linear morphism $\mu : \mathbb{R}\langle \mathcal{Z} \rangle \to \mathcal{V}_Y$, defined by its action on monomials $\mu : z_{\boldsymbol{i}} \mapsto S_{\boldsymbol{i}}Y$, $\boldsymbol{i} \in \mathsf{M}$ and extended to all of $\mathbb{R}\langle \mathcal{Z} \rangle$ by linearity. Using this, we can associate to Y a linear mapping $\mathsf{F}_Y : \mathbb{R}\langle \mathcal{Z} \rangle \to \mathbb{R}\langle\langle \mathcal{Z} \rangle\rangle$ defined by its action on monomials as $\mathsf{F}_Y(z_{\boldsymbol{i}}) := \sum_{\boldsymbol{i}' \in \mathsf{M}} c \circ \mu(z_{\boldsymbol{i}\boldsymbol{i}'}) z_{\boldsymbol{i}'}$ and, again, extended to all of $\mathbb{R}\langle \mathcal{Z} \rangle$ by linearity.

Definition 4. The *Hankel rank* of a Dupire-smooth process Y is the rank of the mapping F_Y: $\rho_H(Y) := \dim \mathsf{F}_Y(\mathbb{R}\langle \mathcal{Z} \rangle)$. The *Lie rank* of Y is the rank of the restriction of F_Y to Lie polynomials: $\rho_L(Y) := \dim \mathsf{F}_Y(\mathcal{L}(\mathcal{Z}))$.

The inequality $\rho_L(Y) \le \rho_H(Y)$ is immediate; moreover, in full analogy with the deterministic case, it is possible for the Lie rank $\rho_L(Y)$ to be finite and for the Hankel rank $\rho_H(Y)$ to be infinite. Again, the key ideas were already present in the work of Hijab in the special case of processes driven by Brownian motion, although Hijab only defined the Lie rank of a process. Once the process Y is represented using the Chen–Fliess series (13), the machinery of formal power series can be applied in a unified manner. In particular, we have the following result, essentially due to Hijab (1983a):

Theorem 2. Let Y admit a state space realization specified by $(n, x_0, g_0, \dots, g_m, h)$, i.e., $Y(t) = h(X(t))$ for all $t \in [0, T]$, where $(X(t))_{t \in [0,T]}$ is an n-dimensional continuous semimartingale that solves the Stratonovich integral equation

$$X(t) = x_0 + \int_0^t g_0(X(s)) \, \mathrm{d}s + \sum_{i=1}^m \int_0^t g_i(X(s)) \circ \mathrm{d}W_i(s)$$

for $0 \le t \le T$. Then $\rho_L(Y) \le n$.

An equivalent representation of the mapping F_Y associated to a Dupire-smooth process Y is given by the (infinite) *Hankel matrix* H_Y with entries indexed by words in M:

$$\mathsf{H}_Y(\boldsymbol{i}, \boldsymbol{i}') := c \circ \mu(z_{\boldsymbol{i}\boldsymbol{i}'}) = c(S_{\boldsymbol{i}\boldsymbol{i}'}Y), \qquad \boldsymbol{i}, \boldsymbol{i}' \in \mathsf{M}.$$

The Hankel rank $\rho_H(Y)$ is then the rank of the Hankel matrix. Thus, for the linear filtering situation considered in Example 2, the only nonzero entries of the Hankel matrix are given by $\mathsf{H}_Y(\boldsymbol{i}, \boldsymbol{i}') = \sum_{j=1}^m h_j^{(k)}(0)$ if $\boldsymbol{i}$ and $\boldsymbol{i}'$ are both strings of 0's and their concatenation has length $k \ge 0$, and $\mathsf{H}_Y(\boldsymbol{i}, \boldsymbol{i}') = h_i^{(k)}(0)$ if $\boldsymbol{i}\boldsymbol{i}' = (0, \dots, 0, i)$ consisting of k 0's followed by a single $i \in \{1, \dots, m\}$. It then follows from the classical realization theory of linear time-invariant systems that $\rho_H(Y) = \operatorname{rank} \mathsf{H}_Y \le n$ if and only if there exist matrices $C \in \mathbb{R}^{1 \times n}$, $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, such that

$$h(t) = \big(h_1(t), \dots, h_m(t)\big) = Ce^{At}B,$$

which gives rise to a linear state space realization

$$X(t) = \int_0^t AX(s) \, \mathrm{d}s + \int_0^t B \, \mathrm{d}W(s), \quad Y(t) = CX(t).$$

