ROUGH MCKEAN–VLASOV DYNAMICS FOR ROBUST ENSEMBLE KALMAN FILTERING

BY MICHELE COGHI[1,a], TORSTEIN NILSSEN[2,b], NIKOLAS NÜSKEN[3,c] AND SEBASTIAN REICH[4,d]

[1]*Department of Mathematics, Università di Trento,* [a]*michele.coghi@unitn.it*

[2]*Institute of Mathematics, University of Agder,* [b]*torstein.nilssen@uia.no*

[3]*Department of Mathematics, King's College London,* [c]*nikolas.nusken@kcl.ac.uk*

[4]*Institute of Mathematics, Universität Potsdam,* [d]*sebastian.reich@uni-potsdam.de*

Motivated by the challenge of incorporating data into misspecified and multiscale dynamical models, we study a McKean–Vlasov equation that contains the data stream as a common driving rough path. This setting allows us to prove well-posedness as well as continuity with respect to the driver in an appropriate rough-path topology. The latter property is key in our subsequent development of a robust data assimilation methodology: We establish propagation of chaos for the associated interacting particle system, which in turn is suggestive of a numerical scheme that can be viewed as an extension of the ensemble Kalman filter to a rough-path framework. Finally, we discuss a data-driven method based on subsampling to construct suitable rough path lifts and demonstrate the robustness of our scheme in a number of numerical experiments related to parameter estimation problems in multiscale contexts.

CONTENTS

1. Introduction. Combining mathematical descriptions of reality with observational data is a key task in economics, science and engineering. In typical applications (such as meteorology [67] or molecular dynamics [79]) there is a hierarchy of models available, ranging from the highly accurate but conceptually and computationally demanding to the approximate but readily interpretable and scalable. Naturally, real-world data is almost always complex, multiscale, and increasingly high-dimensional, whereas the corresponding mathematical abstractions are often preferred to be simple and of comparatively low resolution. The discrepancy between intricate data and reduced-order models poses a significant challenge for their simultaneous treatment, and the failure of some standard statistical approaches in scenarios of this type is well documented [2, 96, 114].

Robustness to model misspecification and perturbation of the data. is a central concept in the design of statistical methodology capable of bridging scales: Consider a statistical model $\mathcal{M}_0$—to be thought of as simple —generating (time-dependent) data $(Y_t^0)_{t\geq 0}$ and a corresponding algorithmic procedure Φ producing the output $\Phi((Y_t^0)_{t\geq 0})$. We expect that Φ deals adequately with complex data in the case when it is continuous in an appropriate sense. Indeed, $\mathcal{M}_0$ might be a simplified description of an underlying model family $(\mathcal{M}_\varepsilon)_{\varepsilon\geq 0}$, where the formal limit $\lim_{\varepsilon\to 0}\mathcal{M}_\varepsilon = \mathcal{M}_0$ encapsulates the passage from a complex to a reduced description. The output $\Phi((Y_t^\varepsilon)_{t\geq 0})$ on "real-world" data $(Y_t^\varepsilon)_{t\geq 0}$ is then close to $\Phi((Y_t^0)_{t\geq 0})$ by continuity even though Φ has been contructed on the basis of $\mathcal{M}_0$.

Unfortunately, in various contexts (for instance, in parameter estimation for diffusions [40] and stochastic filtering [10]) the map Φ is given in terms of stochastic integrals against $(Y_t^0)_{t\geq 0}$ which are well known to be discontinuous with respect to standard topologies [53, 55]. The theory of *rough paths* provides a principled route towards constructing continuous modifications Φ' of Φ (employed, for instance, in [29, 40]) by replacing stochastic integrals in terms of rough integrals defined on appropriately lifted paths $(Y, \mathbb{Y})$. Although the difficulty of obtaining or imposing the additional information $\mathbb{Y}$ is well known (see, however, [9, 49, 80] and Section 6 of this article), recent works have demonstrated the potential of including path signatures into data-driven methods to compress information efficiently or unveil multiscale structure [20, 85]. Moreover, the rough path perspective provides refined insights into misspecification of diffusion models: Indeed, two paths may be "similar in a classical sense" (for instance, in supremum norm) in spite of their associated lifts exhibiting a large rough path distance, in turn leading to inaccurate inferences. In Section 6 we discuss how those discrepancies can be compensated in a systematic and data-driven manner.

In this paper, we follow the rough path paradigm just described and develop a robust version of the *ensemble Kalman filter* (EnKF) [15, 48], drawing on a reformulation of the Kushner–Stratonovich SPDE for stochastic filtering [10] in terms of McKean–Vlasov dynamics (known as the *feedback particle filter* [92, 105, 109]). The EnKF is a versatile approximate procedure for Bayesian inference that is observed to perform particularly well in high-dimensional settings [101] and has recently been applied to problems in machine learning [59, 98]. Our analysis includes the case when the model and observation noises are correlated as this is precisely the setting in which significant obstacles in the construction of robust

filters are known to occur [29]. Furthermore, as shown in [88] and recalled in Appendix B, stochastic filtering with correlated noise is a natural generalisation of maximum likelihood parameter estimation for diffusions in the context of inexact measurements. In this setting, we discuss the relationship between subsampling-based approaches to multiscale parameter estimation [96] to the task of estimating the Lévy area in a rough paths approach.

The construction put forward in this paper requires the formulation of McKean–Vlasov equations in a rough path setting (as studied first in [18] and more recently in the twin papers [7, 8], both considering the dynamics to be driven by a random rough path). The approach introduced in [26] is more suitable for our needs as it allows a clear separation between independent Brownian motions and a common deterministic noise. However, the assumptions in [26] do not cover unbounded coefficients, while the common noise coefficient may depend on both the state of the solution and its law. In equation (11) considered below, the coefficient P only depends on the law of $\widehat{X}$, which simplifies the problem to some extent, but also allows us to use a different approach, where we treat the stochastic and the rough integrals in two separated steps. As will become clear from the proofs in Section 4.4, there is no need to create a joint rough path (or rough driver). All previous works on rough McKean–Vlasov equations deal with bounded globally Lipschitz-continuous (actually smoother) coefficients. These results cannot be applied here, as the coefficient P has linear growth in the measure of the solution and is locally-Lipschitz with Lipschitz constant depending on the moments of the solution.

1.1. *Setting and main results.*

In this section we specify the exact setting and present our main results. Throughout the paper we fix a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ satisfying the usual conditions and consider the following filtering model:

$$\text{(1a)} \qquad \mathrm{d}X_t = f(X_t)\,\mathrm{d}t + G^{1/2}\,\mathrm{d}W_t, \qquad X_0 \sim \pi_0,$$

$$\text{(1b)} \qquad \mathrm{d}Y_t = h(X_t)\,\mathrm{d}t + U\,\mathrm{d}W_t + R^{1/2}\,\mathrm{d}V_t, \qquad Y_0 = 0.$$

In the above display, (1a) represents the (hidden) signal, whereas (1b) specifies the available observations. Note straight away that the signal noise $G^{1/2}\,\mathrm{d}W_t$ and the observation noise $U\,\mathrm{d}W_t + R^{1/2}\,\mathrm{d}V_t$ may be correlated. We assume that the signal is D-dimensional, that is, $X_t \in \mathbb{R}^D$, and that the observations are d-dimensional, $Y_t \in \mathbb{R}^d$. In applications it is often the case that $D \gg d$. The maps $f : \mathbb{R}^D \to \mathbb{R}^D$ and $h : \mathbb{R}^D \to \mathbb{R}^d$ are assumed to be sufficiently regular (see below). We assume $G \in \mathbb{R}^{D \times D}$, $R \in \mathbb{R}^{d \times d}$ to be symmetric nonnegative definite and $U \in \mathbb{R}^{d \times D}$. Furthermore, $(W_t)_{t\geq 0}$ and $(V_t)_{t\geq 0}$ denote independent D- and d- dimensional standard Brownian motions, respectively. Lastly, we assume that the observation covariance

$$\text{(2)} \qquad C = UU^\top + R \in \mathbb{R}^{d \times d}$$

is strictly positive definite. Defining the observation σ-algebras

$$\text{(3)} \qquad \mathcal{Y}_t = \sigma\big\{(Y_s)_{0\leq s\leq t}\big\} \vee \mathcal{N}, \quad t \geq 0,$$

where $\mathcal{N} \subset \mathcal{F}$ is the collection of $\mathbb{P}$-null sets, our objective is to compute or approximate the filtering measures

$$\text{(4)} \qquad \pi_t[\phi] = \mathbb{E}\big[\phi(X_t)|\mathcal{Y}_t\big],$$

for bounded and measurable test functions ϕ; we refer to [10] for technical details.

As a step towards tractable numerical approximations of (4), we follow [88] and introduce the McKean–Vlasov equation

(5)
$$\begin{cases} d\widehat{X}_t = f(\widehat{X}_t)\, dt + G^{1/2}\, d\widehat{W}_t + K_t(\widehat{X}_t)C^{-1} \circ dI_t + \Xi_t(\widehat{X}_t)\, dt, \\ dI_t = dY_t - \left(h(\widehat{X}_t)\, dt + U\, d\widehat{W}_t + R^{1/2}\, d\widehat{V}_t\right), \end{cases}$$

where $(\widehat{W}_t)_{t\geq 0}$ (resp. $(\widehat{V}_t)_{t\geq 0}$) is a given D-dimensional (resp. d-dimensional) Brownian motion, both independent from $(W_t)_{t\geq 0}$ and $(V_t)_{t\geq 0}$. Denoting by $\widehat{\pi} := \mathcal{L}(\widehat{X} \mid \mathcal{Y})$ the conditional law with respect to the common noise Y of solutions $\widehat{X}$ to (5), the coefficients $K_t : \mathbb{R}^D \to \mathbb{R}^{D\times d}$ and $\Xi_t : \mathbb{R}^D \to \mathbb{R}^D$ are required to solve the following partial differential equations:

(6)
$$\mathrm{div}\big(\widehat{\pi}_t(K_t - BC)\big) = -\widehat{\pi}_t\big(h - \widehat{\pi}_t[h]\big),$$

and

(7)
$$\mathrm{div}(\widehat{\pi}_t \Xi_t) = \frac{1}{2}\widehat{\pi}_t\big(\mathrm{Trace}(K_t C^{-1} Dh) - \widehat{\pi}_t[\mathrm{Trace}(K_t C^{-1} Dh)]\big),$$

for $t \geq 0$ and $\mathbb{P}$-almost surely, where $B = G^{1/2}U^\top C^{-1}$. Here div stands for the divergence operator and Dh is the Jacobian matrix of the function h. Moreover, the common noise Y coincides with the observation process (1b) and the integral in $\circ dY$ is understood in the sense of Stratonovich. We would like to remark that in related works the Stratonovich correction is applied to the spatial component of $(t, x) \mapsto K_t(x)$ only, see [99], Remark 4.4, and [92], Section 3, for a discussion. Here, Stratonvich integration is understood in the more usual sense with the semimartingale $t \mapsto K_t(\widehat{X}_t)$ as the integrand (i.e., the Stratonovich interpretation is employed in both t and x, in the terminology of [92]).

The construction of the system (5)–(7) as well as existence of solutions will be detailed in Section 2.4. The filtering problem and the McKean–Vlasov equation (5) are related in the following way.

PROPOSITION (Formal, see Proposition 2.1). *If π_t admits a density and the solution to (5) is unique, then $\pi_t = \widehat{\pi}_t$, $\mathbb{P}$-a.s. for every $t \geq 0$.*

Reformulations of the filtering problem in terms of McKean–Vlasov dynamics similar to (5)–(7) have been introduced in [31, 109], further analysed in [92], and are commonly referred to as *feedback particle filters*; the formulation (5)–(7) combines Stratonovich integration (as in [92]) with a stochastic innovation term dI_t (as in [99], Section 4.2) so as to allow for a transition to geometric rough paths in the setting of correlated noises (as in [88]).

One of the practical challenges posed by the system of equations (5)–(7) is solving the PDEs (6) and (7) in K and Ξ for a given measure $\widehat{\pi}$. As has been shown in [105] for a similar system of equations, replacing K and Ξ by their best constant-in-space approximations in least-square sense recovers a version of the ensemble Kalman filter dynamics [15]. We follow this approach (see Lemma 2.2 below) and replace the coefficient $K_t C^{-1}$ by $P : \mathcal{P}_1(\mathbb{R}^D) \to \mathbb{R}^{D\times d}$, explicitly defined as

(8) $$P(\pi) := \mathrm{Cov}_\pi(x, h)C^{-1} + B := \pi\big[x(h - \pi[h])^\top\big]C^{-1} + B, \quad \pi \in \mathcal{P}_1(\mathbb{R}^D),$$

where, for $\rho \geq 1$, $\mathcal{P}_\rho(\mathbb{R}^D)$ refers to the set of probability measures on $\mathbb{R}^D$ with finite ρth moment. Similarly, we replace Ξ by $\Gamma : \mathcal{P}_1(\mathbb{R}^D) \to \mathbb{R}^D$, defined as

(9) $$\Gamma^\gamma(\pi) = -\frac{1}{2}\mathrm{Trace}\big(P(\pi)\pi\big[x^\gamma(Dh - \pi[Dh])\big]\big), \quad 1 \leq \gamma \leq D, \pi \in \mathcal{P}_1(\mathbb{R}^D),$$

which can be interpreted as an Itô–Stratonovich correction term.[1] Thus, we obtain the system

$$(10) \quad \begin{cases} d\widehat{X}_t = f(\widehat{X}_t)\,dt + G^{1/2}\,d\widehat{W}_t + P(\widehat{\pi}_t) \circ dI_t + \Gamma(\widehat{\pi}_t)\,dt, \\ dI_t = d\mathcal{Y}_t - \big(h(\widehat{X}_t)\,dt + U\,d\widehat{W}_t + R^{1/2}\,d\widehat{V}_t\big), \end{cases}$$

where $\widehat{\pi}_t = \mathcal{L}(\widehat{X}_t|\mathcal{Y}_t)$. This equation is well posed according to Lemma 4.18 below.[2]

In order to construct a robust filter, we replace the common noise in equation (10) by a deterministic rough path $\mathbf{Y} \in \mathscr{C}^\alpha([0, T]; \mathbb{R}^d)$ with regularity $\frac{1}{3} < \alpha \le \frac{1}{2}$. As explained in the Introduction, the rationale is that the solutions to rough differential equations are expected to be continuous in the rough path driver (see [53]), in contrast to the Itô solution map associated to stochastic differential equations. Moreover, using a deterministic path is natural since we are conditioning on the observation, that we can assume to be given and deterministic.

Applying the modifications from the preceding two paragraphs to equation (5) we obtain the system

$$(11) \quad \begin{cases} d\widehat{X}_t = f(\widehat{X}_t)\,dt + G^{1/2}\,d\widehat{W}_t + P(\widehat{\pi}_t)\,dI_t + \Gamma(\widehat{\pi}_t)\,dt, \\ dI_t = d\mathbf{Y}_t - \big(h(\widehat{X}_t)\,dt + U\,d\widehat{W}_t + R^{1/2}\,d\widehat{V}_t\big). \end{cases}$$

It might appear that there is inconsistency between the Itô formulation in equation (11) and the Stratonovich formulation in equation (10). However, one should notice that the Itô and Stratonovich formulation of the integrals with respect to $\widehat{W}$ and $\widehat{V}$ in equation (10) coincide because the integrand is a semimartingale independent of the integrator. Moreover, the integral $d\mathbf{Y}$ in equation (11) is a general rough path integral that corresponds to different stochastic integrals depending on the chosen lift for Y. In particular the case when the Stratonovich lift is chosen corresponds to the integral $\circ dY$ in equation (10). From Lemma 4.13 below, the path $P(\widehat{\pi}.)$ is controlled by $\mathbf{Y}$ with Gubinelli derivative ([53], Definition 4.6),

$$(12) \quad P(\widehat{\pi}_s)' = P(\widehat{\pi}_s)^\top \widehat{\pi}_s[(x - \pi[x])Dh^\top]C^{-1} = P(\widehat{\pi}_s)^\top \mathrm{Cov}_{\widehat{\pi}}(x, Dh)C^{-1},$$

so that the rough integral in equation (11) will make sense (see Section 3.3 below for an overview on controlled rough paths). Moreover, when Y is a semimartingale with covariance $\mathrm{Cov}(Y_t, Y_s) = C(t - s)$, for $s \le t$, the correction between the Stratonovich and Itô rough path lift is given by $-\frac{1}{2}\mathrm{Trace}(P(\widehat{\pi}_t)'C)\,dt$, which corresponds to $\Gamma(\widehat{\pi})\,dt$, see Appendix C.

Our main result is the following well-posedness and stability theorem for (11).

THEOREM 1.1. *Let $1/3 < \alpha < 1/2$ and $\widehat{X}_0 \in L^\rho(\Omega, \mathbb{R}^D)$ with $\rho > 2/(1 - 2\alpha)$. Assume $h \in C_b^2(\mathbb{R}^D, \mathbb{R}^d)$ and that $f : \mathbb{R}^D \to \mathbb{R}^D$ is bounded and Lipschitz-continuous. Then equation (11) admits a unique solution. Moreover, the map $\mathscr{C}^\alpha([0, T]; \mathbb{R}^d) \ni \mathbf{Y} \mapsto \widehat{\pi} \in \mathcal{P}_\rho(C([0, T]; \mathbb{R}^D))$ is continuous. Moreover, if $\mathbf{Y}$ is the Stratonovich lift of Y in (10), then the solutions to (10) and (11) coincide, $\mathbb{P}$-almost surely.*

The proof is a consequence of the results in Section 4.4 and can be found at the end of that section. Associated to the McKean–Vlasov equation (11) we also study the following system of mean-field interacting particles:

$$(13a) \quad d\widehat{X}_t^i = f(\widehat{X}_t^i)\,dt + G^{1/2}\,d\widehat{W}_t^i + P(\mu_t^N)\,dI_t^i + \Gamma(\mu_t^N)\,dt,$$

$$(13b) \quad dI_t^i = d\mathbf{Y}_t - \big(h(\widehat{X}_t^i)\,dt + U\,d\widehat{W}_t^i + R^{1/2}\,d\widehat{V}_t^i\big),$$

---

[1]Indeed, the system (10) can be recognised as the Stratonovich version of the EnKF, see Appendix C.

[2]In this paper, we assume h to be bounded, circumventing difficulties with the non-Lipschitzness of P in the case when h is unbounded. For results addressing this challenge (although in slightly different settings), we refer the reader to [36, 37, 75, 76].

where $\mu_t^N := \frac{1}{N} \sum_{i=1}^N \delta_{X_t^i}$ is the empirical measure of the system and $\mathbf{Y} \in \mathscr{C}^\alpha([0, T]; \mathbb{R}^d)$ is the canonical lift of a differentiable and bounded path $Y : [0, T] \to \mathbb{R}^d$ with bounded càdlàg derivative. Notice that in the particle system the semimartingale decomposition of $P(\mu_t^N)$ is not independent of $(\widehat{W}^i, \widehat{V}^i)$, so that the Stratonovich and Itô versions of (13) do not coincide. However, the empirical measure μ_t^N will be asymptotically independent of $(\widehat{W}, \widehat{V})$.

Indeed, we have the following well-posedness and convergence result for the interacting particles.

THEOREM (see Remark 5.2 and Theorem 5.8). *Under the same assumptions as Theorem 1.1, let $(\mathbf{Y}^\delta)_{\delta>0}$ be a family of lifts of bounded differentiable paths with càdlàg derivatives that approximate $\mathbf{Y}$ in the rough-path metric. Let $\mu^{\delta,N}$ be the empirical measure of (13) driven by $\mathbf{Y}^\delta$. Then there exists a sequence δ_N such that, for every $t \in [0, T]$, $\mu_t^{N,\delta_N} \stackrel{N\to\infty}{\Rightarrow} \widehat{\pi}_t$, in ρ-Wasserstein distance in L^1.*

For results concerning similar mean-field limits in the classical setting we refer to [42, 43, 77]. The preceding two theorems suggest that numerical methods based on the interacting particle system (13) are robust to perturbations in the data, and hence suitable for applications in multiscale contexts as described in the Introduction. Inspired by Davie's work [32], we propose the following recursive numerical scheme:

$$
\begin{aligned}
X_{k+1}^i = X_k^i &+ f(X_k^i)\Delta t + G^{1/2}\sqrt{\Delta t}\xi_k^i \\
&+ \widehat{P}_k\big(\Delta Y_k - (h(X_k^i)\Delta t + U\sqrt{\Delta t}\xi_k^i + R^{1/2}\sqrt{\Delta t}\eta_k^i)\big) \\
&+ \widehat{\mathrm{Cov}}(x, Dh)\widehat{P}_k\Delta \mathbb{Y}_k + \widehat{\Gamma}_k\Delta t,
\end{aligned}
$$
(14a)

in the following referred to as the *rough-path ensemble Kalman filter* (RP-EnKF). Here, $\Delta t > 0$ is the step size, and (ξ_n^i) and (η_n^i) denote independent zero mean Gaussian random variables of dimensions D and d, respectively. The precise form of the estimator versions $\widehat{P}_k$, $\widehat{\mathrm{Cov}}$ and $\widehat{\Gamma}_k$ will be detailed in Section 6. Finally, the first term in (14a) is built after the Gubinelli derivative from (12), with its precise meaning given by

$$
(15) \quad (\widehat{\mathrm{Cov}}_k(x, Dh)\widehat{P}_k\Delta \mathbb{Y}_k)_\gamma = \sum_{j,q=1}^d \sum_{r=1}^D \widehat{\mathrm{Cov}}_k(x, Dh)_{\gamma,j,r}(\widehat{P}_k)_{r,q}(\Delta \mathbb{Y}_k)_{q,j} \quad \gamma = 1, \dots, D.
$$

Crucially, the scheme (14) takes the lifted component $\Delta \mathbb{Y}_k$ (representing iterated integrals of the path $(Y_t)_{t\geq0}$) as an input. This dependence allows our methodology to appropriately take into account multiscale structure and other information encoded in the signature (such as discrepancies due to model misspecification), but also necessitates appropriate ways to estimate $\mathbb{Y}_k$ from data. We will discuss this in more depth in Sections 2.3 and 6, but note here that it is natural to decompose $\Delta \mathbb{Y}_k$ into its symmetric and skew-symmetric part

$$
(16) \qquad\qquad \Delta \mathbb{Y}_k = \Delta \mathbb{Y}_k^{\mathrm{sym}} + \Delta \mathbb{Y}_k^{\mathrm{skew}}.
$$

The form of the symmetric part $\Delta \mathbb{Y}_k^{\mathrm{sym}}$ is suggested by the requirement that the lifted path is geometric. Indeed, geometric rough paths are also weakly geometric, which by definition satisfy

$$
(17) \qquad\qquad \Delta \mathbb{Y}_k^{\mathrm{sym}} = \frac{1}{2}(y_{k+1} - y_k) \otimes (y_{k+1} - y_k);
$$

in particular this expression can readily be computed from discrete-time observations y_k. The difficulty thus resides in estimating the Lévy area contributions (or corrections) $\Delta \mathbb{Y}_k^{\mathrm{skew}}$. We

suggest a subsampling-based method, establishing connections to multiscale parameter estimation as investigated in [96]; see Section 2.3. Other approaches towards obtaining $\Delta \mathbb{Y}_k$ have been developed in [9, 49, 80]. We would like to stress that although estimating $\Delta \mathbb{Y}_k^{\text{skew}}$ works reasonably well in our experiments (see Sections 6.2 and 6.3), in many applications it might yield satisfactory results to neglect the skew-symmetric part, that is, to use the approximation $\Delta \mathbb{Y}_k \approx \Delta \mathbb{Y}_k^{\text{sym}}$, either because the observation path is one-dimensional, or because the Lévy area term is comparatively small (see Section 6.4). In these cases, the RP-EnKF can be implemented straightforwardly without additional estimation steps, and represents a robust Stratonovich-version of the EnKF (see Appendix C).

Itô, Stratonovich and rough integrals. Before concluding this introduction, let us comment on the difference between the RP-EnKF scheme (14) and the more conventional EnKF updates

$$
(18) \quad
\begin{aligned}
X_{k+1}^i = {}& X_k^i + f(X_k^i)\Delta t + G^{1/2}\sqrt{\Delta t}\,\xi_k^i \\
& + \widehat{P}_k\big(\Delta Y_k - \big(h(X_k^i)\Delta t + U\sqrt{\Delta t}\,\xi_k^i + R^{1/2}\sqrt{\Delta t}\,\eta_k^i\big)\big),
\end{aligned}
$$

see [15, 48, 101]. Clearly, (14) and (18) coincide up to the terms in (14a). This difference can be attributed to alternative (but equivalent) perspectives on the underlying continuous-time dynamics, and hence we expect the terms in (14a) to cancel each other in the limit as $\Delta t \to 0$, when ΔY_k is obtained at discrete time-points from (1b) and $\Delta \mathbb{Y}_k$ are the increments of the Stratonovich lift corresponding to piecewise linear approximation (that can be computed according to (17) and setting $\Delta \mathbb{Y}_k^{\text{skew}} = 0$). Indeed, (18) corresponds to the Euler–Maruyama discretisation of the McKean–Vlasov SDE

$$
(19) \quad \mathrm{d}\widehat{X}_t = f(\widehat{X}_t)\,\mathrm{d}t + G^{1/2}\,\mathrm{d}W_t + P(\widehat{\pi}_t)\big(\mathrm{d}Y_t - \big(h(\widehat{X}_t)\,\mathrm{d}t + U\,\mathrm{d}\widehat{W}_t + R^{1/2}\,\mathrm{d}\widehat{V}_t\big)\big),
$$

understood in the sense of Itô. In contrast, the term $\widehat{\Gamma}_k \Delta t$ in (14a) can be seen as an Itô–Stratonovich correction to (19), see Appendix C, while the first term in (14a) arises from a Milstein-type approximation scheme for Stratonovich SDEs [68], Section 10.3, (or as part of the discrete approximation of rough integrals according to Davie [32]). The equivalence of the classical and the rough path version, that is, of (10) and (11), is stated in Lemma 4.18.

The distinction between (14) and (18) becomes important when the data ΔY_k and $\Delta \mathbb{Y}_k$ is described by (1) only in an approximate sense and hence robustness becomes a key issue. Importantly, the passage to rough integrals is indispensable for the continuity statements in Theorem 1.1 and the numerical robustness of the RP-EnKF scheme demonstrated in Section 6. The Stratonovich picture is natural in view of the Wong–Zakai theorem [54], Section 9.2, and determines the symmetric part in (16) as $\mathbf{Y}$ can then be interpreted as a geometric rough path (see (17) and Section 6). Our approach of replacing Itô by Stratonovich and subsequently rough integrals mirrors the construction in [40] in the context of maximum likelihood parameter estimation for diffusions.

Our contributions and structure of the paper. Our main contributions are as follows:

- Based on the prior works [88, 92, 99], we derive the McKean–Vlasov system (5)–(7), incorporating a stochastic innovation term as well as correlated model and observation noise. Crucially, the dynamics (5)–(7) is given entirely in terms of Stratonovich integrals that allow the construction of robust filtering schemes built on geometric rough paths.
- We prove well-posedness of the rough McKean–Vlasov dynamics (11) as well as continuity in the rough driver $\mathbf{Y}$ (see Theorem 1.1).
- We show well-posedness as well as propagation of chaos of the interacting particle approximation (13).

• We suggest the RP-EnKF scheme (14) and in particular devise a subsampling based method to estimate the lift components $\mathbb{Y}_k$ from data (as corrections in a model misspecification scenario). The robustness of the RP-EnKF (and the nonrobustness of the EnKF) is demonstrated using numerical examples in the context of combined state-parameter estimation.

In Section 2 we review the most relevant results in filtering theory, we motivate the use of the McKean–Vlasov equation and explain the concept of a robust filter. In Section 3 we introduce common notation and recall some background on rough paths. In Section 4.4 we present the analysis of the rough McKean–Vlasov equation (11), which includes well-posedness and stability. In Section 5 we treat the interacting particle system (13) and prove well-posedness and propagation of chaos. In Section 6 we detail the construction of the numerical scheme, including a presentation of our subsampling approach towards estimating the Lévy area. Finally, we conclude the paper with some numerical experiments.

2. Background in filtering, robust representations and McKean–Vlasov dynamics. In this section we discuss essential background on robust filtering and put our work into perspective. Section 2.1 will summarise existing approaches towards solving the filtering problem posed by (4), both from a theoretical as well as from an algorithmic perspective. In Section 2.2 we review the challenges to these methods posed by perturbations in the observed data $(Y_s)_{0 \le s \le t}$, leading to the concept of robustness. In Section 2.3 we draw connections of the McKean–Vlasov approach considered in this paper to maximum likelihood based techniques for stochastic differential equations, in particular to the methods developed in [40]. Finally in Section 2.4 we make our McKean–Vlasov formulation as well as the ensemble Kalman approximation precise.

2.1. *Solutions to the filtering problem and algorithms.* It is well known that the measure π_t defined in (4) is a measurable function of the observation path $(Y_s)_{0 \le s \le t}$ and can be obtained as a solution to the *Kushner–Stratonovich SPDE* [10], Section 3.6,

(20)
$$\pi_t[\phi] = \pi_0[\phi] + \int_0^t \pi_s[\mathcal{L}\phi] \, ds$$
$$+ \int_0^t \left(\pi_s[\phi h^\top] - \pi_s[h^\top] \pi_s[\phi] + \pi_s[(B\nabla\phi)^\top] \right) (C^{-1} \, dY_s - \pi_s[h] \, ds),$$

where

(21)
$$\mathcal{L}\phi = \frac{1}{2} \operatorname{Trace}(G D^2 \phi) + f \cdot \nabla\phi$$

denotes the infinitesimal generator associated to the signal process (1a). The term including $\nabla\phi$ in (20) accounts for the correlation between the noises in (1a) and (1b). We refer the reader to the proof of Proposition A.1 in Appendix A for a detailed explanation based on [10].

From the computational viewpoint, numerically solving (20) directly (for instance, using grid-based methods) is usually infeasible, especially when the dimension D is large (see [10], Section 8.5, for a discussion). Many algorithmic approaches therefore rely on the simulation of carefully constructed interacting particle systems, positing the corresponding (possibly weighted) empirical measures $\frac{1}{N} \sum_{i=1}^{N} \delta_{X_t^i}$ as approximations for π_t.

Sequential Monte Carlo methods rely on Bayes' theorem in order to approximate the conditional expectations (4). More precisely, defining the likelihood

(22)
$$l_t = \exp\left(\int_0^t h(X_s) \cdot C^{-1} \, dY_s - \frac{1}{2} \int_0^t h(X_s) \cdot C^{-1} h(X_s) \, ds \right),$$

the filtering measures admit the representation

$$
\pi_t[\phi] = \frac{\mathbb{E}[\phi(X_t)l_t|\mathcal{Y}_t]}{\mathbb{E}[l_t|\mathcal{Y}_t]}, \tag{23}
$$

according to the Kallianpur–Striebel formula [10], Proposition 3.16. Consequently, approximations of π_t can be obtained by sampling from the signal dynamics (1a) in conjunction with appropriate weighting and/or resampling steps on the basis of (22). For detailed accounts, we refer the reader to [38, 45, 101]. While sequential Monte Carlo methods reproduce the filtering measures exactly in the large-particle limit (see, for instance, [10], Theorem 9.15), their finite-particle versions become unstable due to weight collapse (resulting in low effective sample sizes) as the discrepancy between the unconditional and the conditional path space measures increases. This problem typically becomes more pronounced in situations where the signal dynamics is transient, or when the state space dimension of the signal is high.

Ensemble Kalman filters (EnKFs) [101], Section 7.1, can be formulated in terms of interacting or mean-field (McKean–Vlasov) diffusions. In the case when $U = 0$ (i.e., when the signal and observation noises are uncorrelated [102], Section 7.1, [14]) the basic EnKF due to Evensen [16, 47, 48] is given by

$$
\mathrm{d}\widehat{X}_t = f(\widehat{X}_t)\,\mathrm{d}t + G^{1/2}\,\mathrm{d}\widehat{W}_t + P(\pi_t)C^{-1}\big(\mathrm{d}Y_t - \big(h(\widehat{X}_t)\,\mathrm{d}t + R^{1/2}\,\mathrm{d}\widehat{V}_t\big)\big), \tag{24}
$$

with

$$
P(\pi) = \mathrm{Cov}_\pi(x, h), \tag{25}
$$

or a standard particle approximation thereof. The system (24)–(25) is motivated by the fact that the corresponding law reproduces π_t exactly in the linear Gaussian case: If π_0 is Gaussian, $f(x) = Fx$ and $h(x) = Hx$ for appropriate matrices $H \in \mathbb{R}^{d \times D}$ and $F \in \mathbb{R}^{D \times D}$, then π_t remains Gaussian for all $t \geq 0$, and $\mathrm{Law}(\widehat{X}_t) = \pi_t$. In cases where the preceding conditions are not satisfied, the system (24)–(25) becomes an approximation, the accuracy of which is far from well understood theoretically. However, the ensemble Kalman approach has empirically proven to be both fairly reliable in nonlinear settings as well as scalable to high-dimensional scenarios, and therefore nowadays constitutes one of the workhorses in practical data assimilation tasks [101]. We refer to [15] for a recent review of its theoretical properties.

The more recently proposed *feedback particle filters* [31, 105, 109] rely on carefully designed McKean–Vlasov diffusions of the type (24) such that the associated (conditional, nonlinear) Fokker–Planck equation coincides with the Kushner–Stratonovich SPDE (20). By construction, such models are exact, and the conditional laws induced by the solutions to feedback particle filter dynamics provide the filtering measures (4). As an illustration,

$$
\begin{aligned}
\mathrm{d}\widehat{X}_t &= f(\widehat{X}_t)\,\mathrm{d}t + G^{1/2}\,\mathrm{d}\widehat{W}_t + K(\widehat{X}_t, \pi_t)C^{-1}\big(\mathrm{d}Y_t - \big(h(\widehat{X}_t)\,\mathrm{d}t + R^{1/2}\,\mathrm{d}\widehat{V}_t\big)\big) \\
&\quad + \Xi(\widehat{X}_t, \pi_t)\,\mathrm{d}t,
\end{aligned} \tag{26}
$$

was suggested in [99] and extended in [88], where $K(\cdot, \pi) = \nabla\phi(\cdot, \pi)$ is determined from the elliptic PDE

$$
\mathrm{div}(\pi\nabla\phi) = -\pi\big(h - \pi[h]\big), \tag{27}
$$

and Ξ is an appropriate correction term (see Section 2.4 for an in-depth discussion). Clearly, the systems (24)–(25) and (26)–(27) are strongly related in spirit, combining a replication of the signal dynamics (1a) with a data-dependent nudging term so as to match the observations. Reiterating the discussion so far, solutions to (26)–(27) provide exact solutions to the filtering problem (4), while solutions to (24)–(25) lead to approximate ones (except in the

linear Gaussian case). However, as P is given explicitly in (25), the system (24)–(25) lends itself straightforwardly to efficient numerical integration, while the system (26)–(27) poses a formidable numerical challenge in the form of the high-dimensional PDE (27). What is more, well-posedness of systems of the type (26)–(27), with coefficients that depend on the law through the solution of a PDE is currently not well understood. Nevertheless, McKean–Vlasov formulations of the type (26)–(27) conveniently link between the theoretically optimal Kushner–Stratonovich SPDE (20) and the numerically tractable and practically relevant ensemble Kalman dynamics (24)–(25). In this paper, we leverage this viewpoint in order to construct a robust version of (18).

2.2. *Robust filtering.*   In order to model and solve real-life problems it is highly desirable that the conditional law π_t (or any numerical approximation thereof) depends continuously on the observation path $(Y_s)_{0 \leq s \leq t}$: This property would ensure robustness against misspecification of the underlying signal and observation dynamics (as is typical in reduced-order modeling) as well as against anomalies or artefacts in the collection of the data (such as discretisation errors or perturbation by noise), see [10], Chapter 5, for an overview, and Section 2.3 below. Unfortunately, however, the ϕ-dependent measurable map $(Y_s(\cdot))_{s \in [0,t]} \mapsto \pi_t[\phi]$ provided by (4) can be shown to be neither unique nor continuous [29] in standard topologies. At a fundamental level, this problem is due to the appearance of stochastic integrals against $(Y_s)_{0 \leq s \leq t}$ in (20) and (22) which are well known to induce classically discontinuous maps (for instance, in the supremum norm), see [53, 55]. In order to address this issue and to obtain a continuous[3] version of the process $(\pi_t)_{t \geq 0}$, Clarke suggested using (stochastic) integration by parts in (22) in order to eliminate the dY-dependence [21]. Notably, this approach is restricted to the case when the signal and observation noises are independent (i.e., $U = 0$), see [21, 22, 73], or the case when the observation is one-dimensional (i.e., $Y_t \in \mathbb{R}$), see [33–35]. Addressing the situation of multidimensional correlated observations, the authors of [29] showed by means of a counter-example (see [29], Example 1) that continuity as a function of $(Y_s)_{0 \leq s \leq t}$ is impossible to achieve. Instead, they use rough path lifts $C([0, T]; \mathbb{R}^d) \ni (Y_s)_{0 \leq s \leq T} \mapsto (\mathbf{Y}_s)_{0 \leq s \leq T} \in \mathscr{C}^\alpha([0, T]; \mathbb{R}^d)$ and establish continuity when π_t is considered as a function of the augmented observation path. Similar ideas have been pursued in [60], Theorem 5.3, putting forward the notion of "good" approximations of the observation path. As the aforementioned works are concerned with the likelihood (22), these lay the foundations for the development of robust sequential Monte Carlo methods as reviewed in Section 2.1, and the recent preprint [30] explores that direction. We would also like to mention the works [41, 62] that allow treating the Zakai SPDE (governing the unnormalised filtering distribution [10], Section 3.5) in a rough paths framework, however noticing that the numerical treatment of SPDEs is faced with enormous challenges, in particular in high-dimensional settings. Some other works addressing issues in robust or multiscale filtering include [3, 4] (assuming uncertainty in the coefficients) as well as a sequence of works by N. Perkowski and coworkers in the context of averaging and homogenisation. The latter started with the theoretical set-up in [64, 65] and its extension to correlated noises [12, 13] as well as to multiple time scales [12]. In subsequent developments, the incorporation of ideas from importance sampling and control [81, 82] proved advantageous in applications to increasingly complex dynamical systems such as the Lorenz-96 model [110, 111].

In this paper, we instead construct a robust version of the ensemble Kalman filter (24)–(25) on the basis of its connections to feedback particle formulations as in (26)–(27). Before describing our strategy, we review related work on maximum likelihood parameter estimation for stochastic differential equations.

---

[3]As pointed out in [29], the continuity requirement restores the uniqueness of the map $(Y_s(\cdot))_{s \in [0,t]} \mapsto \pi_t[\phi]$.

2.3. *Parameter estimation and filtering in multiscale systems.* In this section we present a prototypical example that illustrates some of the challenges in robust filtering as well as the scope of the methods developed in this paper, following [88]. Consider the SDE

$$\mathrm{d}Z_t = F(Z_t, \theta)\,\mathrm{d}t + \mathrm{d}W_t, \tag{28}$$

where $Z_t \in \mathbb{R}^d$, and $F : \mathbb{R}^d \times \Theta \to \mathbb{R}^d$ is a parameterised drift vector field, with parameter set $\Theta \subset \mathbb{R}^p$. The objective is to find the true parameter $\theta^* \in \Theta$ from a noisy realisation of $(Z_t)_{t\geq 0}$, that is, we assume that the observation process is given by

$$\mathrm{d}Y_t = \mathrm{d}Z_t + R^{1/2}\,\mathrm{d}V_t. \tag{29}$$

As before, $R \in \mathbb{R}^{d\times d}$ denotes the observation noise covariance, and $(V_t)_{t\geq 0}$ stands for a standard d-dimensional Brownian motion. The problem setting (28)–(29) can be brought into the form (1) by elevating θ to a time-dependent variable, that is, by setting $X_t = (Z_t, \theta_t) \in \mathbb{R}^{d+p}$, hence viewing (28)–(29) as a combined state-parameter estimation problem, see [88]. Accordingly, $f : \mathbb{R}^{d+p} \to \mathbb{R}^{d+p}$ and $h : \mathbb{R}^{d+p} \to \mathbb{R}^d$ are then given as $f(z, \theta) = (F(z, \theta), 0)$ and $h(z, \theta) = F(z, \theta)$, and the matrices $G \in \mathbb{R}^{(d+p)\times(d+p)}$ and $U \in \mathbb{R}^{d\times(d+p)}$ take the form

$$G = \begin{pmatrix} I_{d\times d} & 0_{d\times p} \\ 0_{p\times d} & 0_{p\times p} \end{pmatrix}, \qquad U = \begin{pmatrix} I_{d\times d} & 0_{p\times p} \end{pmatrix}. \tag{30}$$

Finally, the filtering formulation is completed by specifying a *prior distribution* on the initial condition (Z_0, θ_0). The resulting filtering measures $\pi_t \in \mathcal{P}(\mathbb{R}^{d+p})$ encode the Bayesian posterior on the combined variable (Z_t, θ_t). Consequently, the means of the θ-marginals provide a posteriori estimates on the parameter of interest while the variances can be used for Bayesian uncertainty quantification.

In the particular case when the path $(Z_t)_{t\geq 0}$ is observed without contamination by noise, that is, $R = 0$, and F is linear[4] in θ, that is, $F(z, \theta) = \theta f(z)$ with f satisfying appropriate nondegeneracy conditions [40], the parameter $\theta \in \Theta$ can be recovered from the maximum likelihood estimator

$$\theta_T^*(Z) = \frac{\int_0^\top f(Z_t)\,\mathrm{d}Z_t}{\int_0^\top |f(Z_t)|^2\,\mathrm{d}t} \tag{31}$$

in the limit when $T \to \infty$, see [74, 83]. Furthermore, in this case the McKean–Vlasov dynamics suggested in this paper can be solved explicitly, and the corresponding means are directly related to (31), see Appendix B and [88]. It is well known that the estimator (31) can be inaccurate when evaluated on paths that only approximately satisfy (28), for instance when (28) represents a reduced description of underlying multiscale dynamics [2, 96, 114]. A common approach towards addressing this problem is to subsample [96] or smooth out [1] the data, see [91, 95] for methodological aspects and [56, 57] for similar ideas in a slightly different context (here the dynamics (28) is replaced by an ODE due to a particular scaling of the multiscale system). For specific applications see [87] (multiscale inverse problems), [2, 90, 114] (economics and finance), and [27, 112] (ocean and atmospheric science). We mention in passing that regression-type schemes for parameter estimation in multiscale systems have been developed in [66, 69, 70].

A different approach towards robustness of the estimator (31) has been taken in [40], addressing the discontinuity of the Itô integral $\int_0^\top f(Z_t)\,\mathrm{d}Z_t$. To resolve this issue, the authors suggest replacing Itô by Stratonovich integration (motivated by the Wong–Zakai theorem [53], Theorem 9.3, and entailing a correction term involving Df), and subsequently

---

[4]For simplicity of the presentation, we also assume here that θ is one-dimensional, that is, $\Theta \subset \mathbb{R}$.

using rough path integration instead of Stratonovich integration (relying on a suitable lift $C([0, T]; \mathbb{R}^d) \ni (Z_s)_{0 \le s \le T} \mapsto (\mathbf{Z}_s)_{0 \le s \le T} \in \mathscr{C}^\alpha([0, T]; \mathbb{R}^d))$. The resulting estimator

$$\theta_T^{\mathrm{RP}}(\mathbf{Z}) = \frac{\int_0^\top f(Z_t) \, \mathrm{d}\mathbf{Z}_t - \frac{1}{2} \int_0^\top \mathrm{Trace}(Df)(Z_t) \, \mathrm{d}t}{\int_0^\top |f(Z_t)|^2 \, \mathrm{d}t} \tag{32}$$

can then be shown to be continuous as a map from $\mathscr{C}^\alpha([0, T]; \mathbb{R}^d)$ to $\mathbb{R}$. The construction of the RP-EnKF dynamics (14) follows a similar line of reasoning, but our approach is applicable to situations where the path $(Z_t)_{t \ge 0}$ is contaminated by noise ($R \ne 0$) and where $F(z, \theta)$ is nonlinear in θ. The latter generalisation makes our method suitable to applications involving deep learning, that is, when the drift in (28) is parameterised by a neural network as in [59]. We discuss the idea of subsampling the observed data path in the context of constructing an appropriate rough path lift in Remark 6.4 below in Section 6.

2.4. *From the filtering problem to the McKean–Vlasov equation.* In this section, we discuss the construction of the McKean–Vlasov system (5)–(7) and the nature of the approximation in (10) and (11). The general idea goes back to [31, 109], and various modifications have been proposed in [88, 92, 99]. Our formulation combines Stratonovich integration (as in [92]) in order to later invoke Wong–Zakai type approximation results with a stochastic innovation term (as in [88, 101]) as required for the case of correlated model and observation noise. For the sake of clarity, we repeat the PDEs (6) and (7) in their respective index forms (using Einstein's summation convention),

$$\partial_i \big(\hat{\pi}_t (K_t^{ij} - (BC)^{ij}) \big) = -\hat{\pi}_t (h^j - \hat{\pi}_t[h^j]), \quad j = 1, \dots, d, \tag{33}$$

and

$$\partial_i \big(\hat{\pi}_t \Xi_t^i \big) = \frac{1}{2} \hat{\pi} \big((K_t C^{-1})^{ij} \partial_i h^j - \hat{\pi}_t[(K_t C^{-1})^{ij} \partial_i h^j] \big). \tag{34}$$

The system (33)–(34) is analogous to the system (2.5)–(2.6) obtained in [92]; note, however, that the form of $\mathrm{d}I_t$ in (5) is different, and so the right-hand sides of (34) and (2.6) in [92] do not coincide, even in the uncorrelated case ($U = 0$). The McKean–Vlasov system (5)–(7) solves the filtering problem in the following sense.

PROPOSITION 2.1. *Let $T > 0$, assume that the system (1) admits a unique solution (X_t, Y_t) and that the Zakai equation associated to the corresponding filtering problem is well posed. Let π be the conditional law of X given Y as defined in (4) and assume that π_t admits a C^1-density with respect to the Lebesgue measure, $\mathbb{P}$-a.s., for all $t \in [0, T]$. Moreover, assume that the McKean–Vlasov equation (5) admits a unique solution $(\hat{X}_t)_{t \in [0, T]}$ such that its conditional law $\hat{\pi}_t$ admits a C^1-density with respect to the Lebesgue measure, $\mathbb{P}$-a.s. Assume that K and Ξ are predictable stochastic processes with values in $C^1(\mathbb{R}^D; \mathbb{R}^{D \times d})$ and $C^1(\mathbb{R}^D; \mathbb{R}^D)$, respectively, independent from $(\hat{V}_t)_{t \ge 0}$ and $(\hat{W}_t)_{t \ge 0}$, and such that the PDEs (6) and (7) are satisfied, $\mathbb{P}$-a.s. Then $\pi_t = \hat{\pi}_t$, for all $t \in [0, T]$.*

To prove Proposition 2.1, we define the unnormalised conditional law associated to the McKean–Vlasov dynamics (5),

$$\hat{\rho}_t[\phi] := \mathbb{E}\big[\phi(\hat{X}_t) l_t | \mathcal{Y}_t\big], \tag{35}$$

where the likelihood l_t is defined in (22). Comparing the evolution of $\hat{\rho}_t$ with the solution of the Zakai equation [10], Section 3.5, allows us to derive the PDEs (6) and (7). This approach allows us to circumvent the stringent regularity condition in [92], Assumption 3.4. For details see Appendix A.

One of the numerical challenges posed by the system of equations (5)–(7) is to obtain (approximate) solutions K and Ξ to the PDEs (6) and (7). We sidestep this problem by using constant-in-space approximations leading to the system (10) of ensemble Kalman filter type. A similar correspondence has been observed in [105] and is optimal in the following sense.