We can now proceed to address the question of nonlinear realization posed in the beginning of this section. Given a Dupire-smooth process Y with finite Hankel rank $\rho_H(Y) = n$, we can follow the approach of Fliess (cf. Theorem 3.4.3 in Isidori (1995)) to construct a *bilinear* realization, i.e., a tuple $(n, x_0, A_0, \dots, A_m, C)$ with $x_0 \in \mathbb{R}^n$, $A_0, \dots, A_m \in \mathbb{R}^{n \times n}$, and $C \in \mathbb{R}^{1 \times n}$, such that

$$c(S_{\boldsymbol{i}}Y) = CA_{i_k} \dots A_{i_1} x_0, \quad \text{for all } \boldsymbol{i} = (i_1, \dots, i_k) \in \mathsf{M}.$$

For a process Y with finite Lie rank $\rho_L(Y) = n$, we would need a convergence condition of the form

$$c(S_{\boldsymbol{i}}Y) \le Ck!r^k, \qquad \text{for all } \boldsymbol{i} = (i_1, \dots, i_k) \in \mathsf{M}$$

for some constants $C, r > 0$. Then, just as in Reutenauer (1986), we could establish the existence of a tuple $(n, x_0, g_0, \dots, g_m, h)$, where $g_0, \dots, g_m$ are analytic vector fields and h is an analytic real-valued function on some neighborhood of x_0, such that (14) holds. However, in order to deduce from the above results the corresponding probabilistic constructions, i.e., either a bilinear state space realization of the form

$$X(t) = x_0 + \sum_{i=1}^m \int_0^t A_i X(s) \circ \mathrm{d}W_i(s), \quad Y(t) = CX(t)$$

when $\rho_H(Y) = n$, or a nonlinear analytic one of the type discussed in Example 3, we would need to address the questions of convergence of the Chen–Fliess series (13), either in a suitable L^p sense, as in Sussmann (1988) or Litterer and Oberhauser (2014), or pathwise in the sense of Föllmer (1981), as in Dupire and Tissot-Daguette (2022). These questions can be rather delicate (see, e.g., the discussion in Sussmann (1976) concerning noise-like "generalized inputs" in the context of bilinear systems), and we leave them for future work.

6. CONCLUSION

Building on the pioneering work of Hijab (1983a,b), we have presented an approach to nonlinear stochastic realization for a class of stochastic processes that can be represented as the outputs of a causal system driven by a continuous semimartingale. The key concept here is that of causal derivative of a process, originating in the functional Itô calculus of Dupire (2009). We have shown that, formally, the questions of existence of finite-dimensional state space realizations can be phrased in terms of the stochastic analogues of the Lie and the Hankel rank from the realization theory for deterministic systems following the ideas of Fliess (1981). Some additional examples and discussion had to be omitted due to space limitations, and can be found in Veeravalli and Raginsky (2024). There are several interesting directions for further research, such as the issues of convergence and an extension to processes driven by general càdlàg semimartingales (such as counting processes).

REFERENCES

Billingsley, P. (1999). *Convergence of Probability Measures*. Wiley, 2nd edition.

Brockett, R.W. (1980). Remarks on finite dimensional nonlinear estimation. *Astérisque*, 75-76, 47–55.

Cont, R. and Fournié, D.A. (2010). Change of variable formulas for non-anticipative functionals on path space. *Journal of Functional Analysis*, 259(4), 1043–1072.

Cont, R. and Fournié, D.A. (2013). Functional Itô calculus and stochastic integral representation of martingales. *The Annals of Probability*, 41(1), 109–133.

Dupire, B. (2009). Functional Itô calculus. Bloomberg Portfolio Research Paper 2009-04.

Dupire, B. and Tissot-Daguette, V. (2022). Functional expansions. arXiv.org preprint. URL `https://arxiv.org/abs/2212.13628`.