LEMMA 2.2. *Let $\pi \in \mathcal{P}_1(\mathbb{R}^D)$ and $K : \mathbb{R}^D \to \mathbb{R}^{D \times d}$ be a (weak) solution to (6) such that $\pi[\|X\|_F^2] < \infty$ where $\|\cdot\|_F$ denotes the Frobenius norm. Denote by $\widetilde{K}$ the best constant approximation in least-squares sense, that is,*

$$
(36) \qquad \widetilde{K} = \underset{\widetilde{K} \in \mathbb{R}^{D \times d}}{\arg\min} \int_{\mathbb{R}^D} \|K(x) - \widetilde{K}\|_F^2 \, d\pi(x).
$$

Then $\widetilde{K}$ is given by

$$
(37) \qquad \widetilde{K} = \int_{\mathbb{R}^D} x\big(h(x) - \pi[h]\big)^\top d\pi(x) + BC \in \mathbb{R}^{D \times d}.
$$

Moreover, let Ξ be a solution to (7) with K replaced by $\widetilde{K}$ such that $\pi[|\Xi|^2] < \infty$. Then the best constant approximation of Ξ in least-squares sense (with respect to the Euclidean norm) is given by

$$
\widetilde{\Xi}^\gamma = -\frac{1}{2} \operatorname{Trace}\big(\widetilde{K} C^{-1} \pi[x^\gamma (Dh - \pi[Dh])]\big), \quad 1 \le \gamma \le D.
$$

The proof of this Lemma is postponed to Appendix A.

Constructing numerical approximations for (6)–(7) beyond the constant-in-space approximation is a topic of ongoing research. We mention in particular the approach developed in [106] and analysed in [93] based on diffusion maps as well as the methods based on the Stein geometry [46, 89].

2.5. *Literature on McKean–Vlasov dynamics.* McKean–Vlasov equations are stochastic differential equations whose coefficients depend on the law of the solution. They are sometimes called law-dependent equations. McKean–Vlasov equations have been the subject of several studies starting from the seminal work of McKean [86] and Dobrushin [44]. McKean–Vlasov equations arise as limits of mean-field interacting particle systems, when the number of particles goes to infinity. For a general introduction on the topic we refer the reader to Sznitman [104].

In recent years there has been an increased interest in mean-field particles with common noise, see [24, 71, 72] or [17] in the case of mean-field games. In these types of systems the particles are subject to the same random perturbation and possibly additional independent noises. There is no averaging effect of the common perturbation when the number of particles increases. The limit object is again a law-dependent SDE, but this time the coefficient depends on the conditional law of the solution given the common noise. This is the case for equation (11), where the coefficients P and Γ depend on the law of X given Y.

McKean–Vlasov equations from a rough path perspective were studied for the first time in [18] and more recently in the twin papers [7, 8]. In both of these works the equation is driven by a random rough path that is quite general and can describe the independent noise, the common noise or both. This gives the additional difficulty of needing to keep track of the rough path as an L^p-valued path. In [18] only the drift of the equation depends on the law of the solution, the coefficients in front of the noise depend only on the state of the solution. The more recent work [7] generalises that approach to include law-dependent coefficients. The authors use the approach by Gubinelli on controlled rough paths (see Section 3.2 for a brief introduction on the topic). In order to do this, they need Lions' approach to calculus in

measure spaces endowed with the Wasserstein metric. The equation is then solved as a fixed-point in the mixed $\mathbb{R}^d$ and L^p-space. In [23] the case of pathwise McKean–Vlasov equation with additive noise is considered. The basic techniques used are similar to the ones used in the rough-path case, but the need for rough paths is removed thanks to the additive noise. See also [107].

McKean–Vlasov equations with a rough common noise have been studied recently in [26]. This is the first time that a rough McKean–Vlasov equation is studied when there is a clear separation between independent Brownian motions and a common deterministic noise. In [26], the common noise coefficient depends on both the state of the solution and its law. In equation (11), the coefficient P only depends on the law of $\widehat{X}$, which simplifies the problem to some extent, but also allows us to use a different approach, where we treat the stochastic and the rough integrals in two separated steps. As it will be clear from the proofs in Section 4.4, there is no need to create a joint rough path (or rough driver). All previous works on rough McKean–Vlasov equations deal with bounded coefficients, which cannot be applied here, as the ceofficient P has linear growth in the measure of the solution.

Very recently the authors of [54] developed a theory of mixed rough and stochastic differential equations under Lipschitz and boundedness conditions on the coefficients and they plan to address the application to McKean–Vlasov equations with common noise in a forthcoming paper.

3. Preliminaries.

3.1. *Notation.* Given a metric space (S, d_S), we call $\mathcal{P}(S)$ the space of probability measures on S. For $\pi \in \mathcal{P}(S)$ we denote by $\pi[x] = \int_S x\pi(\mathrm{d}x)$ the mean of π and by $\pi[\phi] = \int_S \phi(x)\pi(\mathrm{d}x)$ the integral of a measurable function $\phi : S \to \mathbb{R}$ with respect to π.

Let $\rho > 0$, if S is a normed space with norm $|\cdot|$ and $\pi \in \mathcal{P}(S)$, we denote by

$$M^\rho(\pi) = \int_S |x|^\rho \pi(\mathrm{d}x), \qquad \overline{M}^\rho(\pi) = \int_S |x - \pi[x]|^\rho \pi(\mathrm{d}x),$$

the ρ-moment of π and the ρ-central moment of π, respectively. For $\rho \geq 1$, we call $\mathcal{P}_\rho(S) \subset \mathcal{P}(S)$ the space of probability measures π on S such that $M^\rho(\pi) < \infty$. We endow this space with the ρ-Wasserstein metric

$$W^\rho_{\rho,S}(\mu, \nu) := \inf_{m \in \Gamma(\mu,\nu)} \iint_{S \times S} d^\rho_S(x, y) m(\mathrm{d}x, \mathrm{d}y),$$

where $\Gamma(\mu, \nu)$ is the set couplings between μ and ν. Given a function $\varphi \in C(\mathbb{R}^m, \mathbb{R}^n)$ and a path $x \in C([0, T]; \mathbb{R}^m)$, we define

$$[\varphi]^{k,x}_{s,t} := \int_0^1 (1 - \theta)^{k-1} \varphi(x_s + \theta \delta x_{s,t}) \, \mathrm{d}\theta,$$

where $\delta x_{s,t} := x_t - x_s$. If $\varphi \in C^1(\mathbb{R}^m, \mathbb{R}^n)$, we use the following notation for the Taylor expansion:

$$\delta\varphi(x.)_{s,t} = [D\varphi]^{1;x}_{s,t} \delta x_{s,t}, \qquad [\varphi]^{1;x}_{s,t} - \varphi(x_s) = [D\varphi]^{2;x}_{s,t} \delta x_{s,t}.$$

Throughout the paper we use $D\varphi$ for the usual Fréchet derivative and we denote by $\nabla\varphi = D\varphi^\top$ the gradient when the function φ is scalar. Whenever E is a Banach space, we denote by $L^p(E)$ the usual Bochner–Lebesgue spaces. We sometimes use the notation $L^\rho_\omega := L^p(\Omega, \mathbb{R}^d)$. For a stochastic process X on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, we sometimes denote the conditional expectation by $\mathbb{E}_s[X_t] := \mathbb{E}[X_t \mid \mathcal{F}_s]$.

3.2. *Background in rough paths.*   The theory of rough paths is a framework that allows well-posedness and stability properties for equations of the form

(38)
$$\dot{X}_t = \xi(X_t)\dot{Y}_t, \quad X_0 \in \mathbb{R}^d,$$

where Y is a path of regularity lower than the regularity assumptions amenable to classical calculus. For $\alpha > 0$, we denote by $C_2^\alpha([0, T]; \mathbb{R}^d)$ the set of all continuous functions

$$g : \{(s, t) \in [0, T]^2 : s < t\} \to \mathbb{R}^d$$

such that there exists a constant C with $|g_{s,t}| \leq C|t - s|^\alpha$ and we denote by $[g]_\alpha$ the infimum over all such constants. We denote by $\|g\|_\alpha := [g]_\alpha + |g_0|$ the α-Hölder norm. We write $C^\alpha([0, T]; \mathbb{R}^d)$ for the set of paths $f : [0, T] \to \mathbb{R}^d$ such that $\delta f \in C_2^\alpha([0, T]; \mathbb{R}^d)$, where we have defined $\delta f_{s,t} := f_t - f_s$.

A *rough path* is a pair $\mathbf{Y} = (Y, \mathbb{Y}) \in \mathscr{C}^\alpha([0, T]; \mathbb{R}^d) \subset C^\alpha([0, T]; \mathbb{R}^d) \times C_2^{2\alpha}([0, T]; \mathbb{R}^{d \times d})$ such that

(39)
$$\mathbb{Y}_{s,t} - \mathbb{Y}_{s,u} - \mathbb{Y}_{u,t} = Y_{s,u} \otimes Y_{u,t}.$$

We equip $\mathscr{C}^\alpha([0, T]; \mathbb{R}^d)$ with its subset topology which we shall call the rough path topology. Relation (39), commonly referred to as Chen's relation, encodes the algebraic property between a path and its *iterated integral*, viz the formal equality

$$\mathbb{Y}_{s,t} = \int_s^t Y_{s,r} \otimes \mathrm{d}Y_r.$$

When $\alpha \in (\frac{1}{3}, \frac{1}{2}]$ the above integral is in general not canonically defined using functional analysis. However, in the case of Y being a sample path of the Brownian motion, B, one can use probability theory to define iterated integrals using for example, Itô integration or Stratonovich integration. We denote by $\mathbf{B} := (B, \mathbb{B}^{\mathrm{Itô}}) := (B, \int B \otimes \mathrm{d}B)$ and $\mathbf{B}^{\mathrm{Strat}} := (B, \mathbb{B}^{\mathrm{Strat}}) := (B, \int B \otimes \circ \mathrm{d}B)$ these (random) rough paths, respectively. It is classical that

(40)
$$\mathbb{B}_{s,t}^{\mathrm{Itô}} = \mathbb{B}_{s,t}^{\mathrm{Strat}} - \frac{1}{2}(t - s)I_{d \times d}.$$

The Stratonovich rough path is an example of a *geometric* rough path, that is to say it is in the closure in the rough path topology of the image of the mapping

$$Y \mapsto \left(Y, \int Y \otimes \mathrm{d}Y\right)$$

defined on $BV([0, T]; \mathbb{R}^d)$.

Given two rough paths $\mathbf{Y}^1, \mathbf{Y}^2 \in \mathscr{C}^\alpha([0, T]; \mathbb{R}^d)$ we define the following distance:

$$\rho_\alpha(\mathbf{Y}^1, \mathbf{Y}^2) := \|Y^1 - Y^2\|_\alpha + \|\mathbb{Y}^1 - \mathbb{Y}^2\|_{2\alpha}.$$

We refer the reader to [53] for a more comprehensive discussion of the rough paths notation and concepts used in this paper.

3.3. *Controlled rough paths and rough differential equations.*   To use rough paths for a solution theory of equations of the form (38), which we rewrite with the formal expression

(41)
$$\mathrm{d}X_t = \xi(X_t)\,\mathrm{d}\mathbf{Y}_t,$$

we start with the ansatz that the solution X takes the form of a Taylor-like expansion

(42)
$$\delta X_{s,t} = X_s' \delta Y_{s,t} + X_{s,t}^\sharp,$$

where $X^\sharp$ is of higher regularity than X, and X' is the so-called Gubinelli derivative. We denote by $\mathscr{D}_Y^{2\alpha}([0,T];\mathbb{R}^n)$ the set of all pairs (X,X') such that $X^\sharp$ implicitly defined via (42) satisfies $(X',X^\sharp) \in C^\alpha([0,T];\mathcal{L}(\mathbb{R}^d;\mathbb{R}^n)) \times C_2^{2\alpha}([0,T];\mathbb{R}^n)$, which also induces the topology on $\mathscr{D}_Y^{2\alpha}([0,T];\mathbb{R}^n)$, through the norm

$$\|(X,X')\|_{\mathscr{D}_Y^{2\alpha}} := \|X'\|_\alpha + \|X^\sharp\|_{2\alpha}.$$

When $(X,X') \in \mathscr{D}_Y^{2\alpha}([0,T];\mathbb{R}^n)$ and $(\hat{X},\hat{X}') \in \mathscr{D}_{\hat{Y}}^{2\alpha}([0,T];\mathbb{R}^n)$ we define

$$\|X,\hat{X}\|_{Y,\hat{Y},2\alpha} := \|X' - \hat{X}'\|_\alpha + \|X^\sharp - \hat{X}^\sharp\|_{2\alpha}.$$

The sewing lemma provides a continuous integration mapping

$$\mathscr{D}_Y^{2\alpha}([0,T];\mathcal{L}(\mathbb{R}^n,\mathbb{R}^d)) \longrightarrow \mathscr{D}_Y^{2\alpha}([0,T];\mathbb{R}^n)$$

$$(X,X') \longmapsto \left(\int X_r\, d\mathbf{Y}_r, X\right),$$

where

$$X_{s,t}^\flat := \int_s^t X_r\, d\mathbf{Y}_r - X_s Y_{s,t} - X_s' \mathbb{Y}_{s,t}$$

satisfies $|X_{s,t}^\flat| \le C|t-s|^{3\alpha}$ for some constant C only depending on α. A solution of (41) can now be defined as a fixed point of the composition of the mappings

$$\mathscr{D}_Y^{2\alpha}([0,T];\mathbb{R}^n) \longrightarrow \mathscr{D}_Y^{2\alpha}([0,T];\mathcal{L}(\mathbb{R}^n,\mathbb{R}^d)) \longrightarrow \mathscr{D}_Y^{2\alpha}([0,T];\mathbb{R}^n)$$

$$(X,X') \longmapsto (\xi(X),\xi(X)') = (\xi(X), D\xi(X)X') \longmapsto \left(\int \xi(X_r)\, d\mathbf{Y}_r, \xi(X)\right).$$

From the sewing lemma and the definition of the integration mapping we see that we could equivalently define the solution of (41) as a path $X : [0,T] \to \mathbb{R}^n$ such that

$$X_{s,t}^\flat := \delta X_{s,t} - \xi(X_s)Y_{s,t} - D\xi(X_s)\xi(X_s)\mathbb{Y}_{s,t}$$

satisfies $|X_{s,t}^\flat| \le C|t-s|^{3\alpha}$. The latter formulation is usually referred to as Davie's expansion/solution.

One of the remarkable properties of rough path equations is the continuity of the Itô–Lyons map,

$$\mathscr{C}^\alpha([0,T];\mathbb{R}^d) \longrightarrow C^\alpha([0,T];\mathbb{R}^n)$$
$$\mathbf{Y} \longmapsto X,$$

where $X = X^{\mathbf{Y}}$ denotes the solution of (41), provided f is regular enough. In fact, Theorem 1.1 is an analogous result for the McKean–Vlasov dynamics of the ensemble Kalman filter treated in this paper.

4. Stochastic rough McKean–Vlasov equations. In this section, unless otherwise specified we fix $T > 0$ and $\rho \ge 1$. Moreover, $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge 0}, \mathbb{P})$ is a complete filtered probability space that supports a standard m-dimensional Brownian motion W. Recall that for $\pi \in \mathcal{P}_\rho(\mathbb{R}^D)$, we call $M^\rho(\pi)$ the ρ-moment of π and $\overline{M}^\rho(\pi)$ the ρ-central moment of π.

4.1. *McKean–Vlasov SDEs with linear growth in the measure.* Consider the measurable functions $b : [0, T] \times \mathbb{R}^D \times \mathcal{P}_\rho(\mathbb{R}^D) \to \mathbb{R}^D$ and $\sigma : [0, T] \times \mathbb{R}^D \times \mathcal{P}_\rho(\mathbb{R}^D) \to \mathbb{R}^{D \times m}$ satisfying the following assumptions.

ASSUMPTION 1. Let $\rho \geq 1$. Assume that there exists a constant $C > 0$ such that:

(i) (linear growth) $\forall t \in [0, T]$, $\forall x \in \mathbb{R}^D$, $\forall \pi \in \mathcal{P}_\rho(\mathbb{R}^D)$,

$$(43) \qquad \left| b(t, x, \pi) \right|, \left| \sigma(t, x, \pi) \right| \leq C \left(1 + M^\rho(\pi) \right)^{\frac{1}{\rho}},$$

(ii) (locally Lipschitz) $\forall t \in [0, T]$, $\forall x, y \in \mathbb{R}^D$, $\forall \pi, \nu \in \mathcal{P}_\rho(\mathbb{R}^D)$,

$$\left| b(t, x, \pi) - b(t, y, \nu) \right|, \left| \sigma(t, x, \pi) - \sigma(t, y, \nu) \right|$$

$$(44) \qquad \leq C \left(1 + M^\rho(\pi) \right)^{\frac{1}{\rho}} \left(|x - y| + W_{\rho, \mathbb{R}^D}(\pi, \nu) \right).$$

REMARK 4.1. Notice that, if $\rho \geq \rho'$, Assumption 1 with ρ' implies Assumption 1 with ρ.

Consider the following McKean–Vlasov equation:

$$(45) \qquad dX_t = b\left(t, X_t, \mathcal{L}(X_t) \right) dt + \sigma\left(t, X_t, \mathcal{L}(X_t) \right) dW_t, \qquad X_t|_{t=0} = X_0 \in L_\omega^2,$$

where W is an m-dimensional Brownian motion.

LEMMA 4.2. *Assume that b and σ satisfy Assumption 1 with $\rho \geq 2$, equation (45) admits a unique strong solution in the classical sense. Moreover, the solution has continuous sample paths.*

PROOF. To prove well-posedness we use a Picard-type argument. Let $\pi_0 := \mathcal{L}(X_0)$. For $T \geq 0$ and $K > 2M^\rho(\pi_0) \vee 1$, we define

$$B_K := \left\{ \pi \in \mathcal{P}\left(C([0, T]; \mathbb{R}^D) \right) \mid M^\rho(\pi_t) \leq K \ \forall t \in [0, T], M^\rho(\pi_0) \leq K/2 \right\}.$$

Let $\pi \in B_K$. We define the following stochastic differential equation:

$$(46) \qquad dX_t = b(t, X_t, \pi_t) dt + \sigma(t, X_t, \pi_t) dW_t, \qquad \tilde{X}_t|_{t=0} = X_0 \sim \pi_0.$$

By our choice of π and Assumption 1, the coefficients $b(\cdot, \cdot, \pi_\cdot)$ and $\sigma(\cdot, \cdot, \pi_\cdot)$ are bounded and Lipschitz. By standard theory this equation admits a unique strong solution on the interval $[0, T]$, which we denote by X^π. This solution has continuous sample paths, $\mathbb{P}$-a.s. We are ready to define the map

$$\Phi : B_K \to \mathcal{P}\left(C([0, T]; \mathbb{R}^D) \right)$$
$$\pi \mapsto \mathcal{L}(X^\pi).$$

Using standard stochastic calculus estimates, Assumption 1 and Gronwall's lemma we obtain the following upper bound on the ρth moment:

$$(47) \qquad \left\| \sup_{t \in [0, T]} |X_t| \right\|_{L_\omega^\rho} \leq K/2 + TC(1 + K),$$

where $C > 0$ is a generic constant that only depends on the coefficients b and σ. Let us now take $T = 1/(4C)$, with this choice we have $\| \sup_{t \in [0, T]} |X_t| \|_{L_\omega^\rho} \leq K$. This implies $\Phi(B_K) \subset B_K$. Using again stochastic calculus and Gronwall's lemma, we have that for each $t \in [0, T]$,

$$W_{\rho, C([0,t], \mathbb{R}^D)}\left(\Phi(\pi), \Phi(\pi') \right)^\rho$$

$$(48) \qquad \leq \left\| \sup_{s \in [0,t]} \left| X_s^\pi - X_s^{\pi'} \right| \right\|_{L_\omega^\rho}^\rho$$

$$\leq C(1 + K) e^{C(1+K)t} \int_0^t W_{\rho, \mathbb{R}^D}\left(\pi_s, \pi'_s \right)^\rho ds, \qquad \pi, \pi' \in B_K.$$

Here $C > 0$ is again a generic constant depending only on b and σ, possibly different than before. Also notice that π, π' are measures on the path space $C([0, T]; \mathbb{R}^D)$ and in the left-hand side of (48) we are considering, with an abuse of notation, their projections on $C([0, t], \mathbb{R}^D)$.

Notice that, for every $s \in [0, T]$, $W_{\rho, \mathbb{R}^D}(\pi_s, \pi'_s)^\rho \leq W_{\rho, C([0,s], \mathbb{R}^D)}(\pi, \pi')$, for any two measures in $\pi, \pi' \in B_K$. For every $n \in \mathbb{N}$, let Φ^n the nth composition of ϕ with itself. We can thus apply (48) n times and obtain

$$W_{\rho, C([0,T]; \mathbb{R}^D)}\big(\Phi^n(\pi), \Phi^n(\pi')\big)^\rho$$

$$\leq C^n (1 + K)^n e^{nC(1+K)T} \int_0^t \int_0^{t_1} \cdots \int_0^{t_{n-1}} W_{\rho, \mathbb{R}^D}(\pi_s, \pi'_s)^\rho \, \mathrm{d}s \, \mathrm{d}t_{n-1} \cdots \mathrm{d}t_1$$

$$\leq \frac{e^{(n+1)C(1+K)T}}{n!} W_{\rho, C([0,T]; \mathbb{R}^D)}(\pi, \pi')^\rho, \quad \pi, \pi' \in B_K.$$

In the last inequality we bounded from above the Wasserstein distance in $\mathbb{R}^d$ with the Wasserstein distance on the path space and grouped it out of the integral. Then we computed the iterated integral.

For the choice of $T = 1/(4C)$, if we take n large enough, we have that the map Φ^n is a contraction on B_K. By the Banach fixed point theorem, Φ^n has a unique fixed point on B_K, which is also the unique fixed point of Φ (see [23], Proposition 6), which is the unique solution to equation (45) up to time T.

Global existence and uniqueness follow by iterating this argument on intervals of fixed length $1/(4C)$, which does not depend on the value of the initial condition, but only on the assumptions on the coefficients of the equation. $\quad\square$

4.2. *The common noise case.* Consider measurable functions b and σ as in the previous section and $\beta : [0, T] \times \mathcal{P}_\rho(\mathbb{R}^D) \to \mathbb{R}^{D \times m_1}$, each of which satisfying the following assumption on their respective domain.

ASSUMPTION 2. Let $\rho \geq 1$. Let $(V, |\cdot|)$ be a Banach space, for $f : [0, T] \times \mathbb{R}^D \times \mathcal{P}_\rho(\mathbb{R}^D) \to V$ assume that there exists a constant $C > 0$ such that:

(i) (linear growth) $\forall x \in \mathbb{R}^D, \forall \pi \in \mathcal{P}_\rho(\mathbb{R}^d), \forall t \in [0, T]$,

$$(49) \qquad \big|f(t, x, \pi)\big| \leq C\big(1 + \overline{M}^\rho(\pi)\big)^{\frac{1}{\rho}},$$

(ii) (locally Lipschitz) $\forall x, y \in \mathbb{R}^D, \forall \pi, \nu \in \mathcal{P}_\rho(\mathbb{R}^D), \forall t \in [0, T]$,

$$(50) \qquad \big|f(t, x, \pi) - f(t, y, \nu)\big| \leq C\big(1 + \overline{M}^\rho(\pi)\big)^{\frac{1}{\rho}}\big(|x - y| + W_{\rho, \mathbb{R}^D}(\pi, \nu)\big).$$

Notice that, since β is independent of the space variable $x \in \mathbb{R}^D$, condition 2(ii) reduces to local Lipschitz continuity in the measure variable.

Consider the following McKean–Vlasov equation:

$$(51) \qquad \begin{aligned} &\mathrm{d}X_t = b\big(t, X_t, \mathcal{L}(X_t \mid \mathcal{B}_t)\big) \mathrm{d}t + \sigma\big(t, X_t, \mathcal{L}(X_t \mid \mathcal{B}_t)\big) \mathrm{d}W_t + \beta\big(t, \mathcal{L}(X_t \mid \mathcal{B}_t)\big) \mathrm{d}B_t, \\ &X_t|_{t=0} = X_0 \in L_\omega^2, \end{aligned}$$

where W is the m-dimensional Brownian motion fixed at the beginning of the section and B is an m_1-dimensional Brownian motion adapted to $(\mathcal{F}_t)_{t \geq 0}$. Assume that X_0, W, B are independent. In (51), $\mathcal{L}(X \mid \mathcal{B}_t)$ is the conditional law of the solution X given the filtration $\mathcal{B}_t := \sigma(B_s \mid 0 \leq s \leq t)$ generated by the common noise B.

LEMMA 4.3.   *Assume that b, σ and β satisfy Assumption* 2 *for some $\rho \geq 2$. Let X and Z be any two solutions to equation* (51) *starting from the same initial condition $X_0 = Z_0 \in L_\omega^\rho$. Then X and Z are indistinguishable.*

PROOF.   It follows from the independence of W, B and X_0 that

$$\mathbb{E}[X_t \mid \mathcal{B}_t] = \mathbb{E}[X_0] + \int_0^t \mathbb{E}\big[b\big(s, X_s, \mathcal{L}(X_s \mid \mathcal{B}_s)\big) \mid \mathcal{B}_s\big] \, ds + \int_0^t \mathbb{E}\big[\beta\big(s, \mathcal{L}(X_s \mid \mathcal{B}_s)\big) \mid \mathcal{B}_s\big] \, dB_s.$$

Let $\overline{X}_t := X_t - \mathbb{E}[X_t \mid \mathcal{B}_t]$, we have

$$\overline{X}_t = \overline{X}_0 + \int_0^t \big(b\big(s, X_s, \mathcal{L}(X_s \mid \mathcal{B}_s)\big) - \mathbb{E}\big[b\big(s, X_s, \mathcal{L}(X_s \mid \mathcal{B}_s)\big) \mid \mathcal{B}_s\big]\big) \, ds$$

$$+ \int_0^t \sigma\big(s, X_s, \mathcal{L}(X_s \mid \mathcal{B}_s)\big) \, dW_s.$$

Set $N_t := \int_0^t \sigma(s, X_s, \mathcal{L}(X_s \mid \mathcal{B}_s)) \, dW_s$, as an application of Itô's formula we have

$$\mathbb{E}\big[|N_t|^\rho \mid \mathcal{B}_t\big] \leq \rho \mathbb{E}\bigg[\int_0^t |N_s|^{\rho-2} N_s \cdot \sigma\big(s, X_s, \mathcal{L}(X_s \mid \mathcal{B}_s)\big) \, dW_s \mid \mathcal{B}_t\bigg]$$

$$+ C \int_0^t \mathbb{E}\big[|N_s|^{\rho-2}|\sigma\sigma^\top|\big(s, X_s, \mathcal{L}(X_s \mid \mathcal{B}_s)\big) \mid \mathcal{B}_t\big] \, ds$$

$$\leq C \int_0^t \mathbb{E}\big[|N_s|^{\rho-2} \mid \mathcal{B}_t\big]\big(1 + \overline{M}^\rho\big(\mathcal{L}(X_s \mid \mathcal{B}_s)\big)\big)^{\frac{2}{\rho}} \, ds$$

$$\leq C \int_0^t \mathbb{E}\big[|N_s|^{\rho-2} \mid \mathcal{B}_t\big]^{\frac{\rho}{\rho-2}} + \big(1 + \overline{M}^\rho\big(\mathcal{L}(X_s \mid \mathcal{B}_s)\big)\big)^{\frac{2}{\rho} \cdot \frac{\rho}{2}} \, ds,$$

where in the last line we used Young inequality with $\frac{\rho-2}{\rho} + \frac{2}{\rho} = 1$. We use Jensen's inequality and Gronwall's lemma to obtain

$$\mathbb{E}\big[|N_t|^\rho \mid \mathcal{B}_t\big] \leq C \int_0^t \big(1 + \overline{M}^\rho\big(\mathcal{L}(X_s \mid \mathcal{B}_s)\big)\big) \, ds.$$

By using Assumption 2(i) also on the drift, we obtain

$$\overline{M}^\rho\big(\mathcal{L}(X_t \mid \mathcal{B}_t)\big) = \mathbb{E}\big[|\overline{X}_t|^\rho \mid \mathcal{B}_t\big] \leq \overline{M}^\rho\big(\mathcal{L}(X_0 \mid \mathcal{B}_0)\big) + C \int_0^t \big(1 + \overline{M}^\rho\big(\mathcal{L}(X_s \mid \mathcal{B}_s)\big)\big) \, ds.$$

Gronwall's lemma gives $\overline{M}^\rho\big(\mathcal{L}(X_t \mid \mathcal{B}_t)\big) \leq C_T$, where $C_T > 0$ is a constant depending on T. A similar bound can be obtained for Z.

Combining the upper bounds on the central moment with Assumption 2, we have global Lipschitz continuity and boundedness of the coefficients, in the sense that b, σ, β satisfy

$$\big|f(t, x, \pi) - f(t, y, \nu)\big| \leq C_T\big(|x - y| + W_{\rho, \mathbb{R}^D}(\pi, \nu)\big).$$

Using this bound, Burkholder–Davis–Gundy inequality and standard estimates we have

$$\mathbb{E}\Big[\sup_{t \leq T} |X_t - Z_t|^\rho\Big] \leq C_T \int_0^T \mathbb{E}\Big[\sup_{s \leq t} |X_s - Z_s|^\rho\Big] + \mathbb{E}\Big[\sup_{s \leq t} W_{\rho, \mathbb{R}^D}\big(\mathcal{L}(X_s \mid \mathcal{B}_s), \mathcal{L}(Z_s \mid \mathcal{B}_s)\big)^\rho\Big] \, dt.$$

Notice that

$$W_{\rho, \mathbb{R}^D}\big(\mathcal{L}(X_s \mid \mathcal{B}_s), \mathcal{L}(Z_s \mid \mathcal{B}_s)\big)^\rho \leq \mathbb{E}\big[|X_s - Z_s|^\rho \mid \mathcal{B}_s\big].$$

We can now use the tower property of the conditional expectation and Gronwall's lemma to obtain

$$\mathbb{E}\Big[\sup_{t \leq T} |X_t - Z_t|^\rho\Big] = 0,$$

which implies indistinguishability of the processes $(X_t)_{t \in [0,T]}$ and $(Z_t)_{t \in [0,T]}$. $\square$

REMARK 4.4.   Weak existence follows from [61] and one can apply Yamada–Watanabe to obtain well-posedness of (51). However, well-posedness will also follow from our results on mixed rough and stochastic McKean–Vlasov equations in the special case $\beta(t, \pi) = P(\pi)$ defined in (61) below.

4.3. *McKean–Vlasov with continuous deterministic forcing.*   Let $b : \mathbb{R}^D \times \mathcal{P}_\rho(\mathbb{R}^D) \to \mathbb{R}^D$ and $\sigma : \mathbb{R}^D \times \mathcal{P}_\rho(\mathbb{R}^D) \to \mathbb{R}^{D \times m}$ be measurable functions satisfying the following assumptions.

ASSUMPTION 3.   Let $\rho \geq 1$. Assume that there exists a constant $C > 0$ such that:

(i) (linear growth) $\forall x \in \mathbb{R}^D, \forall \pi \in \mathcal{P}_\rho(\mathbb{R}^d)$,

$$\tag{52} \left| b(x, \pi) \right|, \left| \sigma(x, \pi) \right| \leq C \left(1 + \overline{M}^\rho(\pi) \right)^{\frac{1}{\rho}},$$

(ii) (Lipschitz continuity) $\forall x, y \in \mathbb{R}^d, \forall \pi, \nu \in \mathcal{P}_\rho(\mathbb{R}^D)$,

$$\tag{53} \left| b(x, \pi) - b(y, \nu) \right|, \left| \sigma(x, \pi) - \sigma(y, \nu) \right| \leq C \left(1 + \overline{M}^\rho(\pi) \right)^{\frac{1}{\rho}} \left(|x - y| + W_{\rho, \mathbb{R}^D}(\pi, \nu) \right).$$

Let $F : [0, T] \to \mathbb{R}^D$ be a continuous bounded function, $X_0 \in L_\omega^\rho$ and consider the following stochastic differential equation:

$$\tag{54} \mathrm{d}X_t = b(X_t, \mathcal{L}(X_t)) \, \mathrm{d}t + \sigma(X_t, \mathcal{L}(X_t)) \, \mathrm{d}W_t + \mathrm{d}F_t, \qquad X_t|_{t=0} = X_0.$$

DEFINITION 4.5.   A stochastic process $(X_t)_{t \in [0,T]}$ on $\mathbb{R}^D$ is a solution for equation (54) with initial condition X_0 if $(\sigma(X_t, \mathcal{L}(X_t)))_{t \in [0,T]}$ is predictable and for every $t \in [0, T]$, the following integral equation is satisfied $\mathbb{P}$-a.s.:

$$X_t - F_t = X_0 - F_0 + \int_0^t b(X_s, \mathcal{L}(X_s)) \, \mathrm{d}s + \int_0^t \sigma(X_s, \mathcal{L}(X_s)) \, \mathrm{d}W_s.$$

REMARK 4.6.   In the following, we will construct an adapted solution $(X_t)_{t \in [0,T]}$ with continuous sample paths. Since σ is Lipschitz continuous, we immediately have that $(\sigma(X_t, \mathcal{L}(X_t)))_{t \in [0,T]}$ is predictable and the Itô integral is well defined.

Let X_t be a solution to equation (54), we define $\overline{X}_t := X_t - \mathbb{E}[X_t]$. We start with some preliminary expansions and estimates for $\overline{X}$. We have, for $s, t \in [0, T]$,

$$\tag{55} \delta \overline{X}_{s,t} = \int_s^t \left(b(X_r, \mathcal{L}(X_r)) - \mathbb{E}[b(X_r, \mathcal{L}(X_r))] \right) \mathrm{d}r + \int_s^t \sigma(X_r, \mathcal{L}(X_r)) \, \mathrm{d}W_r.$$

LEMMA 4.7.   *Let* $\rho \geq 2$ *and* $X_0 \in L_\omega^\rho$. *There exists a constant* C_T *such that* $C_T \to 0$ *as* $T \to 0$ *and*

$$\left\| \sup_{t \in [0,T]} |\overline{X}_t| \right\|_{L_\omega^\rho} \leq e^{C_T} \left[C_T + \| \overline{X}_0 \|_{L_\omega^\rho} \right].$$

Moreover, for $\alpha < \frac{1}{2}$ *and* $\rho > \frac{2}{1 - 2\alpha}$, *we have*

$$\left\| [\overline{X}]_\alpha \right\|_{L_\omega^\rho} \leq C_T \left(1 + \| \overline{X}_0 \|_{L_\omega^\rho} \right).$$

PROOF.   Using standard estimates and the Burkholder–Davis–Gundy (BDG) inequality on equation (55) we have

$$\left\| \sup_{t \in [0,T]} |\overline{X}_t| \right\|_{L_\omega^\rho}^\rho \leq \| \overline{X}_0 \|_{L_\omega^\rho}^\rho + C \int_0^T \left(1 + \overline{M}^\rho(\mathcal{L}(X_t)) \right) \mathrm{d}t.$$

The first inequality follows from a standard application of Gronwall's lemma. For the second inequality we use again BDG and Jensen's inequality to obtain

$$\mathbb{E}\big[|\delta\overline{X}_{s,t}|^{\rho}\big] \leq |t-s|^{\rho/2-\epsilon} C_T\big(1 + \|\overline{X}_0\|_{L^{\rho}_{\omega}}\big),$$

for some $\epsilon > 0$ arbitrarily small. The second inequality as well as the condition on ρ follows from the Kolmogorov continuity theorem [53], Theorem 3.1. $\quad\square$

REMARK 4.8.   The bounds in Lemma 4.7 do not depend on the forcing term F. This gives us a good a priori bound on the solution. Let $\overline{T} > 0$ be fixed and arbitrarily large and assume that there exists $M_0 > 0$ such that $\|\overline{X}_0\|_{L^{\rho}} \leq M_0$. Then there exists a global constant $M(M_0, \overline{T})$ such that

$$(56) \qquad \Big\| \sup_{t \in [0,T]} |\overline{X}_t| \Big\|_{L^{\rho}_{\omega}}, \big\|[\overline{X}]_{\alpha}\big\|_{L^{\rho}_{\omega}} \leq M.$$

We have the following a priori estimates.

LEMMA 4.9.   *Let $\rho \geq 2$. Given $X_0 \in L^{\rho}_{\omega}$ and $F \in C([0,T]; \mathbb{R}^D)$, we call $X(F, X_0)$ a solution to equation* (54) *with forcing F and initial condition X_0.*

Let b, σ satisfy Assumption 3. Then there exists a constant $C_T := C(T, \rho) > 0$ such that $C_T \to 0$ as $T \to 0$ and

$$(57) \qquad \Big\| \sup_{t \in [0,T]} |X_t(F, X_0)| \Big\|_{L^{\rho}_{\omega}} \leq e^{C_T}\big[C_T + \|F\|_{C_b} + \|X_0\|_{L^{\rho}_{\omega}}\big].$$

Given $X_0, Y_0 \in L^{\rho}_{\omega}$ and $F, G \in C([0,T]; \mathbb{R}^D)$, there exists a positive constant $C > 0$ such that

$$(58) \quad \Big\| \sup_{t \in [0,T]} |X_t(F, X_0) - X_t(G, Y_0)| \Big\|_{L^{\rho}_{\omega}} \leq e^{C(1+\|\overline{X}_0\|_{L^{\rho}_{\omega}})}\big(\|F - G\|_{C_b} + \|X_0 - Y_0\|_{L^{\rho}_{\omega}}\big).$$

PROOF.   We write X and omit here the dependence of the process on F and X_0 as there is no possibility of confusion in the first part of the proof. Using Jensen's inequality, the Burkholder–Davis–Gundy inequality and Assumption 3 we have the following estimate for $s, t \in [0, T]$:

$$\mathbb{E} \sup_{t \in [0,T]} |X_t|^{\rho} \leq \mathbb{E}|X_0|^{\rho} + CT^{\rho-1} \int_0^{\top} \mathbb{E}|b(X_r, \mathcal{L}(X_r))|^{\rho} \, dr$$

$$+ CT^{\rho/2-1} \int_0^{\top} \mathbb{E}|\sigma(X_r, \mathcal{L}(X_r))|^{\rho} \, dr + C \sup_{t \in [0,T]} |F_t|^{\rho}$$

$$\leq \mathbb{E}|X_0|^{\rho} + C(T^{\rho-1} + T^{\rho/2-1}) \int_0^{\top} (1 + \mathbb{E}|\overline{X}_r|^{\rho}) \, dr + C \sup_{t \in [0,T]} |F_t|^{\rho},$$

where the constant C depends only on ρ, b and σ. Equation (57) follows immediately using the bounds in Lemma 4.7.

We now prove inequality (58). Applying a similar reasoning as before, we obtain

$$\mathbb{E} \sup_{t \in [0,T]} |X_t(F, X_0) - X_t(G, Y_0)|^{\rho} \leq C(T^{\rho} + T^{\rho/2}) \int_0^T (1 + \mathbb{E}|\overline{X}_t(F, X_0)|^{\rho})$$

$$\cdot \Big(\mathbb{E} \sup_{r \in [0,t]} |X_r(F, X_0) - X_r(G, Y_0)|^{\rho} + W_{\rho, \mathbb{R}^D}\big(\mathcal{L}(X_t(F, X_0)), \mathcal{L}(X_t(G, Y_0))\big)^{\rho}\Big) \, dt$$

$$+ \|F - G\|_{C_b}.$$

The ρ-Wasserstein distance is controlled by the L_ω^ρ norm of the difference of the processes, so we can apply Gronwall's lemma and Lemma 4.7 to obtain the desired inequality (58). $\quad\square$

In addition to the previous bounds, if the forcing term is α-Hölder continuous, we have that also the solution is α-Hölder continuous.

LEMMA 4.10.   *Let $\alpha \in (\frac{1}{3}, \frac{1}{2})$ and $\rho > 2/(1 - 2\alpha)$. Assume that $F, G \in C^\alpha([0, T]; \mathbb{R}^D)$ and $X_0, Y_0 \in L_\omega^\rho$. Then there exist $C_T, C := C(\rho, T, \alpha) > 0$ such that $C_T \to 0$ as $T \to 0$ and*

$$\big\| [X(F, X_0)]_\alpha \big\|_{L_\omega^\rho} \le C_T \big(1 + \|\overline{X}_0\|_{L_\omega^\rho}\big) + C[F]_\alpha. \tag{59}$$

Moreover,

$$\begin{aligned} \big\| [X(F, X_0) - X(G, Y_0)]_\alpha \big\|_{L_\omega^\rho} &\le C_T e^{C(1 + \|\overline{X}_0\|_{L_\omega^\rho})} \big(\|F - G\|_{C_b} + \|X_0 - Y_0\|_{L_\omega^\rho}\big) \\ &\quad + C[F - G]_\alpha. \end{aligned} \tag{60}$$

PROOF.   We write X and omit here the dependence of the process on F, X_0 as there is no possibility of confusion in the first part of the proof. Using Jensen's inequality, the Burkholder–Davis–Gundy inequality and Assumption 3 we have the following estimate for $s, t \in [0, T]$:

$$\begin{aligned} &\mathbb{E}\big[|\delta(X - F)_{s,t}|^\rho\big] \\ &\quad \le C|t - s|^{\rho - 1} \int_s^t \mathbb{E}|b(X_r, \mathcal{L}(X_r))|^\rho \, dr + C|t - s|^{\rho/2 - 1} \int_s^t \mathbb{E}|\sigma(X_r, \mathcal{L}(X_r))|^\rho \, dr \\ &\quad \le C\big(|t - s|^{\rho - 1} + |t - s|^{\rho/2 - 1}\big) \int_s^t \big(1 + \mathbb{E}|\overline{X}_r|^\rho\big) \, dr \\ &\quad \le C\big(|t - s|^{\rho - 1} + |t - s|^{\rho/2 - 1}\big)|t - s|\Big(1 + \mathbb{E}\Big| \sup_{r \in [0, T]} \overline{X}_r \Big|^\rho\Big) \\ &\quad \le C|t - s|^{\rho/2 - \epsilon}\Big[C_T\Big(1 + \mathbb{E}\Big[\sup_{r \in [0, T]} |\overline{X}_r|^\rho\Big]\Big)\Big]. \end{aligned}$$

The constant C_T is such that $C_T \to 0$, as $T \to 0$. We apply Lemma 4.7 to obtain the following estimate:

$$\mathbb{E}\big[|\delta(X - F)_{s,t}|^\rho\big] \quad \le C_T|t - s|^{\rho/2 - \epsilon}\big(1 + \|\overline{X}_0\|_{L_\omega^\rho}\big),$$

for some $\epsilon > 0$ arbitrarily small. Using the Kolmogorov continuity theorem we obtain that $X - F$ is α-Hölder continuous for $\rho > 2/(1 - 2\alpha)$ and

$$\big\| [X]_\alpha \big\|_{L_\omega^\rho} \le \big\| [X - F]_\alpha \big\|_{L_\omega^\rho} + C[F]_\alpha \le C_T\big(1 + \|\overline{X}_0\|_{L_\omega^\rho}\big) + C[F]_\alpha.$$

We now prove inequality (60). Arguing as in the first half of the proof we obtain

$$\begin{aligned} &\mathbb{E}\big|\delta\big(X(F, X_0) - X(G, Y_0) - (F - G)\big)_{s,t}\big|^\rho \\ &\quad \le C\big(|t - s|^{\rho - 1} + |t - s|^{\rho/2 - 1}\big) \\ &\qquad \cdot \int_s^t \big(1 + E|\overline{X}_r|^\rho\big)\big(\mathbb{E}|X_r(F, X_0) - X_r(G, Y_0)|^\rho \\ &\qquad + W_\rho\big(\mathcal{L}(X_r(F, X_0)), \mathcal{L}(X_r(G, Y_0))\big)^\rho\big) \, dr. \end{aligned}$$

We apply Lemma 4.7 and (58) and use the fact that the L^ρ-norm of the difference of the process controls the ρ-Wasserstein distance, to obtain

$$\mathbb{E}\big|\delta\big(X(F, X_0) - X(G, Y_0) - F + G\big)_{s,t}\big|^\rho$$

$$\leq C_T|t - s|^{\alpha\rho}e^{C_T(1+\|\overline{X}_0\|_{L_\omega^\rho})}\big(\|F - G\|_{C_b} + \|X_0 - Y_0\|_{L_\omega^\rho}\big)^\rho.$$

Equation (60) follows as before from the Kolmogorov continuity theorem, for $\rho > 2/(1 - 2\alpha)$. $\square$

LEMMA 4.11.   *Let $F \in C_b([0, T]; \mathbb{R}^D)$ and $X_0 \in L_\omega^2$. Let b and σ satisfy Assumption 3. Then equation* (54) *admits a unique strong solution.*

PROOF.   To prove existence we define

$$\tilde{b}(t, x, \mu) := b\big(x + F_t, (\tau_{F_t})_\#\mu\big), \quad t \in [0, T], x \in \mathbb{R}^D, \mu \in \mathcal{P}(\mathbb{R}^D),$$

where $\tau_z : \mathbb{R}^D \to \mathbb{R}^D$ is the translation by $z \in \mathbb{R}^D$. Similarly we define $\tilde{\sigma}$. It is immediate to see that, if b, σ satisfy Assumption 3, then $\tilde{b}, \tilde{\sigma}$ satisfy Assumption 1.

By Lemma 4.2 the following equation admits a unique global strong solution $\tilde{X}$:

$$d\tilde{X}_t = \tilde{b}\big(t, \tilde{X}_t, \mathcal{L}(\tilde{X}_t)\big)\,dt + \tilde{\sigma}\big(t, \tilde{X}_t, \mathcal{L}(\tilde{X}_t)\big)\,dW_t, \qquad \tilde{X}_t|_{t=0} = X_0 - F_0.$$

The stochastic process $X = \tilde{X} + F$ solves equation (54). Uniqueness follows from the a priori estimates given in Lemma 4.9. $\square$

4.4. *Rough McKean–Vlasov.*

ASSUMPTION 4.   Let $h \in C_b^2(\mathbb{R}^D, \mathbb{R}^d)$.

For a given probability measure $\pi \in \mathcal{P}(\mathbb{R}^D)$, we define for $1 \leq l \leq D, 1 \leq j \leq d$,

$$(61) \qquad P(\pi) := \mathrm{Cov}_\pi(x, h)A + B, \qquad \mathrm{Cov}_\pi(x, h)^{l,j} = \int_{\mathbb{R}^D}(x^l - \pi[x^l])h^j(x)\pi(dx),$$

where $A \in \mathbb{R}^{d\times d}$ and $B \in \mathbb{R}^{D\times d}$ are given matrices. Notice that in the definition of the covariance we can choose to centre either one or both of the variables without changing the result. Indeed, in other sections of the paper we work with different formulations.

Let $\mathbf{Y} := (Y, \mathbb{Y}) \in \mathscr{C}^\alpha([0, T]; \mathbb{R}^d)$ and $X_0 \in L_\omega^2$. We study the following equation:

$$(62a) \qquad dX_t = b\big(X_t, \mathcal{L}(X_t)\big)\,dt + \sigma\big(X_t, \mathcal{L}(X_t)\big)\,dW_t + dF_t, \qquad X_t|_{t=0} = X_0,$$

$$(62b) \qquad dF_t = P\big(\mathcal{L}(X_t)\big)\,d\mathbf{Y}_t.$$

DEFINITION 4.12.   A couple $(X, F) : [0, T] \times \Omega \to \mathbb{R}^D \times \mathbb{R}^D$ is a solution to equation (62) if:

- F is a continuous path and X solves equation (62a) in the sense of Definition 4.5;
- $P(\mathcal{L}(X_t)) \in \mathscr{D}_\mathbf{Y}^{2\alpha}([0, T]; \mathscr{L}(\mathbb{R}^d, \mathbb{R}^D))$ and F is the rough integral in (62b).

Next we prove that, if X solves (62a) for a fixed controlled F, then $P(\mathcal{L}(X_t))$ is a controlled path, which makes the rough integral (62b) well defined.

LEMMA 4.13. *Let* $\alpha \in (\frac{1}{3}, \frac{1}{2})$ *and* $\rho > 2/(1 - 2\alpha)$. *Let* $\mathbf{Y} := (Y, \mathbb{Y}) \in \mathscr{C}^\alpha([0, T]; \mathbb{R}^d)$, $F \in \mathscr{D}_{\mathbf{Y}}^{2\alpha}([0, T]; \mathbb{R}^D)$ *and* $X_0 \in L_\omega^\rho(\mathbb{R}^D)$.

If π_t *is the law of the solution process* $X_t(F, X_0)$ *to equation* (54) *with forcing* F *and initial condition* X_0, *then*

$$P(\pi.) \in \mathscr{D}_{\mathbf{Y}}^{2\alpha}([0, T]; \mathscr{L}(\mathbb{R}^d, \mathbb{R}^D)),$$

with Gubinelli derivative

(63)
$$P(\pi_s)' = (F_s')^\top \pi [(x - \pi[x]) Dh^\top] A.$$

Moreover, we have the bound

$$\|P(\pi.)\|_{\mathscr{D}_{\mathbf{Y}}^{2\alpha}} \le C \|h\|_{C_b^2} (1 + \|\overline{X}_0\|_{L_\omega^\rho} + [F]_\alpha)(C_T(1 + \|\overline{X}_0\|_{L_\omega^\rho}) + \|F'\|_{C_b}[Y]_\alpha + [R^F]_{2\alpha}),$$

where $C_T \to 0$ *as* $T \to 0$.