Fliess, M. (1981). Fonctionnelles causales non linéaires et indéterminées non commutatives. *Bulletin de la Société Mathématique de France*, 109, 3–40.

Fliess, M. (1983). On the concept of derivatives and Taylor expansions for nonlinear input-output systems. In *The 22nd IEEE Conference on Decision and Control*, 643–646.

Föllmer, H. (1981). Calcul d'Itô sans probabilités. *Séminaire de probabilités de Strasbourg*, 15, 143–150.

Georgiou, T.T. and Lindquist, A. (2013). The separation principle in stochastic control, redux. *IEEE Transactions on Automatic Control*, 58(10), 2481–2494.

Hazewinkel, M. and Marcus, S.I. (1981). Some results and speculations on the role of lie algebras in filtering. In M. Hazewinkel and J.C. Willems (eds.), *Stochastic Systems: The Mathematics of Filtering and Identification and Applications*, 591–604. Springer Netherlands.

Hijab, O. (1983a). Finite dimensional causal functionals of brownian motion. In R.S. Bucy and J.M.F. Moura (eds.), *Nonlinear Stochastic Problems*, 425–435. Springer Netherlands.

Hijab, O. (1983b). A realization theory for nonlinear stochastic systems. In *The 22nd IEEE Conference on Decision and Control*, 898–903.

Isidori, A. (1995). *Nonlinear Control Systems*. Springer London.

Knight, F.B. (1975). A predictive view of continuous time processes. *The Annals of Probability*, 3(4), 573–596.

Lindquist, A., Mitter, S., and Picci, G. (1982). Toward a theory of nonlinear stochastic realization. In D. Hinrichsen and A. Isidori (eds.), *Feedback Control of Linear and Nonlinear Systems*, 175–189. Springer.

Lindquist, A. and Picci, G. (2015). *Linear Stochastic Systems: A Geometric Approach to Modeling, Estimation and Identification*. Springer.

Litterer, C. and Oberhauser, H. (2014). On a Chen–Fliess approximation for diffusion functionals. *Monatshefte für Mathematik*, 175(4), 577–593.

Mitter, S.K. (1979). On the analogy between mathematical problems of non-linear filtering and quantum physics. *Ricerche di Automatica*, 10(2), 163–216.

Protter, P.E. (2005). *Stochastic Integration and Differential Equations*. Springer Berlin Heidelberg.

Reutenauer, C. (1986). The local realization of generating series of finite Lie rank. In M. Fliess and M. Hazewinkel (eds.), *Algebraic and Geometric Methods in Nonlinear Control Theory*, 33–43. Springer Netherlands.

Stroock, D.W. (1987). Homogeneous chaos revisited. *Séminaire de probabilités de Strasbourg*, 21, 1–7.

Sussmann, H.J. (1976). Semigroup representations, bilinear approximation of input-output maps, and generalized inputs. In G. Marchesini and S.K. Mitter (eds.), *Mathematical Systems Theory*, 172–191. Springer Berlin Heidelberg.

Sussmann, H.J. (1981). Rigorous results on the cubic sensor problem. In M. Hazewinkel and J.C. Willems (eds.), *Stochastic Systems: The Mathematics of Filtering and Identification and Applications*, 637–648. Springer Netherlands.

Sussmann, H.J. (1988). Product expansions of exponential Lie series and the discretization of stochastic differential equations. In W. Fleming and P.L. Lions (eds.), *Stochastic Differential Systems, Stochastic Control Theory and Applications*, 563–582. Springer New York.

Taylor, T.J. and Pavon, M. (1988). A solution of the nonlinear stochastic realization problem. *Systems Control Letters*, 11(2), 117–121.

Taylor, T.J. and Pavon, M. (1989). On the nonlinear stochastic realization problem. *Stochastics and Stochastic Reports*, 26(2), 65–79.

Veeravalli, T. and Raginsky, M. (2024). Revisiting stochastic realization theory using functional Itô calculus. arXiv.org preprint. URL `https://arxiv.org/abs/2402.10157`.

Willems, J.C. and van Schuppen, J.H. (1980). Stochastic systems and the problem of state space realization. In C.J. Byrnes and C.F. Martin (eds.), *Geometrical Methods for the Theory of Linear Systems*, 283–313. Springer Netherlands.