PROOF. To simplify the notation, we write X_t for $X_t(F, X_0)$. Let $\overline{X}_t := X_t - \mathbb{E}[X_t]$. We have the following expansion for $h(X.)$, $s, t \in [0, T]$:

(64)
$$\delta h(X.)_{s,t} = [Dh]_{s,t}^{1,x} \left(\int_s^t b(X_r, \mathcal{L}(X_r)) \, dr + R_{s,t}^F \right)$$
$$+ ([D^2 h]_{s,t}^{2,x} \delta X_{s,t} + Dh(X_s)) \left(\int_s^t \sigma(X_r, \mathcal{L}(X_r)) \, dW_r + F_s' \delta Y_{s,t} \right).$$

Let $s, t \in [0, T]$. From the definition of $P(\pi)$, equation (61), we have

$$\delta P(\pi.)_{s,t} = \delta \mathbb{E}[\overline{X}. h(X.)^\top]_{s,t} A$$
$$= \mathbb{E}[\delta \overline{X}_{s,t} h(X_s)^\top] A + \mathbb{E}[\delta \overline{X}_{s,t} \delta h(X.)_{s,t}^\top] A$$
$$+ \mathbb{E}[\overline{X}_s \delta h(X.)_{s,t}^\top] A$$
$$=: I_1 + I_2 + I_3.$$

We expand I_3 even further using (64)

$$I_3 = \mathbb{E}[\overline{X}_s (Dh(X_s) F_s' \delta Y_{s,t})^\top] A + \mathbb{E}\left[\overline{X}_s \left(Dh(X_s) \int_s^t \sigma(X_r, \mathcal{L}(X_r)) \, dW_r \right)^\top\right] A$$

$$+ \mathbb{E}\left[\overline{X}_s \left([Dh]_{s,t}^{1,x} \left(\int_s^t b(X_r, \mathcal{L}(X_r)) \, dr + R_{s,t}^F \right)\right)^\top \right]$$

$$+ \mathbb{E}\left[\overline{X}_s \left([D^2 h]_{s,t}^{2,x} \delta X_{s,t} \left(\int_s^t \sigma(X_r, \mathcal{L}(X_r)) \, dW_r + F_s' \delta Y_{s,t} \right)\right)^\top \right] A$$

$$=: \mathbb{E}[\overline{X}_s (Dh(X_s) F_s' \delta Y_{s,t})^\top] A + I_4.$$

We can write

$$\delta P(\pi.)_{s,t} = \mathbb{E}[\overline{X}_s (Dh(X_s) F_s' \delta Y_{s,t})^\top] A + R_{s,t}^P,$$

where $R_{s,t}^P = I_1 + I_2 + I_4$. For $1 \le l \le D$ and $1 \le j \le d$ we write the (l, j) entry of the matrix P as

$$\delta P^{l,j}(\pi.)_{s,t} = (P^{l,j}(\pi.))_s' \cdot \delta Y_{s,t} + R_{s,t}^{P^{l,j}},$$

where, for $s \in [0, T]$, $P(\pi_s)' \in \mathscr{L}(\mathbb{R}^d, \mathscr{L}(\mathbb{R}^d, \mathbb{R}^D)) \cong \mathbb{R}^{D \times d \times d}$ is the Gubinelli derivative given as

$$(P^{l,j}(\pi_s))' := \sum_{k=1}^D \mathbb{E}[\overline{X}_s^l (Dh(X_s)^\top A)^j](F_s')^k = \sum_{k=1}^D \sum_{i=1}^d \mathbb{E}[\overline{X}_s^l \partial_k h^i(X_s)] A^{i,j} (F_s')^k \in \mathbb{R}^d.$$

We check now that R^P is a remainder with regularity 2α. Using the martingale property of the stochastic integral in (55), we have

$$I_1 = \mathbb{E}\big[\mathbb{E}_s[\delta\overline{X}_{s,t}]h(X_s)^\top\big]A = \mathbb{E}\bigg[\int_s^t \big(b(X_r, \mathcal{L}(X_r)) - \mathbb{E}[b(X_r, \mathcal{L}(X_r))]\big)\,\mathrm{d}r\,h(X_s)^\top\bigg]A.$$

We apply Lemma 4.7 to obtain

$$|I_1| \le |t - s|C\|h\|_\infty\big(C_T + \|X_0\|_{L_\omega^\rho}\big),$$

where $C > 0$ is a global constant. For I_2 we use Lemma 4.7 and (59) to obtain, for any $\rho > 3$,

$$|I_2| = \big|\mathbb{E}[\delta\overline{X}_{s,t}([Dh]_{s,t}^{1;x}\delta X_{s,t})^\top]\big| \le \|Dh\|_\infty\|[\overline{X}]_\alpha\|_{L_\omega^\rho}\|[X]_\alpha\|_{L_\omega^\rho}$$

$$\le \|Dh\|_\infty C_T\big(1 + \|\overline{X}_0\|_{L_\omega^\rho}\big)\big[C_T\big(1 + \|\overline{X}_0\|_{L_\omega^\rho}\big) + [F]_\alpha\big]|t - s|^{2\alpha}.$$

We proceed by finding a bound for I_4. Notice that the term $E[\overline{X}_s(Dh(X_s)\int_s^t \sigma(X_r, \mathcal{L}(X_r)\,\mathrm{d}W_r)^\top]$ vanishes thanks to the martingale property of the stochastic integral. Using Lemma 4.7 and inequality (59), we get

$$I_4 \le C|t - s|^{2\alpha}\|Dh\|_\infty\Big\|\sup_{t\in[0,T]}\overline{X}_t\Big\|_{L_\omega^\rho}\Big(C_T\Big(1 + \Big\|\sup_{t\in[0,T]}\overline{X}_t\Big\|_{L_\omega^\rho}\Big) + [R^F]_{2\alpha}\Big)$$

$$+ C|t - s|^{2\alpha}\|D^2h\|_\infty\|[X]_\alpha\|_{L_\omega^\rho}\Big(C_T\Big(1 + \Big\|\sup_{t\in[0,T]}\overline{X}_t\Big\|_{L_\omega^\rho}\Big) + \|F'\|_{C_b}[Y]_\alpha\Big)$$

$$\le C|t - s|^{2\alpha}\|Dh\|_\infty\big(C_T + \|\overline{X}_0\|_{L_\omega^\rho}\big)\big(C_T\big(1 + \|\overline{X}_0\|_{L_\omega^\rho}\big) + [R^F]_{2\alpha}\big)$$

$$+ C|t - s|^{2\alpha}\|D^2h\|_\infty\big(1 + \|\overline{X}_0\|_{L_\omega^\rho} + [F]_\alpha\big)\big(C_T\big(1 + \|\overline{X}_0\|_{L_\omega^\rho}\big) + \|F'\|_{C_b}[Y]_\alpha\big)$$

$$\le C|t - s|^{2\alpha}\big(\|Dh\|_\infty + \|D^2h\|_\infty\big)\big(1 + \|\overline{X}_0\|_{L_\omega^\rho} + [F]_\alpha\big)$$

$$\cdot\big(C_T\big(1 + \|\overline{X}_0\|_{L_\omega^\rho}\big) + \|F'\|_{C_b}[Y]_\alpha + [R^F]_{2\alpha}\big). \qquad \square$$

We now set up a contraction argument that we will use to prove the well-posedness of equation (62). Let $f \in \mathscr{D}_\mathbf{Y}^{2\alpha}([0, T]; \mathcal{L}(\mathbb{R}^d, \mathbb{R}^D))$. We define, for $t \in [0, T]$,

$$(65) \qquad (F_t, F_t') := \bigg(\int_0^t f_r\,\mathrm{d}\mathbf{Y}_r, f_t\bigg) \in \mathscr{D}_\mathbf{Y}^{2\alpha}([0, T]; \mathbb{R}^D).$$

Now we plug this as the forcing into equation (54) with initial condition $X_0 \in L_\omega^\rho(\mathbb{R}^D)$ and call X the solution with law π. We define the map

$$(66) \qquad \begin{array}{ccc} \Gamma : \mathscr{D}_\mathbf{Y}^{2\alpha}([0, T]; \mathcal{L}(\mathbb{R}^d, \mathbb{R}^D)) & \to & \mathscr{D}_\mathbf{Y}^{2\alpha}([0, T]; \mathcal{L}(\mathbb{R}^d, \mathbb{R}^D)) \\ f & \mapsto & P(\pi.). \end{array}$$

Using the following estimate:

$$f_t = f_t - f_0 + f_0 = f_0'Y_{0t} + R_{0t}^f + f_0 \le |f_0'|[Y]_\alpha t^\alpha + |f_0| + t^{2\alpha}[R^f]_{2\alpha}$$

we get

$$\|f\|_{C_b} \le C\big(1 + \|\mathbf{Y}\|_\alpha\big)\big(|f_0| + |f_0'| + \|f\|_{\mathscr{D}_\mathbf{Y}^{2\alpha}}\big).$$

From [53], Theorem 4.10, we have the following estimates:

$$[R^F]_{2\alpha} \le C\|\mathbf{Y}\|_\alpha\big(|f_0| + |f_0'| + \|f\|_{\mathscr{D}_\mathbf{Y}^{2\alpha}}\big),$$

$$[F]_\alpha, \|F\|_{\mathscr{D}_\mathbf{Y}^{2\alpha}} \le C\big(1 + \|\mathbf{Y}\|_\alpha\big)\big(|f_0| + |f_0'| + \|f\|_{\mathscr{D}_\mathbf{Y}^{2\alpha}}\big).$$

It follows from Lemma 4.13 that Γ is well defined. Let us prove that there exists a small time, such that the following set is invariant under Γ:

$$D(\Gamma) := \{f \in \mathscr{D}_{\mathbf{Y}}^{2\alpha} \mid |f_0| + |f_0'| + \|f\|_{\mathscr{D}_{\mathbf{Y}}^{2\alpha}} \le 2\|h\|_\infty M\},$$

where M is the global constant of Remark 4.8.

If we assume that $Y \in \mathscr{C}^{\bar{\alpha}}$ with $\alpha < \bar{\alpha} < 1/2$, we have the following estimates:

$$|\Gamma(f)_0| = |\mathbb{E}[\overline{X}_0 h(X_0)]A + B| \le \|h\|_\infty M,$$

$$\|\Gamma(f)\|_{\mathscr{D}_{\mathbf{Y}}^{2\alpha}} = \|P(\pi.)\|_{\mathscr{D}_{\mathbf{Y}}^{2\alpha}}$$

$$\le C\|h\|_{C_b^2}(1 + \|\overline{X}_0\|_{L_\omega^\rho} + \|\mathbf{Y}\|_\alpha(|f_0| + |f_0'| + \|f\|_{\mathscr{D}_{\mathbf{Y}}^{2\alpha}}))$$

$$\cdot (C_T(1 + \|\overline{X}_0\|_{L_\omega^\rho}) + \|\mathbf{Y}\|_\alpha(|f_0| + |f_0'| + \|f\|_{\mathscr{D}_{\mathbf{Y}}^{2\alpha}}))$$

$$\le \|h\|_{C_b^2} C_T(1 + M + 2\|h\|_\infty(\|\mathbf{Y}\|_\alpha \vee \|\mathbf{Y}\|_{\bar{\alpha}}))^2$$

$$\le \|h\|_\infty M,$$

where in the last step we chose T small enough such that

$$C_T \le \frac{\|h\|_\infty M}{\|h\|_{C_b^2}(1 + M + 2M\|h\|_\infty(\|\mathbf{Y}\|_\alpha \vee \|\mathbf{Y}\|_{\bar{\alpha}}))^2}.$$

Adding the previous estimates we proved that $\Gamma(D(\Gamma)) \subset D(\Gamma)$. Since C_T only depends on global quantities, we can divide each interval $[0, \overline{T}]$ into smaller intervals of length T and in each one apply the contraction argument. We prove in the following that Γ is a contraction on $D(\Gamma)$. We start by showing that Γ is Lipschitz continuous.

LEMMA 4.14. *Let $1/3 < \alpha < 1/2$. Let $\rho > 2/(1-2\alpha)$ and $X_0, Z_0 \in L_\omega^\rho$. Let $\mathbf{Y}, \hat{\mathbf{Y}} \in \mathscr{C}^{\bar{\alpha}}$ and $F \in \mathscr{D}_{\mathbf{Y}}^{2\alpha}, G \in \mathscr{D}_{\hat{\mathbf{Y}}}^{2\alpha}$. Assume that there exists a universal constant $C > 0$ such that*

$$\|F\|_{\mathscr{D}_{\mathbf{Y}}^{2\alpha}} \le C\|\mathbf{Y}\|_\alpha, \qquad \|G\|_{\mathscr{D}_{\hat{\mathbf{Y}}}^{2\alpha}} \le C\|\hat{\mathbf{Y}}\|_\alpha, \qquad \|\overline{X}_0\|_{L_\omega^\rho}, \|\overline{Z}_0\|_{L_\omega^\rho} \le C,$$

where $\overline{X}_0 = X_0 - \mathbb{E}[X_0]$ and $\overline{Z}_0 = X_0 - \mathbb{E}[Z_0]$. We call $X_t(F, X_0)$ (resp. $X_t(G, Z_0)$) the solution to equation (54) *with inputs F and X_0 (resp. G and Z_0). We call R the difference of the remainders of $P(\pi^x)$ and $P(\pi^z)$. There exists a global constant $C_T > 0$ such that $C_T \to 0$ as $T \to 0$ and*

$$(67) \quad \|P(\pi^x), P(\pi^z)\|_{Y, \hat{Y}, 2\alpha}$$

$$(68) \quad \begin{aligned} &\le (C_T + \|\mathbf{Y}\|_\alpha)(\|X_0 - Z_0\|_{L_\omega^\rho} + \|F - G\|_{C_b} + [F - G]_\alpha + \|F' - G'\|_{C_b} \\ &\quad + [F' - G']_\alpha) + \|G'\|_{C_b}[\mathbf{Y} - \hat{\mathbf{Y}}]_\alpha. \end{aligned}$$

A similar estimate can be obtained for the difference of the Gubinelli derivatives.

PROOF. To simplify the notation, we write X_t for $X(F, X_0)_t$ and Z_t for $X(G, Z_0)_t$. Let $\overline{X}_t := X_t - E[X_t]$, similarly we define $\overline{Z}_t$. We have the following expansion for $h(X.) - h(Z.)$, $s, t \in [0, T]$, $\mathbb{P}$-a.s.:

$$(69a) \quad \begin{aligned} &\delta(h(X.) - h(Z.))_{s,t} \\ &= Dh(X_s)F'\delta Y_{s,t} - Dh(Z_s)G_s'\delta\hat{Y}_{s,t} \end{aligned}$$

$$(69\text{b}) \qquad \begin{aligned} &+ Dh(X_s)\left(\int_s^t \sigma(X_r, \mathcal{L}(X_r))\,\mathrm{d}W_r\right) \\ &- Dh(Z_s)\left(\int_s^t \sigma(Z_r, \mathcal{L}(Z_r))\,\mathrm{d}W_r\right) \end{aligned}$$

$$(69\text{c}) \qquad \begin{aligned} &+ [Dh]_{s,t}^{1;z}\left(\int_s^t [b(X_r, \mathcal{L}(X_r)) - b(Z_r, \mathcal{L}(Z_r))]\,\mathrm{d}r \right.\\ &\left.+ R_{s,t}^F - R_{s,t}^G\right) \end{aligned}$$

$$(69\text{d}) \qquad + ([Dh]_{s,t}^{1;x} - [Dh]_{s,t}^{1;z})\left(\int_s^t b(X_r, \mathcal{L}(X_r))\,\mathrm{d}r + R_{s,t}^F\right)$$

$$(69\text{e}) \qquad + [D^2h]_{s,t}^{2;z}\delta Z_{s,t}\left(\int_s^t [\sigma(X_r, \mathcal{L}(X_r)) - \sigma(Z_r, \mathcal{L}(Z_r))]\,\mathrm{d}W_r\right)$$

$$(69\text{f}) \qquad + [D^2h]_{s,t}^{2;z}\delta Z_{s,t}(F_s'\delta Y_{s,t} - G_s'\delta \hat{Y}_{s,t})$$

$$(69\text{g}) \qquad \begin{aligned} &+ ([D^2h]_{s,t}^{2;z}\delta(X-Z)_{s,t} + ([D^2h]_{s,t}^{2;x} - [D^2h]_{s,t}^{2;z})\delta X_{s,t}) \\ &\cdot \left(\int_s^t \sigma(X_r, \mathcal{L}(X_r))\,\mathrm{d}W_r + F_s'\delta Y_{s,t}\right). \end{aligned}$$

Let $s, t \in [0, T]$. We call $\pi^x := \mathcal{L}(X)$ and $\pi^z := \mathcal{L}(Z)$. From the definition of $P(\pi)$, equation (61), we have

$$\begin{aligned} &\delta\big(P(\pi_{\cdot}^x) - P(\pi_{\cdot}^z)\big)_{s,t} \\ &= \mathbb{E}[\delta(\overline{X}_{\cdot} - \overline{Z}_{\cdot})_{s,t}h(Z_s)^\top]A + \mathbb{E}[\delta(\overline{X}_{\cdot} - \overline{Z}_{\cdot})_{s,t}\delta h(Z_{\cdot})_{s,t}^\top]A \\ &\quad + \mathbb{E}[(\overline{X}_s - \overline{Z}_s)\delta h(Z_{\cdot})_{s,t}^\top]A + \mathbb{E}[\delta\overline{X}_{s,t}(h(X_s) - h(Z_s))]A \\ &\quad + \mathbb{E}[\delta\overline{X}_{s,t}\delta(h(X_{\cdot}) - h(Z_{\cdot}))_{s,t}]A + \mathbb{E}[\overline{X}_s\delta(h(X_{\cdot}) - h(Z_{\cdot}))_{s,t}]A \\ &=: I_{11} + I_{12} + I_{13} + I_{21} + I_{22} + I_{23}. \end{aligned}$$

Using the same expansion for $h(Z)$ as in (64), we obtain

$$\begin{aligned} I_{13} &= \mathbb{E}[(\overline{X}_s - \overline{Z}_s)(Dh(Z_s)G_s'\delta\hat{Y}_{s,t})^\top]A + 0 \\ &\quad + \mathbb{E}\left[(\overline{X}_s - \overline{Z}_s)\left([Dh]_{s,t}^{1;z}\left(\int_s^t b(Z_r, \mathcal{L}(Z_r))\,\mathrm{d}r + R_{s,t}^G\right)\right)^\top\right]A \\ &\quad + \mathbb{E}\left[(\overline{X}_s - \overline{Z}_s)\left([D^2h]_{s,t}^{2;z}\delta Z_{s,t}\left(\int_s^t \sigma(Z_r, \mathcal{L}(Z_r))\,\mathrm{d}W_r + G_s'\delta\hat{Y}_s\right)\right)^\top\right]A \\ &= \mathbb{E}[(\overline{X}_s - \overline{Z}_s)(Dh(Z_s)G_s'\delta\hat{Y}_{s,t})^\top]A + I_{14} + I_{15}. \end{aligned}$$

The second term in the first line vanishes thanks to the martingale property of the stochastic integral. Similarly, we expand I_{23} using (69). Notice that the term (69b) vanishes because the stochastic integral is a martingale. The term (69a) produces a term of regularity α, the others have regularity 2α.

$$I_{23} = \mathbb{E}[\overline{X}_s(Dh(X_s)F'\delta Y_{s,t} - Dh(Z_s)G_s'\delta\hat{Y}_{s,t})^\top]A + I_{24},$$

where $I_{24} := \mathbb{E}[\overline{X}_s((69\text{c}) + \cdots + (69\text{g}))]A$. We thus obtain that

$$\begin{aligned} \delta\big(P(\pi_{\cdot}^x) - P(\pi_{\cdot}^z)\big)_{s,t} &= \mathbb{E}[(\overline{X}_s - \overline{Z}_s)(Dh(Z_s)G_s'\delta\hat{Y}_{s,t})^\top]A \\ &\quad + \mathbb{E}[\overline{X}_s(Dh(X_s)F'\delta Y_{s,t} - Dh(Z_s)G_s'\delta\hat{Y}_{s,t})]A + R_{s,t} \end{aligned}$$

$$= \mathbb{E}\big[\overline{X}_s \big(Dh(X_s)F'\delta Y_{s,t}\big)^\top\big]A - \mathbb{E}\big[\overline{Z}_s \big(Dh(Z_s)G'_s\delta\hat{Y}_{s,t}\big)^\top\big]A + R_{s,t}$$

$$= P\big(\pi_s^x\big)'\delta Y_{s,t} + P\big(\pi_s^z\big)'\delta\hat{Y}_{s,t} + R_{s,t}$$

with $R_{s,t} = I_{11} + I_{12} + I_{14} + I_{15} + I_{21} + I_{22} + I_{24}$. For $1 \le l \le D$ and $1 \le j \le d$ we write the (l, j) entry of the matrix as

$$\delta\big(P(\pi^x) - P(\pi^z)\big)_{s,t}^{l,j} = \sum_{k=1}^{D} \mathbb{E}\big[\overline{X}_s^l (Dh(X_s)^\top A)^{i,k}(F'_s)^k\big] \cdot \delta Y_{s,t}$$

$$- \sum_{k=1}^{D} \mathbb{E}\big[\overline{Z}_s^l (Dh(Z_s)^\top A)^{i,k}(G'_s)^k\big] \cdot \delta\hat{Y}_{s,t} + R_{s,t}^{P^{l,j}}.$$

We must now find estimates for $[R]_{2\alpha}$. We start by some estimates of the processes in L^ρ. From equation (58) we have

$$(70) \qquad \Big\| \sup_{t \in [0,T]} |X_t - Z_t| \Big\|_{L_\omega^\rho} \le C\big(\|F - G\|_{C_b} + \|X_0 - Z_0\|_{L_\omega^\rho}\big).$$

It follows from the equations for $\overline{X}$ and $\overline{Z}$ as well as estimate (70) that

$$\big\|[\overline{X} - \overline{Z}]_\alpha\big\|_{L_\omega^\rho} \le C_T\big(1 + \|\overline{X}_0\|_{L_\omega^\rho}\big)\Big\| \sup_{t \in [0,T]} (X_t - Z_t)\Big\|_{L_\omega^\rho}$$

$$\le C_T\big(\|F - G\|_{C_b} + \|X_0 - Z_0\|_{L_\omega^\rho}\big).$$

Using Lemma 4.7 we also have

$$\big\|[\overline{X} - \overline{Z}]_\alpha\big\|_{L_\omega^\rho} \le \big\|[\overline{X}]_\alpha\big\|_{L_\omega^\rho} + \big\|[\overline{Z}]_\alpha\big\|_{L_\omega^\rho} \le C_T.$$

From (60) we have

$$\big\|[X - Z]_\alpha\big\|_{L_\omega^\rho} \le C_T\big(\|F - G\|_{C_b} + \|X_0 - Z_0\|_{L_\omega^\rho}\big) + C[F - G]_\alpha.$$

From (59) we obtain

$$\big\|[X]_\alpha\big\|_{L_\omega^\rho}, \big\|[Z]_\alpha\big\|_{L_\omega^\rho} \le C_T.$$

Using the previous estimates we obtain

$$\delta(I_{11} + I_{21})_{s,t} \le C_T \|h\|_{C_b^1}\Big\| \sup_{t \in [0,T]} |X_t - Z_t|\Big\|_{L_\omega^\rho}|t - s|^{2\alpha},$$

$$\delta(I_{12} + I_{22})_{s,t} \le C_T \|h\|_{C_b^1}\big\|[X - Z]_\alpha\big\|_{L_\omega^\rho}|t - s|^{2\alpha},$$

$$\delta(I_{14} + I_{15})_{s,t} \le \|h\|_{C_b^2}\Big\|\Big[\sup_{t \in [0,T]} |X_t - Z_t|\Big\|_{L_\omega^\rho}\big(C_T + [R^G]_{2\alpha} + \|G'\|_{C_b}[\hat{Y}]_\alpha\big)|t - s|^{2\alpha}.$$

For I_{24} we have a combination of the above and an extra term

$$[I_{24}]_{2\alpha} \le \cdots + \big\|F' - G'\big\|_{C_b}[\mathbf{Y}]_\alpha + \big\|G'\big\|_{C_b}[\mathbf{Y} - \hat{\mathbf{Y}}]_\alpha.$$

Summing up, we get

$$[R]_{2\alpha} \le C\big(\|F - G\|_{C_b} + \|X_0 - Y_0\|_{L_\omega^\rho} + [F - G]_\alpha\big)\big(C_T + [R^G]_{2\alpha} + \big\|G'\big\|_{C_b}[\hat{Y}]_\alpha\big)$$

$$+ \big\|F' - G'\big\|_{C_b}[\mathbf{Y}]_\alpha + \big\|G'\big\|_{C_b}[\mathbf{Y} - \hat{\mathbf{Y}}]_\alpha,$$

which gives the desired estimate using the assumption $\|G\|_{\mathscr{D}_{\hat{\mathbf{Y}}}^{2\alpha}} \le C\|\hat{\mathbf{Y}}\|_\alpha$.

We now look at the difference between the Gubinelli derivatives. From (63) we get

$$\delta\big(P(\pi^x) - P(\pi^z)\big)_{s,t} = \delta\mathbb{E}\big[(\overline{X}_\cdot - \overline{Z}_\cdot)Dh(X_\cdot)F_\cdot'\big]_{s,t} + \delta\mathbb{E}\big[\overline{Z}_\cdot(Dh(X_\cdot) - Dh(Z_\cdot))F_\cdot'\big]_{s,t}$$
$$+ \delta\mathbb{E}\big[\overline{Z}_\cdot Dh(Z_\cdot)(F_\cdot' - G_\cdot')\big]_{s,t}.$$

By analysing each term independently and with similar reasoning as in the case of the remainder, we obtain

$$\big[P(\pi^x) - P(\pi^z)\big]_\alpha \leq (C_T + \|\mathbf{Y}\|_\alpha)\big(\|F - G\|_{C_b} + \|X_0 - Z_0\|_{L^\rho} + \big[F' - G'\big]_\alpha\big).$$

Estimate (67) is obained by putting together the estimates on the remainder and the Gubinelli derivatives. $\square$

We are now ready to prove the main well-posedness result.

THEOREM 4.15. *Let* $1/3 < \alpha < \bar\alpha < 1/2$. *Let* $\rho > 2/(1 - 2\alpha)$ *and* $X_0, Z_0 \in L^\rho$. *Let* $\mathbf{Y} \in \mathscr{C}^{\bar\alpha}([0, T]; \mathbb{R}^d)$. *If* b, σ *and* h *satisfy Assumptions* 3 *and* 4, *then equation* (62) *admits a unique solution* $(X, F) \in L^\rho(C([0, T]; \mathbb{R}^D)) \times \mathscr{D}_{\mathbf{Y}}^{2\alpha}([0, T]; \mathbb{R}^D)$ *in the sense of Definition* 4.12.

PROOF. If X, F solves equation (62), then necessarily $F \in C([0, T]; \mathbb{R}^D)$ and we can apply the a priori bound in Lemma 4.9 to see that $X \in L^\rho(C([0, T]; \mathbb{R}^d))$. A process $X : \Omega \to C([0, T]; \mathbb{R}^D)$ is a solution to equation (62) if and only if $P(\mathcal{L}(X))$ is a fixed point of Γ, defined in (66).

We want to prove that $\Gamma : D(\Gamma) \to D(\Gamma)$ is a contraction. If $f, g \in D(\Gamma)$ we have that F, G defined as in (65) satisfy the assumptions of Lemma 4.14 with a global constant $C = C(h, M)$. By taking $\mathbf{Y} = \hat{\mathbf{Y}}$ and $X_0 = Z_0$ the estimate in Lemma 4.14 reduces to

$$\|\Gamma(f) - \Gamma(g)\|_{\mathscr{D}_{\mathbf{Y}}^{2\alpha}} \leq C_T(M, \|\mathbf{Y}\|_{\bar\alpha})\big(\|f - g\|_{\mathscr{D}_{\mathbf{Y}}^{2\alpha}} + \|f_0 - g_0\|_{C_b}\big), \quad f, g \in D(\Gamma).$$

Choosing now T_0 small enough such that $C_{T_0} < 1$, we have that Γ is a contraction on the closed subset $D(\Gamma) \subset \mathscr{D}_{\mathbf{Y}}^{2\alpha}([0, T_0], \mathcal{L}(\mathbb{R}^d, \mathbb{R}^D))$. Hence, Γ admits a unique fixed point.

Since the small constant C_{T_0} in the definition of the domain $D(\Gamma)$ of Γ and in the contraction argument only depends on the global quantities h, M, $\mathbf{Y}$ and not on the initial condition X_0, we can construct a finite family of subsequent time intervals of size T_0 that covers $[0, T]$. On each of these time interval we construct the solution as a fixed point of Γ. $\square$

COROLLARY 4.16. *Let* $1/3 < \bar\alpha < 1/2$ *and* $\rho > 2/(1 - 2\bar\alpha)$. *Let* $X_0 \in L^\rho$ *and* $\mathbf{Y}^1, \mathbf{Y}^2 \in \mathscr{C}^{\bar\alpha}$. *For* $i = 1, 2$, *let* X^i *be the solution to equation* (62) *with driver* $\mathbf{Y}^i$ *and initial condition* X_0. *Call* $\pi^i := \mathcal{L}(X^i) \in \mathcal{P}_\rho(C([0, T]; \mathbb{R}^D))$. *There exists a positive constant* $C > 0$ *such that*

$$W_{\rho, C([0, T]; \mathbb{R}^D)}(\pi^1, \pi^2) \leq C\rho_{\bar\alpha}(\mathbf{Y}^1, \mathbf{Y}^2).$$

PROOF. It follows from Lemma 4.9 and Lemma 4.14 that the solution map $\mathbf{Y} \mapsto X$ of equation (62) is a Lipschitz-continuous function between the spaces $\mathscr{C}^{\bar\alpha}([0, T]; \mathbb{R}^d)$ and $L^\rho(C([0, T]; \mathbb{R}^D))$. The corollary follows as the Wasserstein distance between π^1 and π^2 is always less than or equal to the L^ρ distance between X^1 and X^2. $\square$

We now proceed to prove a Wong–Zakai type result when X is the solution to equation (62) driven by the Itô lift $\mathbf{Y}$ of a Brownian motion. We introduce the approximation X^n as the solution of

$$(71a) \qquad dX_t^n = b(X_t^n, \mathcal{L}(X_t^n))\, dt + \sigma(X_t^n, \mathcal{L}(X_t^n))\, dW_t + dF_t^n, \qquad X_t|_{t=0} = X_0,$$

$$(71b) \qquad \dot{F}_t^n = P(\mathcal{L}(X_t^n))\dot{Y}_t^n - \frac{1}{2}P(\mathcal{L}(X_t^n))^\top \mathbb{E}\big[\overline{X}_t^n Dh(X_t^n)^\top\big]A_\cdot,$$

where Y^n is a piecewise linear approximation of Y. We have the following result.

THEOREM 4.17.   *The solution X^n converges to X in the Wasserstein distance, viz $\mathbb{P}$-a.s. we have*

$$W_{\rho, C([0,T];\mathbb{R}^D)}(\mathcal{L}(X^n), \mathcal{L}(X)) \to 0$$

as $n \to \infty$.

PROOF.   We start by noting that X^n solves (71) if and only if it solves the rough path equation

$$dX_t^n = b(X_t^n, \mathcal{L}(X_t^n)) \, dt + \sigma(X_t^n, \mathcal{L}(X_t^n)) \, dW_t + dF_t^n, \qquad X_t|_{t=0} = X_0,$$

$$dF_t^n = P(\mathcal{L}(X_t^n)) \, d\mathbf{Y}_t^{n,\text{Itô}},$$

where

$$\mathbf{Y}_{st}^{n,\text{Itô}} = \left(\delta Y_{s,t}^n, \int_s^t \delta Y_{s,r}^n \otimes \dot{Y}_r^n \, dr - \frac{1}{2}(t-s) I_{d \times d} \right).$$

Indeed, the local approximation of F^n is

(72) $$F_{s,t}^n \approx P(\mathcal{L}(X_s^n)) \delta Y_{s,t}^n + P'(\mathcal{L}(X_s^n)) \left(\int_s^t \delta Y_{s,r}^n \otimes \dot{Y}_r^n \, dr - \frac{1}{2}(t-s) I_{d \times d} \right).$$

The first term in the right hand side contributes for the first term in the right hand side of (71b). The second term in the right hand side of (72) has Hölder regularity greater than 1 and vanishes when summing over $[s,t]$ in a partition of $[0,T]$ and sending the mesh of the partition to 0. Finally the third term in the right hand side of (72) contributes for the second term in the right hand side of (71b), indeed from (63) with $\pi_s = \mathcal{L}(X_s^n)$ and $F_s' = P(\mathcal{L}(X_s^n))$ we have

(73) $$\frac{1}{2}(t-s) P(\mathcal{L}(X_s^n))' I_{d \times d} = \frac{1}{2} P(\mathcal{L}(X_s^n))^\top \mathbb{E}[\overline{X}_s^n Dh(X_s^n)^\top] A(t-s).$$

It is well known that the canonical lift of Y^n,

$$\mathbf{Y}_{st}^{n,Str} = \left(\delta Y_{s,t}^n, \int_s^t \delta Y_{s,r}^n \otimes \dot{Y}_r^n \, dr \right),$$

converges $\mathbb{P}$-a.s. in the rough path topology to the rough path

$$\mathbf{Y}_{st}^{Str} = \left(\delta Y_{s,t}, \int_s^t \delta Y_{s,r} \otimes \text{o} dY_r \right),$$

where the latter integral is the Stratonovich integral. From this it follows immediately that

$$\mathbf{Y}_{st}^{n,\text{Itô}} \to \left(\delta Y_{s,t}, \int_s^t \delta Y_{s,r} \otimes \text{o} dY_r - \frac{1}{2}(t-s) I_{d \times d} \right) = \left(\delta Y_{s,t}, \int_s^t \delta Y_{s,r} \otimes dY_r \right),$$

where the latter integral is the Itô integral. The result now follows from Corollary 4.16.   □

4.5. *Proof of Theorem* 1.1.   We can now proceed with the proof of the main theorem.

Well-posedness of the rough stochastic McKean–Vlasov equation (11).   Given the coefficients of equation (11) we transform it into equation (62) by defining the following coefficients:

$$b(x, \pi) := f(x) - P(\pi)h(x),$$

$$\sigma(x, \pi) := \left(G^{\frac{1}{2}} - P(\pi)U, \quad -P(\pi)R^{\frac{1}{2}} \right),$$

$$P(\pi) := \text{Cov}_\pi(x, h)C^{-1} + B.$$

The m-Brownian motion W in (62) stands for the paired independent Brownian motions $(\widehat{W}, \widehat{V})$ of equation (11), with $m = D + d$. Moreover, if $\mathbf{Y} = (Y, \mathbb{Y})$ is any given rough path, we can modify it by a bounded variation term to obtain

$$\overline{\mathbf{Y}}_{s,t} := \left(Y_{s,t}, \mathbb{Y}_{s,t} - \frac{1}{2}(t-s)C\right).$$

We have that equation (62) driven by $\overline{\mathbf{Y}}$ corresponds to equation (11) driven by $\mathbf{Y}$ and the term $\frac{1}{2}(t-s)C$ generates the correction term $\Gamma(\pi)$. Hence, the proof of well-posedness and stability of equation (11) follows from Theorem 4.15 and Corollary 4.16.

Well-posedness of the McKean–Vlasov equation with common noise (10). The uniqueness of equation (10) follows from Lemma 4.3, as the coeffects of equation (10) satisfy Assumption 2. The following lemma gives existence.

LEMMA 4.18.   *Let $X^{\mathbf{y}}$ denote the solution of* (11) *and let $\mathbf{Y}(\omega)$ denote the Stratonovich rough path lift of the Brownian motion Y. Then $X^{\mathbf{Y}}$ is the solution of* (10) *and we have*

(74) $$\mathbb{E}[\phi(X_t^{\mathbf{Y}})|\mathcal{Y}_t](\omega) = \mathbb{E}[\phi(X_t^{\mathbf{y}})]\big|_{\mathbf{y}=\mathbf{Y}(\omega)|_{[0,t]}}, \quad \mathbb{P}_Y\text{-a.e. } \omega$$

for any $\phi \in C_b(\mathbb{R}^D, \mathbb{R})$.

PROOF.   Since the Stratonovich rough path lift of $Y|_{[0,t]}$ coincides with $(Y, \mathbb{Y})|_{[0,t]}$, we notice that by uniqueness, we have

$$X_t^{\mathbf{Y}} = X_t^{\mathbf{Y}|_{[0,t]}}.$$

Since $\mathbf{Y}|_{[0,t]}$ is independent of $\sigma(\widehat{W}_s, \widehat{V}_s, X_0, 0 \le s \le t)$, then it is also independent of $X_t^{\mathbf{y}}$ for every $\mathbf{y} \in \mathscr{C}^\alpha([0, t]; \mathbb{R}^d)$. Using the monotone class theorem for functions we get (74).

From [53], Corollary 5.2, we get that

$$\int P(\hat{\pi}(\omega)) \, d\mathbf{Y}(\omega) = \left(\int P(\hat{\pi}) \circ dY\right)(\omega) \quad \text{for } \mathbb{P}\text{-a.e. } \omega,$$

where $\hat{\pi}_t = \mathcal{L}(X^{\mathbf{Y}} \mid \mathcal{Y}_t) = \mathcal{L}(X^{\mathbf{y}})|_{\mathbf{y}=\mathbf{Y}}$, the second equality is by definition (74). Hence $X^{\mathbf{Y}}$ satisfies (10). $\square$

5. The interacting particle system.   In this section we prove well-posedness and convergence of the interacting particle system (13) in the case when Y is a path of bounded variation.

To compress the notation and have a slightly more general result we study the following mean-field system:

(75) $$\begin{aligned} dX_t^{i,N} &= [b(X_t^{i,N}, \mu_t^N) + P(\mu_t^N)\dot{Y}_t + \Gamma(\mu_t^N)] \, dt + \sigma(X_t^{i,N}, \mu_t^N) \, dW_t^i, \\ X_t^{i,N}|_{t=0} &= X_0^i, \quad i = 1, \ldots, N, \end{aligned}$$

where $\mu_t^N := \frac{1}{N} \sum_{i=1}^N \delta_{X_t^{i,N}}$. The variable $X_t^{i,N}$ is in the state space $\mathbb{R}^D$. $(W^i)_{i \in \mathbb{N}}$ is a family of independent m-dimensional Brownian motions and $(X_0^i)_{i \in \mathbb{N}}$ is a family of independent and identically distributed initial conditions with law $\pi_0 \in \mathcal{P}(\mathbb{R}^D)$. Moreover, assume that $\dot{Y} : [0, T] \to \mathbb{R}^D$ is càdlàg and bounded. We assume that the coefficients b, σ and P satisfy Assumption 3 and 4, but notice that P only depends on the measure. For the coefficient Γ we assume the following.

ASSUMPTION 5.   Let $\rho \ge 1$, assume that there exists a constant $C > 0$ such that:

(i) (linear growth) $\forall \pi \in \mathcal{P}_\rho(\mathbb{R}^d)$,

$$|\Gamma(\pi)| \le C(1 + \overline{M}^\rho(\pi))^{\frac{2}{\rho}},$$

(ii) (Lipschitz continuity) $\forall \pi, \nu \in \mathcal{P}_\rho(\mathbb{R}^D)$,

$$|\Gamma(\pi) - \Gamma(\nu)| \le C(1 + \overline{M}^\rho(\pi)^2 + \overline{M}^\rho(\nu))^{\frac{1}{\rho}} W_{\rho, \mathbb{R}^D}(\pi, \nu).$$

REMARK 5.1. Notice that the growth is quadratic in the central moments of the measure. Also, the best Lipschitz constant we can hope for is the square of the central moment of one measure and the central moment of the other.

REMARK 5.2. By doing the same substitution as in Section 4.5 and taking (9) for Γ we recover the interacting particle system (13).

LEMMA 5.3. *Let $T > 0$ and $\rho \ge 1$. If the initial distribution π_0 has finite ρ-moment, then equation* (75) *admits a pathwise unique strong solution on $[0, T]$. Moreover, there exists $C = C(\rho, \pi_0, T) > 0$ such that*

$$(76) \qquad \frac{1}{N} \sum_{j=1}^N \mathbb{E}\Big[\sup_{t \in [0,T]} |\overline{X}_t^{j,N}|^\rho \Big] \le C, \qquad \max_{i=1,\dots,N} \mathbb{E}\Big[\sup_{t \in [0,T]} |X_t^{i,N}|^\rho \Big] \le C(1 + \|\dot{Y}\|_\infty),$$

where $\overline{X}^{i,N} = X^{i,N} - \frac{1}{N} \sum_{j=1}^N X^{j,N}$, for $i = 1, \dots, N$.

PROOF. Under Assumptions 3, 4 and 5, the coefficients are locally Lipschitz and there is classically strong existence and pathwise uniqueness for the SDE (75) up to an explosion time τ. We want to prove that $\tau > T$. Let us call the solution $X_t^{(N)} = (X_t^{i,N}, \dots, X_t^{N,N})$. Since the coefficents P and Γ do not depend on the state variable x we have the following identity for $\overline{X}^{i,N}$, $i = 1, \dots, N$:

$$d\overline{X}_t^{i,N} = \left(b(X_t^{i,N}, \mu_t^N) - \frac{1}{N} \sum_{j=1}^N b(X^{j,N}, \mu_t^N) \right) dt + \sigma(X_t^{i,N}, \mu_t^N) \, dW_t^i$$

$$- \frac{1}{N} \sum_{j=1}^N \sigma(X^{j,N}, \mu_t^N) \, dW_t^j.$$

Notice that

$$\overline{M}^\rho(\mu_t^N) = \mathbb{E}_{\mu_t^N}[|X - \mathbb{E}_{\mu_t^N}[X]|^\rho] = \frac{1}{N} \sum_{i=1}^N |\overline{X}_t^{i,N}|^\rho.$$

Using Assumption 3 we obtain that for every $\rho \ge 2$ there exists a constant $C > 0$ independent of $\dot{Y}$ and N such that

$$\mathbb{E}\Big[\sup_{s \le t} |\overline{X}_s^{i,N}|^\rho \Big] \le \mathbb{E}[|\overline{X}_0^{i,N}|^\rho] + C \int_0^t \left(1 + \frac{1}{N} \sum_{j=1}^N \mathbb{E}[|\overline{X}_s^{j,N}|^\rho] \right) ds.$$

Taking the mean over i and using Gronwall's lemma we obtain the first estimate in (76). We can now estimate the moments of $X^{i,N}$ as follows:

$$\mathbb{E}\Big[\sup_{t \in [0,T]} |X_t^{i,N}|^\rho \Big] \le C(1 + \|\dot{Y}\|_\infty) \int_0^T \mathbb{E}[(1 + \overline{M}^\rho(\mu_s^N))]^2 \, ds \le (1 + \|\dot{Y}\|_\infty)C,$$

where $C = C(\rho, \pi_0, T)$ can change from one inequality to the next. Since $X^{(N)}$ has finite L_ω^ρ-norm on the interval $[0, T]$, it means that the explosion time satisfies $\tau > T$. □

REMARK 5.4.   Notice that in Lemma 5.3 the choice of $T > 0$ was arbitrary, which implies that the system of interacting particles is well posed on intervals of any length.

5.1. *Propagation of chaos.*   For $i = 1, \ldots, N$, we introduce the following McKean–Vlasov equation:

$$
\text{(77)} \qquad
\begin{aligned}
\mathrm{d}X_t^i &= \big[b(X_t^i, \pi_t) + P(\pi_t)\dot{Y}_t + \Gamma(\pi_t)\big]\mathrm{d}t + \sigma(X_t^i, \pi_t)\,\mathrm{d}W_t^i, \\
\pi_t &= \mathcal{L}(X_t^i), \qquad X_t^i|_{t=0} = X_0^i \sim \pi_0.
\end{aligned}
$$

By construction the random variables $X^1, \ldots, X^N$ are independent and identically distributed.

Equation (77) is well-posed because it corresponds to equation (62) driven by the *rough path*

$$
\mathbf{Y}_{st} := \left(Y_{s,t}, \int_s^r Y_{s,r} \otimes \dot{Y}_r \, \mathrm{d}r - \frac{1}{2}(t - s)I_{d \times d}\right),
$$

where $Y_t := \int_0^t \dot{Y}_r \, \mathrm{d}r$. Notice that, since $\dot{Y}$ is càdlàg and bounded, the path $\mathbf{Y}$ is Lipschitz continuous which is much more regular than α-Hölder with $\alpha \in (\frac{1}{3}, \frac{1}{2}]$.

Let $\overline{X}^i := X^i - \mathbb{E}[X^i]$ and notice that, by definition $\overline{M}^\rho(\pi_t) = \|\overline{X}_t^i\|_{L_\omega^\rho}^\rho$. Using the a priori estimate in Lemma 4.7 and Assumptions 3 and 5 we can see that, for any $\rho \geq 2$ we have an estimate on the central moments of π_t,

$$
\overline{M}^\rho(\pi_t) \leq \Big\| \sup_{t \in [0,T]} |\overline{X}_t^i| \Big\|_{L_\omega^\rho}^\rho \leq C\big[1 + \overline{M}^\rho(\pi_0)\big]^\rho.
$$

From this one can easily recover the following estimate for the moments of π_t:

$$
\text{(78)} \qquad
\begin{aligned}
M^\rho(\pi_t)^{\frac{1}{\rho}} &\leq C\big(1 + \|\dot{Y}\|_\infty\big)\big(M^\rho(\pi_0) + \overline{M}^{2\rho}(\pi_t)\big)^{\frac{1}{\rho}} \\
&\leq C\big(M^\rho(\pi_0) + \|\dot{Y}\|_\infty\big)\big(1 + \overline{M}^{2\rho}(\pi_0)\big)^2,
\end{aligned}
$$

where $C = C(\rho, T)$.

REMARK 5.5.   Notice that the ρ-moment of π_t is only bounded by the 2ρ-central moment of π_0 because of the quadratic growth of Γ from Assumption 5.

In the following we will hide the dependence on π_0 in the constant C and only focus on the explicit dependence on $\|\dot{Y}\|_\infty$. We define the empirical measure associated with the independent particles X^i as $\overline{\mu}_t^N := \frac{1}{N}\sum_{i=1}^N \delta_{X_t^i}$. It follows from [50], Theorem 1, that for any $\overline{\rho} > \rho$ there exists an explicit rate of convergence $H(N) = H(N, \rho, \overline{\rho})$ such that $H(N) \to 0$ as $N \to \infty$ and

$$
\text{(79)} \qquad
\mathbb{E}\big[W_{\rho, \mathbb{R}^D}(\overline{\mu}_t^N, \pi_t)^\rho\big] \leq M^{\overline{\rho}}(\pi_t)^{\frac{\rho}{\overline{\rho}}} H(N) \leq C\big(1 + \|\dot{Y}\|_\infty\big)H(N),
$$

where in the last inequality we used (78) and C depends on the moments of π_0.

Moreover, the rate is optimal and explicitly given as

$$
H(N) =
\begin{cases}
N^{-\frac{1}{2}} + N^{1 - \frac{\rho}{\overline{\rho}}}, & \rho > \dfrac{D}{2}, \overline{\rho} \neq 2\rho, \\[2mm]
N^{-\frac{\rho}{D}} + N^{1 - \frac{\rho}{\overline{\rho}}}, & 1 \leq \rho < \dfrac{D}{2}, \overline{\rho} \neq \dfrac{D}{D-2}.
\end{cases}
$$

Notice that if we have enough moments the optimal rate of convergence for independent particles is $H(N) = \frac{1}{\sqrt{N}}$. Furthermore, we have the following estimate, because μ^N and $\overline{\mu}^N$ are both empirical measures:

$$W_{\rho,\mathbb{R}^D}(\mu_t^N, \overline{\mu}_t^N)^\rho \leq \frac{1}{N}\sum_{i=1}^N |X_t^{i,N} - X_t^i|^\rho,$$

so that we have from the triangle inequality

(80) $$\sup_{t\in[0,T]} \mathbb{E}[W_{\rho,\mathbb{R}^D}(\mu_t^N, \pi_t)^\rho] \leq C(1 + \|\dot{Y}\|_\infty)H(N) + \sup_{t\in[0,T]} \frac{1}{N}\sum_{i=1}^N \mathbb{E}|X_t^{i,N} - X_t^i|^\rho,$$

where C depends on the $2\overline{\rho}$-moment of π_0, for $\overline{\rho} > \rho$. Now that we have stated most of the preliminaries we can prove the following convergence result.

PROPOSITION 5.6. *Let $\overline{\rho} > \rho \geq 2$. If $X_0 \in L_\omega^{2\overline{\rho}}$, there is a function $J(N) = O(\log(N)^{-1})$ such that*

$$\max_{i=1,\dots,N} \mathbb{E}\left[\sup_{t\in[0,T]} |X_t^{i,N} - X^i|^\rho\right] \lesssim e^{\|\dot{Y}\|_\infty} J(N).$$

PROOF. We call $\overline{X}^{i,N} := X^{i,N} - \frac{1}{N}\sum_{j=1}^N X^{j,N}$, for $i = 1, \dots, N$ and recall that the central moment for the empirical measure is $\overline{M}^\rho(\mu_t^N) = \frac{1}{N}\sum_{i=1}^N |\overline{X}_t^{i,N}|^\rho$. For $R > 0$, we define the stopping time $T_R := \inf\{t \geq 0 : \overline{M}^\rho(\mu_t^N) \geq R\}$.

We set $Z^i := X^{i,N} - X^i$ and using Assumptions 3 and 5 as well as inequality (80) and Lemma 4.7 we compute the following:

$$\max_{i=1,\dots,N} \mathbb{E}[|Z_t^i|^\rho 1_{\{T_R>t\}}]$$

$$\leq C\mathbb{E}\left[\int_0^{t\wedge T_R} (1 + \overline{M}^{2\rho}(\pi_s)\|\dot{Y}\|_\infty + \overline{M}^\rho(\mu_s^N))(|Z_s^i| + W_{\rho,\mathbb{R}^D}(\pi_s, \mu_s^N))^\rho \, ds\right]$$

$$\leq C(1 + \|\dot{Y}\|_\infty + R)\left[\int_0^t \max_{i=1,\dots,N} \mathbb{E}[|Z_s^i|^\rho 1_{\{T_R\geq s\}}]\,ds + H(N)\right].$$

Using Gronwall's inequality we obtain

$$\max_{i=1,\dots,N} \mathbb{E}[|Z_t^i|^\rho 1_{\{T_R>t\}}] \leq C(1 + \|\dot{Y}\|_\infty)e^{C(1+\|\dot{Y}\|_\infty)R}e^{CR}H(N).$$

Now we compute the following using Cauchy–Schwarz and Markov inequalities as well as (76):

$$\mathbb{E}[|Z_t^i|^\rho 1_{\{T_R\leq t\}}] \leq C(\mathbb{E}[|Z_t^i|^{2\rho}])^{\frac{1}{2}}\mathbb{P}(T_R \leq t)^{\frac{1}{2}} \leq C\mathbb{P}\left(\sup_{s\leq t}\overline{M}^\rho(\mu_s^N) \geq R\right)^{\frac{1}{2}}$$

$$\leq \frac{C}{R^{\frac{1}{2}}}\left(\frac{1}{N}\sum_{i=1}^N \mathbb{E}\left[\sup_{t\in[0,T]}|\overline{X}_t^{i,N}|^\rho\right]\right)^{\frac{1}{2}} \leq \frac{C}{R^{\frac{1}{2}}}.$$

Notice that we used $\mathbb{E}[|Z_t^i|^{2\rho}] \leq C(\mathbb{E}[|X_t^{i,N}|^{2\rho}] + \mathbb{E}[|X_t^i|^{2\rho}]) \leq C(1 + \|\dot{Y}\|_\infty)$, hence we need from Remark 5.5 that the initial measure π_0 has finite 2ρ moments. We can put together

the estimates to obtain

$$\max_{i=1,\ldots,N} \mathbb{E}\big[|Z_t^i|^\rho\big] \le \frac{C}{R^{\frac{1}{2}}} + C(1 + \|\dot{Y}\|_\infty + R)e^{C(1+\|\dot{Y}\|_\infty)}e^{CR}H(N)$$

$$\le Ce^{C(1+\|\dot{Y}\|_\infty)}\Big(\frac{1}{R^{\frac{1}{2}}} + e^{CR}H(N)\Big),$$

where we changed the constants from one line to the next. Now choose $R = R(N)$ such that

$$R(N) \to \infty, \qquad e^{CR(N)}H(N) \to 0 \quad \text{as } N \to \infty.$$

Remember that $H(N) \approx N^{-\gamma}$ with $\gamma < \frac{1}{2}$ so that choosing $R(N) \approx \log(N^{\overline{\gamma}})$ for some $0 < \overline{\gamma} < \frac{\gamma}{C}$ we have a rate of convergence

$$J(N) \approx (\overline{\gamma}\log(N))^{-1} + N^{\overline{\gamma}C-\gamma} \approx \log(N)^{-1}. \qquad \square$$

REMARK 5.7.   Notice that the rate of convergence is far from the optimal $\frac{1}{\sqrt{N}}$ of a sample of independent and identically distributed random variables. This is due to the nonlocal Lipschitz condition on Γ in Assumption 5.

Finally, we can put together Corollary 4.16 and Proposition 5.6.

THEOREM 5.8.   *Let $\delta > 0$ and $\mathbf{Y}^\delta$ be a bounded differentiable approximation of a geometric rough path $\mathbf{Y} = (Y, \mathbb{Y}) \in \mathscr{C}^\alpha$. Let $\mu^{N,\delta}$ be the empirical measure of the system of mean-field particles (75) with input $\mathbf{Y}^\delta$. Moreover, let π be the law of the solution to equation (62) driven by $(Y, \mathbb{Y} - \frac{1}{2}(t-s)\mathrm{Id}) \in \mathscr{C}^\alpha$.*
Then there exists $\rho > 0$ and a sequence $\delta(N)$ such that $\delta(N) \to 0$ and

$$\sup_{t\in[0,T]} \mathbb{E}\big[W_{\rho,\mathbb{R}^D}(\mu_t^{N,\delta(N)}, \pi_t)^\rho\big] \to 0 \quad \text{as } N \to \infty.$$

PROOF.   Let $\delta \equiv \delta(N) > 0$ to be chosen later. Let $\pi^{\delta(N)}$ be the law of a solution to (77). By the triangle inequality, Proposition 5.6 and Corollary 4.16 (which can be applied for every $\rho > 2/(1-2\bar{\alpha})$)

$$\sup_{t\in[0,T]} \mathbb{E}\big[W_{\rho,\mathbb{R}^D}(\mu_t^{N,\delta(N)}, \pi_t)^\rho\big] \lesssim \sup_{t\in[0,T]} \mathbb{E}\big[W_{\rho,\mathbb{R}^D}(\mu_t^{N,\delta(N)}, \pi_t^{\delta(N)})^\rho\big]$$

$$+ W_{\rho,C([0,T];\mathbb{R}^D)}(\pi^{\delta(N)}, \pi)$$

$$\lesssim e^{\|\dot{Y}^{\delta(N)}\|_\infty}J(N) + \rho_{\overline{\alpha}}(\mathbf{Y}^{\delta(N)}, \mathbf{Y}).$$

Choosing $\delta(N)$ such that $\|\dot{Y}^{\delta(N)}\|_\infty = o(\log(J(N)^{-1})) = o(\log(\log(N)))$ we have that the right hand side vanishes as $N \to \infty$. $\square$

6. The numerical scheme, construction of the lift and examples.   In this section we derive the RP-EnKF (rough path ensemble Kalman filter) numerical scheme alluded to in the Introduction (see equation (14)), discuss the construction of appropriate rough path lifts, and provide details concerning the implementation. Furthermore, we demonstrate its effectiveness in the context of misspecified and multiscale models by means of a few examples in the context of parameter estimation. This setting provides a convenient testbed for the scenario where the model and observation noises are correlated, and we expect our conclusions regarding (non-)robustness to be relevant more generally.

6.1. *Numerical scheme.*

Discretising in time. A natural scheme for approximating the dynamics of the interacting particle system (13) is given by

(81a)
$$\begin{aligned} X_{k+1}^i = {} & X_k^i + f(X_k^i)\Delta t + G^{1/2}\sqrt{\Delta t}\xi_k^i \\ & + \widehat{P}_k\big(\Delta Y_k - (h(X_k^i)\Delta t + U\sqrt{\Delta t}\xi_k^i + R^{1/2}\sqrt{\Delta t}\eta_k^i)\big) \\ & + \widehat{\mathrm{Cov}}_k(x, Dh)\widehat{P}_k\Delta\mathbb{Y}_k + \widehat{\Gamma}_k\Delta t, \end{aligned}$$

where $\Delta t > 0$ is the step size, and (ξ_n^i) and (η_n^i) denote independent zero mean Gaussian random variables with unit variance of dimensions D and d, respectively. In the above display, $\widehat{P}_k := \widehat{\mathrm{Cov}}_k(x, h)C^{-1} + B$ and $\widehat{\mathrm{Cov}}_k(x, Dh)$ refer to the standard unbiased empirical estimators of the covariance, that is,

(82a)
$$\widehat{\mathrm{Cov}}_k(x, h) = \frac{1}{N-1}\sum_{i=1}^{N}(X_k^i - \bar{X}_k) \otimes h(X_k^i) \in \mathbb{R}^{D\times d},$$

(82b)
$$\widehat{\mathrm{Cov}}_k(x, Dh) = \frac{1}{N-1}\sum_{i=1}^{N}(X_k^i - \bar{X}_k) \otimes Dh(X_k^i) \in \mathbb{R}^{D\times d\times D},$$

where

(83)
$$\bar{X}_k = \frac{1}{N}\sum_{i=1}^{N}X_k^i$$

refers to the empirical mean. With (82) in place, the standard empirical estimator for Γ as defined in (9) is given by

(84)
$$\widehat{\Gamma}_{k,\gamma} = -\frac{1}{2}\mathrm{Trace}\big(\widehat{\mathrm{Cov}}_{k,\gamma}(x, Dh)\widehat{P}_k\big),$$

where $\widehat{\Gamma}_{k,\gamma}$ denotes the γth component of $\widehat{\Gamma}_k \in \mathbb{R}^D$, and a similar convention is used for $\widehat{\mathrm{Cov}}_{k,\gamma}(x, Dh)$. The precise meaning of the first term in equation (81a) is

$$\big(\widehat{\mathrm{Cov}}_k(x, Dh)\widehat{P}_k\Delta\mathbb{Y}_k\big)_\gamma = \sum_{j,q=1}^{d}\sum_{r=1}^{D}\widehat{\mathrm{Cov}}_k(x, Dh)_{\gamma,j,r}(\widehat{P}_k)_{r,q}(\Delta\mathbb{Y}_k)_{q,j}, \quad \gamma = 1, \ldots, D,$$

and analogously in equation (84) with $\Delta\mathbb{Y}_k$ replaced by the d-dimensional identity matrix.

REMARK 6.1 (Gubinelli derivative). The first term in (81a) is modelled after the Gubinelli derivative $P(\widehat{\pi}_s)'$ in (12). We would like to stress that a standard time discretisation of the interacting particle system (13) according to Davie [32] would involve further contributions accounting for correlations between the particles as well as for cross terms induced by the joint lift $(Y, W, V) \mapsto ((Y, W, V), (\mathbb{Y}, \mathbb{W}, \mathbb{V}))$. At least formally, these additional terms vanish in the limit as $N \to \infty$, and our numerical experiments have not shown noticeable benefits of including them.

Constructing the lift $\mathbb{Y}$. In order to implement the RP-EnKF scheme defined in (81), we need to posit the discrete-time second order increments $\Delta\mathbb{Y}_k \in \mathbb{R}^{d\times d}$, given discrete-time samples $y_0, y_1, \ldots, y_n \in \mathbb{R}^d$ from $(Y_t)_{0\le t\le T}$. In what follows we will denote the piecewise-linear interpolation of $y_0, \ldots, y_n$ with step size Δt by $(Y_t^{(\Delta t)})_{0\le t\le T}$ and consider the decomposition of $\Delta\mathbb{Y}_k$ into symmetric and skew-symmetric parts,

(85)
$$\Delta\mathbb{Y}_k = \Delta\mathbb{Y}_k^{\mathrm{sym}} + \Delta\mathbb{Y}_k^{\mathrm{skew}},$$

where

$$\Delta \mathbb{Y}_k^{\mathrm{sym}} := \mathrm{sym}(\Delta Y_k) := \frac{1}{2}\big(\Delta \mathbb{Y}_k + \Delta \mathbb{Y}_k^\top\big),$$

(86)

$$\Delta \mathbb{Y}_k^{\mathrm{skew}} := \mathrm{skew}(\Delta Y_k) := \frac{1}{2}\big(\Delta \mathbb{Y}_k - \Delta \mathbb{Y}_k^\top\big).$$

For the symmetric part, we set

$$(87) \qquad \Delta \mathbb{Y}_k^{\mathrm{sym}} = \mathrm{sym}\bigg(\int_{t_k}^{t_{k+1}} Y_{t_k,r}^{(\Delta t)} \otimes \mathrm{d}Y_r^{(\Delta t)}\bigg) = \frac{1}{2}(y_{k+1} - y_k) \otimes (y_{k+1} - y_k),$$

maintaining structural similarities to the defining algebraic relation of weakly geometric rough paths [53], Section 2.2. We also note that (87) can be thought of as an estimator for the quadratic variation of Y on the time interval $[t_k, t_{k+1}]$. Defining the skew-symmetric part is more challenging since[5]

$$(88) \qquad \qquad \mathrm{skew}\bigg(\int_{t_k}^{t_{k+1}} Y_{t_k,r}^{(\Delta t)} \otimes \mathrm{d}Y_r^{(\Delta t)}\bigg) = 0,$$

indicating that information on the enclosed area between two neighbouring points is inevitably lost by the discretisation. However, Wong–Zakai type results on piecewise linear interpolations of semimartingales [28], Proposition 2, as well as the convergence of simplified Euler schemes [39, 52] suggest that the scheme (81) with $\Delta \mathbb{Y}_k^{\mathrm{skew}} = 0$ and $\Delta \mathbb{Y}_k^{\mathrm{sym}}$ as defined in (87), based on data $(Y_t)_{0 \le t \le T}$ obtained from (1b), recovers (11) in the limit $\Delta t \to 0$, and our numerical experiments support this conjecture. We leave a detailed analysis for future work and refer to [9, 49, 80] for specific approaches towards estimating Lévy areas. We next discuss the setting when the data is only approximately obtained from (1a), in which case the skew-symmetric contributions $\Delta \mathbb{Y}_k^{\mathrm{skew}}$ will play a crucial role:

Correcting $\mathbb{Y}$ in the context of model misspecification.   Let us consider the case when instead of observations $(Y_t)_{0 \le t \le T}$ from (1), we have access to perturbed or modified data $(Y_t^\varepsilon)_{0 \le t \le T}$, so that $Y^\varepsilon \to Y$ in $C^\alpha([0, T]; \mathbb{R}^d)$ for some $\alpha \in (\frac{1}{3}, \frac{1}{2}]$, almost surely. Particular instances of this situation occur when the filtering model (1) arises as a simplified description of a more elaborate (possibly multiscale) model in the limit as $\varepsilon \to 0$ (with ε referring to a scale separation parameter in the multiscale scenario). For specific examples we refer to Sections 6.2–6.4 below. Assuming for a moment that $(Y_t^\varepsilon)_{0 \le t \le T}$ is in $C^\gamma([0, T]; \mathbb{R}^d)$ for $\gamma > \frac{1}{2}$, the canonical lift $\mathbb{Y}_{s,t}^\varepsilon = \int_s^t Y_{s,r}^\varepsilon \, \mathrm{d}Y_r^\varepsilon$ is well defined, and it might seem natural to use $\mathbf{Y}^\varepsilon := (Y^\varepsilon, \mathbb{Y}^\varepsilon)$ in the rough McKean–Vlasov dynamics (11), or a discretised version thereof in the scheme (14). However, it is then clearly possible that $\mathbf{Y}^\varepsilon \not\to \mathbf{Y} := (Y, \mathbb{Y}^{\mathrm{Strat}})$ in $\mathscr{C}^\alpha([0, T]; \mathbb{R}^d)$, with $\mathbb{Y}^{\mathrm{Strat}}$ denoting the Stratonovich lift associated to (1b). In this case, the continuity statement of Theorem 1.1 suggests that the solutions to (11) driven by $\mathbf{Y}$ and $\mathbf{Y}^\varepsilon$ may be substantially different, in general, and hence inference based on $\mathbf{Y}^\varepsilon$ may be erroneous.[6] In this case, we say that the model providing $(Y_t^\varepsilon)_{0 \le t \le T}$ is *misspecified* with respect to the filtering model (1) on which the schemes developed in this paper are built. The degree of misspecification can be quantified using the rough path metric $\rho_\alpha(\mathbf{Y}, \mathbf{Y}^\varepsilon)$, see Section 3.2.

In order to devise a practical method to address the problem exposed in the preceding paragraph, let us assume that there exists $\overline{\mathbf{Y}} = (Y, \overline{\mathbb{Y}}) \in \mathscr{C}^\alpha([0, T]; \mathbb{R}^d)$ such that $\mathbf{Y}^\varepsilon \to \overline{\mathbf{Y}}$. Noting

---

[5]This relation expresses the fact that the area enclosed by a straight line with itself is zero.

[6]We remind the reader that the system (11) has been derived on the basis of the approximation stated in Lemma 2.2 which is only exact in the case when f and h are affine, and π_0 is Gaussian. For the sake of discussion in this section, we assume that the incurred error is negligible and the solution to (11) driven by $\mathbf{Y}$ provides sufficiently accurate estimates.

that $\overline{\mathbf{Y}}$ and $\mathbf{Y} = (Y, \mathbb{Y}^{\text{Strat}})$ lie above the same path Y, we see that the *model discrepancy* is captured by

$$(89) \qquad \mathbb{A}_{s,t} := \mathbb{Y}_{s,t}^{\text{Strat}} - \overline{\mathbb{Y}}_{s,t}, \quad s, t, \in [0, T].$$

Introducing a sequence of equidistant grids with mesh sizes $\delta > 0$, the discrepancy $\mathbb{A}$ can be recovered as follows, commuting the limits $\delta \to 0$ and $\varepsilon \to 0$.

LEMMA 6.2. *Let Y be a semimartingale, and consider its Stratonovich lift $\mathbf{Y} = (Y, \mathbb{Y}^{\text{Strat}})$. Assume that there exists a convergent family of rough paths $\mathbf{Y}^{\varepsilon} = (Y^{\varepsilon}, \mathbb{Y}^{\varepsilon}) \subset \mathscr{C}^{\alpha}([0, T]; \mathbb{R}^d)$,*

$$(90) \qquad \mathbf{Y}^{\varepsilon} \xrightarrow{\varepsilon \to 0} \overline{\mathbf{Y}} = (\overline{Y}, \overline{\mathbb{Y}}) \in \mathscr{C}^{\alpha}([0, T]; \mathbb{R}^d), \quad \mathbb{P}\text{-}a.s.,$$

so that the limit $\overline{\mathbf{Y}}$ lies above Y, that is, $\overline{Y} = Y$, $\mathbb{P}$-a.s. Let us futher denote by $Y^{\varepsilon,(\delta)}$ and $Y^{(\delta)}$ the piecewise linear interpolations of Y^{ε} and Y, respectively, on an equidistant grid of mesh size $\delta > 0$. The corresponding (canonical) rough path lifts are given by $\mathbb{Y}^{(\delta)} = \int Y^{(\delta)} \otimes dY^{(\delta)}$ and $\mathbb{Y}^{\varepsilon,(\delta)} = \int Y^{\varepsilon,(\delta)} \otimes dY^{\varepsilon,(\delta)}$, respectively. Moreover, assume that for fixed $\varepsilon > 0$,

$$(91) \qquad \mathbb{Y}^{\varepsilon,(\delta)} \xrightarrow{\delta \to 0} \mathbb{Y}^{\varepsilon} \quad \text{in } C_2^{2\alpha}([0, T]; \mathbb{R}^{d \times d}), \mathbb{P}\text{-}a.s.$$

Then the discrepancy $\mathbb{A} = \overline{\mathbb{Y}} - \mathbb{Y}^{\text{Strat}}$ satisfies

$$(92) \qquad \mathbb{A}_{s,t} = \lim_{\varepsilon \to 0} \lim_{\delta \to 0} \mathbb{Y}_{s,t}^{\varepsilon,(\delta)} - \lim_{\delta \to 0} \lim_{\varepsilon \to 0} \mathbb{Y}_{s,t}^{\varepsilon,(\delta)}, \quad s, t \in [0, T], \mathbb{P}\text{-}a.s.,$$

and, furthermore, $\mathbb{A}$ is skew-symmetric, that is, $\mathbb{A}_{s,t}^{\top} = -\mathbb{A}_{s,t}$, $\mathbb{P}$-a.s.

REMARK 6.3. Let us briefly discuss the two main assumptions stated in (90) and (91). The convergence in (90) expresses the condition that the misspecified model family can asymptotically be described in terms of rough paths (that do not in general need to agree with the Stratonovich lift associated to (1b)). The condition in (91) demands that the lift $\mathbb{Y}^{\varepsilon}$ can be recovered from iterated integrals along piecewise linear interpolations. This is in particular the case when $\mathbb{Y}^{\varepsilon}$ is the canonical lift of a sufficiently regular path Y^{ε}, and also, by virtue of the Wong-Zakai theorem, for Stratonovich lifts. Although the limits in Lemma 6.2 are assumed to hold in $\mathbb{P}$-almost sure sense, an inspection of the proof below reveals that analogous versions of this result can straightforwardly be stated in terms of convergence in distribution or in $L^1(\mathbb{P})$. In our numerical experiments in the sections below, the assumptions of Lemma 6.2 are satisfied (with an appropriate notion of convergence).

PROOF. We have

$$(93) \qquad \lim_{\varepsilon \to 0} \lim_{\delta \to 0} \mathbb{Y}^{\varepsilon,(\delta)} = \overline{\mathbb{Y}}$$

in $C_2^{2\alpha}([0, T]; \mathbb{R}^{d \times d})$ by the assumptions in (90) and (91). On the other hand, we have

$$(94) \qquad \lim_{\delta \to 0} \lim_{\varepsilon \to 0} \mathbb{Y}^{\varepsilon,(\delta)} = \lim_{\delta \to 0} \mathbb{Y}^{(\delta)} = \mathbb{Y}^{\text{Strat}}$$

in $C_2^{2\alpha}([0, T]; \mathbb{R}^{d \times d})$, where the first limit follows from explicitly evaluating the iterated integrals $\mathbb{Y}^{\varepsilon,(\delta)}$ and using the fact that $Y^{\varepsilon} \to Y$ in $C^{\alpha}([0, T]; \mathbb{R}^d)$. The second limit follows from the Wong–Zakai theorem, see, for instance, [53], Theorem 9.3. To see that $\mathbb{A}$ is skew-symmetric, it suffices to observe that both $(Y, \mathbb{Y}^{\text{Strat}})$ and $\overline{\mathbf{Y}}$ are geometric rough paths. □

From Lemma 6.2, setting $\Delta \mathbb{Y}_k^{\mathrm{skew}} = \mathbb{A}_{t_k,t_{k+1}}$ is expected to correct the model misspecification error (see also [40], Section 8.2, and [100]). We also note that Chen's relation together with $Y = \overline{Y}$ implies

$$\mathbb{A}_{s,t} = \mathbb{A}_{0,t} - \mathbb{A}_{0,s}, \quad s, t \in [0, T], \tag{95}$$

so that it is sufficient to estimate the function $t \mapsto \mathbb{A}_{0,t}$. In practice, a realisation of $(Y_t^\varepsilon)_{0 \leq t \leq T}$ will only be available for one specific (small) value of ε. To mimic (92), we may however compute the difference between $\mathbb{Y}^{\varepsilon,(\delta_1)}$ and $\mathbb{Y}^{\varepsilon,(\delta_2)}$ for $\delta_1 \ll \delta_2$, so that δ_1 is small and δ_2 is large in comparison with ε.

Recalling that $(Y_t)_{0 \leq t \leq T}$ is typically available in the form of a discrete time series $y_0, \ldots, y_n$ (to which we can associate a grid with mesh size $\delta_1 = \Delta t$ coinciding with the grid for the numerical scheme (81) and with piecewise linear interpolation $(Y_t^{(\Delta t)})_{0 \leq t \leq T}$), we are naturally led to the idea of subsampling the data in order to obtain a coarser grid with mesh size $\delta_2 = \tau \delta_1 = \tau \Delta t$. More precisely, for a specific time-lag $\tau \in \mathbb{N}_{\geq 1}$, consider the subsampled sequence $y_0, y_\tau, y_{2\tau}, \ldots$, as well as the associated piecewise linear interpolation $(Y_t^{(\tau \Delta t)})_{0 \leq t \leq T}$. The time-lag τ shall be chosen in such a way that the corresponding area paths

$$t \mapsto \mathrm{skew}\left(\int_0^t Y_{0,r}^{(\Delta t)} \otimes \mathrm{d}Y_r^{(\Delta t)} \right) =: \mathbb{Y}_{0,t}^{\mathrm{skew},(\Delta t)},$$

$$(96) \qquad t \mapsto \mathrm{skew}\left(\int_0^t Y_{0,r}^{(\tau \Delta t)} \otimes \mathrm{d}Y_r^{(\tau \Delta t)} \right) =: \mathbb{Y}_{0,t}^{\mathrm{skew},(\tau \Delta t)}$$

are "as distinct as possible" (attempting to realise the limiting regimes in (92)), while maintaining $(Y_t^{(\Delta t)})_{0 \leq t \leq T} \approx (Y_t^{(\tau \Delta)})_{0 \leq t \leq T}$. The latter desideratum is motivated by the fact that Lemma 6.2 holds under the assumption that $\mathbf{Y}$ and $\overline{\mathbf{Y}}$ are over the space path Y. The comparisons between $\mathbb{Y}^{\mathrm{skew},(\Delta t)}$ and $\mathbb{Y}^{\mathrm{skew},(\tau \Delta t)}$ as well as between $Y^{\Delta t}$ and $Y^{\tau \Delta t}$ can be made in supremum norm, for instance. We then set

$$\Delta \mathbb{Y}_k^{\mathrm{skew}} = \big(\mathbb{Y}_{0,t_{k+1}}^{\mathrm{skew},(\Delta t)} - \mathbb{Y}_{0,t_k}^{\mathrm{skew},(\Delta t)} \big) - \big(\mathbb{Y}_{0,t_{k+1}}^{\mathrm{skew},(\tau \Delta t)} - \mathbb{Y}_{0,t_k}^{\mathrm{skew},(\tau \Delta t)} \big) \approx \mathbb{A}_{t_k,t_{k+1}}, \tag{97}$$

for the correction in the numerical scheme (81), relying on (95) and (92). Equation (97) compares the area contributions associated to the original interpolation $Y^{(\Delta t)}$ and the subsampled interpolation $Y^{(\tau \Delta t)}$. The requirement $(Y_t^{(\Delta t)})_{0 \leq t \leq T} \approx (Y_t^{(\tau \Delta t)})_{0 \leq t \leq T}$ is meant to ensure that $Y^{(\Delta t)}$ and $Y^{(\tau \Delta t)}$ mainly differ at the second-order level $\mathbb{Y}$; visually, the subsampling operation may be understood as "straightening out" Y and measuring the area difference accumulated thereby.

REMARK 6.4 (Relationship to subsampling in multiscale parameter estimation). As mentioned in Section 2.3, ideas related to subsampling have been considered extensively in the context of multiscale parameter estimation (without observational noise), see, for instance, [91, 95]. The method proposed in this section is different in that we use the subsampled paths in order to estimate the Lévy area correction, but otherwise input the original data path into the RP-EnKF dynamics (14). Moreover, the motivations are distinct: While subsampling in the aforementioned works is used in order to eliminate small-scale fluctuations, our method is specifically designed to estimate Lévy area correction terms. Our method may constitute a convenient alternative in complex settings where multiple effects would require competing subsampling frequencies, posing a challenge to tradional subsampling strategies. As our approach ultimately uses the resolution of the original data, we would expect to be able to include effects on multiple scales rather seamlessly into the procedure put forward in this section. Moreover, our arguments in this section suggest a principled approach towards

choosing appropriate subsampling frequencies (see also [5, 6] for alternatives in the more traditional setting of subsampling). We leave a detailed exploration of the connection between both methods for future work and refer the reader to the follow-up paper [100].

6.2. *Physical Brownian motion in a magnetic field.* In a first example, we consider a parameter estimation problem where the dynamics of interest is driven by a physical Brownian motion subject to a magnetic field. More precisely, physical Brownian motion $(W_t^\varepsilon)_{t \geq 0}$ is defined in terms of the unique strong solution to the following system of SDEs:

$$\text{(98a)} \qquad dW_t^\varepsilon = \frac{1}{\varepsilon} M P_t^\varepsilon \, dt, \quad W_0^\varepsilon = 0,$$

$$\text{(98b)} \qquad dP_t^\varepsilon = -\frac{1}{\varepsilon} M P_t^\varepsilon \, dt + dW_t^0, \quad P_0^\varepsilon = 0,$$

where $W_t^\varepsilon, P_t^\varepsilon \in \mathbb{R}^2$, $(W_t^0)_{t \geq 0}$ is a standard (mathematical) two-dimensional Brownian motion, and $\varepsilon \ll 1$ is a small parameter, the limit $\varepsilon \to 0$ corresponding to the regime of negligible particle mass. Furthermore, the matrix M is given by

$$\text{(99)} \qquad M = \begin{pmatrix} 1 & \gamma \\ -\gamma & 1 \end{pmatrix},$$

with $\gamma \in \mathbb{R}$ being a real-valued parameter associated to the strength of the magnetic field. For fixed $\alpha \in (1/3, 1/2)$ and $T > 0$, it is known that $(W^\varepsilon, \mathbb{W}^\varepsilon) \to (W^0, \mathbb{W}^{\text{phys}}(\gamma))$ in $\mathscr{C}^\alpha([0, T]; \mathbb{R}^d)$ and L^1 as $\varepsilon \to 0$, where

$$\text{(100)} \qquad \mathbb{W}_{s,t}^\varepsilon = \int_s^t W_{s,r}^\varepsilon \otimes dW_r^\varepsilon$$

denotes the canonical lift, and

$$\text{(101)} \qquad \mathbb{W}_{s,t}^{\text{phys}}(\gamma) = \int_s^t W_{s,r}^0 \otimes \circ dW_r^0 + (t - s)D,$$

with area correction

$$\text{(102)} \qquad D = \frac{1}{2} \begin{pmatrix} 0 & \gamma \\ -\gamma & 0 \end{pmatrix},$$

see [51] and [53], Section 3.4. The setting is reminiscent of the passage between underdamped and overdamped Langevin dynamics, see [94], Section 6.5.1, and [95], Section 2.2. Similar to [40], Section 8.2, we consider the problem of estimating the parameter $\theta \in \mathbb{R}$ in

$$\text{(103)} \qquad dZ_t^\varepsilon = \theta f(Z_t^\varepsilon) \, dt + dW_t^\varepsilon, \quad Z_0^\varepsilon = 0,$$

given noisy observations of the path $(Z_t^\varepsilon)_{0 \leq t \leq T}$, that is, given a path $(Y_t^\varepsilon)_{0 \leq t \leq T}$ of the solution to

$$\text{(104)} \qquad dY_t^\varepsilon = dZ_t^\varepsilon + R^{1/2} \, dV_t, \quad Y_0 = 0,$$

see Section 2.3. In (103) and (104), we allow for both $\varepsilon = 0$ and $\varepsilon > 0$, that is, we consider the dynamics driven by both mathematical and physical Brownian motion. Note that in the noiseless case $R = 0$ our setting coincides with the one discussed in [40, 100], see also Appendix B. Standard arguments show that both $(Z^\varepsilon, \mathbb{Z}^\varepsilon)$ and $(Y^\varepsilon, \mathbb{Y}^\varepsilon)$ converge in $\mathscr{C}^\alpha([0, T]; \mathbb{R}^d)$ with a nontrivial area correction akin to (101), where in the latter case, $\mathbb{Y}^\varepsilon$ refers to the Stratonovich iterated integrals. As an illustration, we plot sample paths of $t \mapsto Z_t^\varepsilon$ and an off-diagonal component of $t \mapsto \int_0^t Z_s^\varepsilon \otimes dZ_s^\varepsilon$ in Figure 1, comparing the cases $\varepsilon = 0$ and $\varepsilon = 10^{-2}$, for the same realisation of (a discretised version of) $(W_t^0)_{0 \leq t \leq T}$. Crucially, the convergence towards a nontrivially lifted path as expressed in (101) manifests itself in the
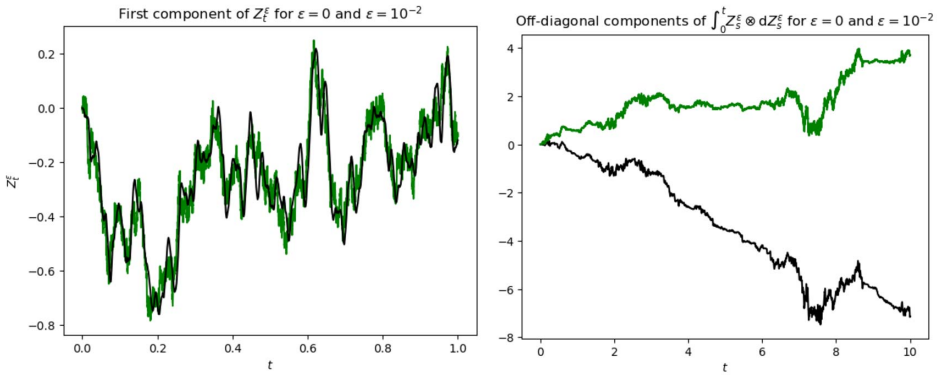
(a) Comparison of the paths $t \mapsto (Z_t^\varepsilon)_{0 \le t \le T}$, for mathematical Brownian motion ($\varepsilon = 0$, green) and physical Brownian motion ($\varepsilon = 10^{-2}$, black).

(b) Comparison of the iterated integrals (area processes) $t \mapsto \int_0^t Z_s^\varepsilon \otimes dZ_s^\varepsilon$.

FIG. 1. $t \mapsto Z_t^\varepsilon$ and $t \mapsto \int_0^t Z_s^\varepsilon \otimes dZ_s^\varepsilon$ for mathematical and physical Brownian motion.

offset between the paths in Figure 1b, corresponding to a component of $\mathbb{A}$ in (92). Throughout this section, we choose a fine time step of $\Delta t = 10^{-4}$ for all the involved approximations, $N = 10$ for the number of particles, $\theta_{\text{true}} = 1/2$ for the parameter to be recovered, $\gamma = -2.0$ for the strength of the magnetic field, $R = 0.1$ for the variance of the observation noise, and $f(z_1, z_2) = -(z_1 - z_2, z_1 + z_2)^\top$ for the drift in (103).

To test the robustness of the EnKF scheme (18) and the RP-EnKF scheme (14), we generate data according to (103) and (104) for both $\varepsilon = 0$ (mathematical Brownian motion) and $\varepsilon = 10^{-2}$ (physical Brownian motion). We would like to stress that the filtering methodology (expressed in terms of the schemes (18) and (14)) is however based on the model (1) and therefore tailored to the case $\varepsilon = 0$. In Figure 2, we show the empirical mean of the $\widehat{\theta}$-components for the output of the EnKF-dynamics (18), considering both mathematical and physical Brownian motion as drivers in (103). Evidently, the EnKF is not robust, in the sense that it fails to recover the true parameter θ_{true} in the case when $\varepsilon = 10^{-2}$, see Figure 2b.

We proceed by showing the corresponding results for the RP-EnKF scheme defined by (81) in Figure 3, demonstrating the robustness promised by Theorem 4.17. For the required
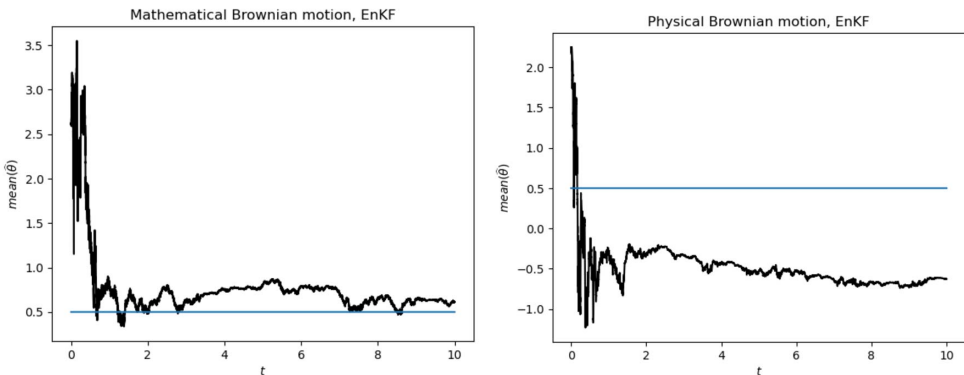
(a) mathematical Brownian motion, $\varepsilon = 0$.

(b) physical Brownian motion, $\varepsilon = 10^{-2}$.

FIG. 2.   *Output (empirical mean of the $\widehat{\theta}$-components) of the EnKF-scheme* (18), *for data obtained from the dynamics* (103) *driven by either mathematical or physical Brownian motion. The blue line indicates the true value* $\theta_{\text{true}} = \frac{1}{2}$. *To suppress sampling error, we plot averaged results computed from 5 independent runs.*
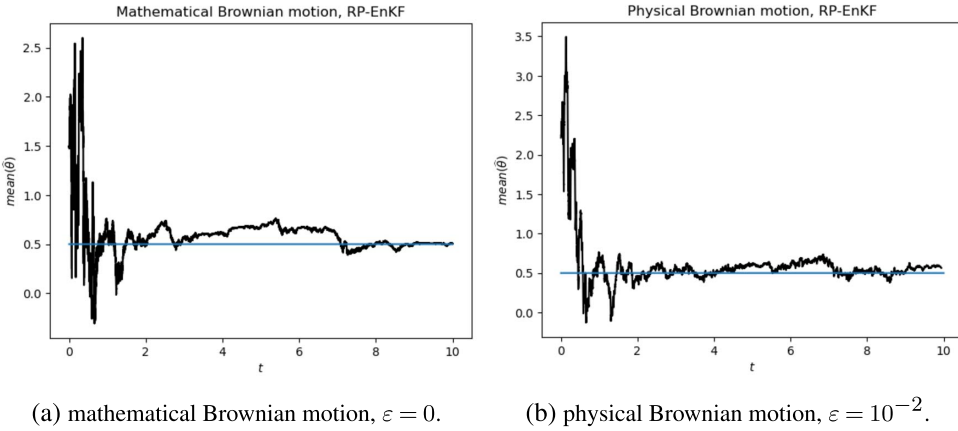
(a) mathematical Brownian motion, $\varepsilon = 0$. (b) physical Brownian motion, $\varepsilon = 10^{-2}$.

FIG. 3.   *Output (empirical mean of the $\widehat{\theta}$-components) of the RP-EnKF-scheme* (14), *for data obtained from the dynamics* (103) *driven by either mathematical of physical Brownian motion. The blue line indicates the true value $\theta_{\mathrm{true}} = \frac{1}{2}$. To suppress sampling error, we plot averaged results computed from 5 independent runs.*

(discrete-time) lift $\Delta\mathbb{Y}$, we use the construction detailed in Section 6.1. More precisely, in the case of physical Brownian motion, we set the time-lag to $\tau = 700$. In the case of mathematical Brownian motion, we set the skew-symmetric part in (85) to zero, $\Delta\mathbb{Y}_k^{\mathrm{skew}} = 0$, corresponding to the choice $\tau = 1$. These choices have been made on the basis of the subsampled area processes $t \mapsto \mathrm{skew}(\int_0^t Y_s^{\varepsilon,(\tau\Delta t)} \otimes dY_s^{\varepsilon,(\tau\Delta t)})$ depicted in Figure 4. More precisely, Figures 4a and 4c show the dependence $t \mapsto \mathrm{skew}(\int_0^t Y_s^{\varepsilon,(\tau\Delta t)} \otimes dY_s^{\varepsilon,(\tau\Delta t)})$ with different values of the time-lag τ, for physical Brownian motion ($\varepsilon = 10^{-2}$, Figure 4a) and mathematical Brownian motion ($\varepsilon = 0$, Figure 4c). While the subsampling only minimally effects the area process associated to the process driven by mathematical Brownian motion (Figure 4c), we observe a systematic shift in the case of physical Brownian motion (Figure 4c), revealing the latent multiscale structure. The difference between the original and subsampled area processes for physical Brownian motion is shown in Figure 4b. As the time-lag τ increases, said difference approaches the theoretically expected area correction implied by (101)–(102). The fact that the difference between the original and the subsampled area process reaches a plateau at around $\tau = 700$ is illustrated in Figure 4e, as opposed to the difference between the original and the subsampled paths, see Figure 4d. Consequently, the choice $\tau = 700$ strikes a balance between separating the original and subsampled area processes as much as possible while maintaining similarity between the original and subsampled paths (as suggested by the discussion motivating (97)).

6.3. *Fast chaotic dynamics—Lorenz-63.*   In this example, we consider the rescaled Lorenz ordinary differential equations [103] for $L_t^\varepsilon = (L_t^{(1),\varepsilon}, L_t^{(2),\varepsilon}, L_t^{(3),\varepsilon}) \in \mathbb{R}^3$,

$$(105a) \qquad \dot{L}_t^{(1),\varepsilon} = \frac{\sigma}{\varepsilon^2}(L^{(2),\varepsilon} - L^{(1),\varepsilon}), \qquad L_0^{(1),\varepsilon} = l_0^{(1)},$$

$$(105b) \qquad \dot{L}_t^{(2),\varepsilon} = \frac{1}{\varepsilon^2}(\rho L^{(1),\varepsilon} - L^{(2),\varepsilon} - L_t^{(1),\varepsilon} L_t^{(3),\varepsilon}), \qquad L_0^{(2),\varepsilon} = l_0^{(2)},$$

$$(105c) \qquad \dot{L}_t^{(3),\varepsilon} = \frac{1}{\varepsilon^2}(L_t^{(1),\varepsilon} L_t^{(2),\varepsilon} - \beta L_t^{(3),\varepsilon}), \qquad L_0^{(3),\varepsilon} = l_0^{(3)},$$

with the standard parameters $\sigma = 10$, $\rho = 28$ and $\beta = \frac{8}{3}$ as an example of fast chaotic dynamics approximating Brownian noise (with nontrivial area correction). We would like to stress that the phenomena described in this section can be considered generic across a wide range of fast chaotic deterministic dynamical systems, and refer the reader to [19] for an overview.
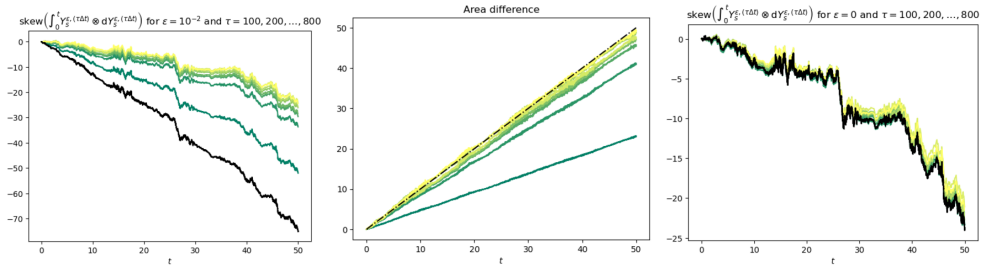
(a) Area processes for physi- (b) Differences of subsampled (c) Area processes for mathemat-
cal Brownian motion ($\varepsilon =$ and original area processes ical Brownian motion ($\varepsilon = 0$).
10^{-2}). ($\varepsilon = 10^{-2}$).
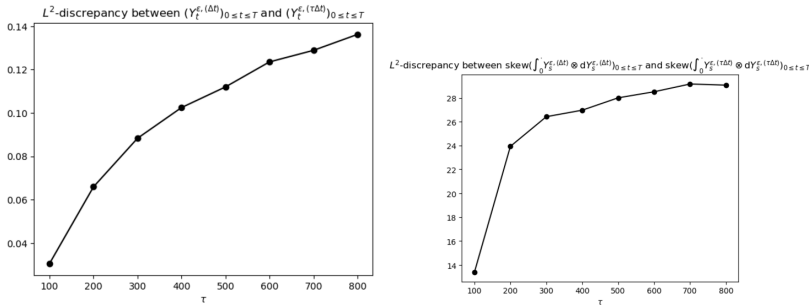
(d) L^2-discrepancy between original (e) L^2-discrepancy between original and
and subsampled path. subsampled area process.

FIG. 4. *Illustration of the subsampling procedure from Section* 6.1 *for constructing the discrete-time lift* $\Delta\mathbb{Y}$.
(4a): *Area processes* $t \mapsto \mathrm{skew}(\int_0^t Y_s^{\varepsilon,(\tau\Delta)} \otimes dY_s^{\varepsilon,(\tau\Delta t)})$ *(off-diagonal component) for physical Brownian mo-*
tion ($\varepsilon = 10^{-2}$) *and time-lags* $\tau = 100$(*green*), ..., 800 (*yellow*). *The area process associated to the path with-*
out subsampling is plotted in black. (4b): *Differences of original and subsampled area processes in the case*
of physical Brownian motion ($\varepsilon = 10^{-2}$), *that is,* $t \mapsto \int_0^t (Y_s^{\varepsilon,(\tau\Delta t)} \otimes dY_s^{\varepsilon,(\tau\Delta t)} - Y_s^{\varepsilon,(\Delta t)} \otimes dY_s^{\varepsilon,(\Delta t)})$, *in the*
spirit of (97). *The dashed line represents the theoretically expected area correction according to* (101). (4c):
Area processes associated to mathematical Brownian motion ($\varepsilon = 0$), *same colour scheme as for Figure 4a.* (4d):
L^2-*discrepancy between the original path* $(Y_t^{(\Delta t)})_{0 \leq t \leq T}$ *and the subsampled path* $(Y_t^{(\tau\Delta t)})_{0 \leq t \leq T}$ *as a function*
of the time-lag τ *for physical Brownian motion* ($\varepsilon = 10^{-2}$). (4e): L^2-*discrepancy between the original area pro-*
cess $\mathrm{skew}(\int_0^{\cdot} Y_s^{(\Delta t)} \otimes dY_s^{(\Delta t)})_{0 \leq t \leq T}$ *and the subsampled area process* $\mathrm{skew}(\int_0^{\cdot} Y_s^{(\tau\Delta t)} \otimes dY_s^{(\tau\Delta t)})_{0 \leq t \leq T}$ *as a*
function of the time-lag τ *for physical Brownian motion* ($\varepsilon = 10^{-2}$).

With $\varepsilon = 1$, the system (105) has originally been proposed as a simplified model for at-
mospheric convection [84], and can serve as a prototype for the study of chaotic ODEs. As
is well known, (105) possesses a "strange" chaotic attractor Λ equipped with a unique SRB
(Sinai–Ruelle–Bowen) measure[7] μ, see [113]. For $\varepsilon \ll 1$, a random initial condition $L_0^\varepsilon \in \mathbb{R}^3$
and after appropriate centering and rescaling, the solution $(L_t)_{t \geq 0}$ is well approximated by a
Brownian motion with a nontrivial area correction in the sense of rough paths:

For Hölder-continuous observables $v : \mathbb{R}^3 \to \mathbb{R}^m$ that are μ-centered (i.e., $\int_{\mathbb{R}^3} v \, d\mu = 0$),
a functional CLT (or *weak invariance principle*) holds for

$$(106) \qquad W_t^\varepsilon := \frac{1}{\varepsilon} \int_0^t v(L_s^\varepsilon) \, ds,$$

assuming that L_0^ε is initialised randomly according to μ. More precisely, there exists a Brow-
nian motion W (possibly on an extended probability space) with appropriate covariance

---

[7]The notion of SRB measures provides a suitable generalisation of ergodic measures.

$\Sigma \in \mathbb{R}^{m \times m}$ such that $W^{\varepsilon} \to W$ weakly in $C([0, T]; \mathbb{R}^m)$ as $\varepsilon \to 0$, see [63], Theorem 1.5.[8]
Moreover, it was shown in [11], Theorem 1.6, that a so-called *iterated weak invariance principle* holds for the iterated integrals

$$(107) \qquad \mathbb{W}_t^{\varepsilon} = \int_0^t W_s^{\varepsilon} \otimes dW_s^{\varepsilon},$$

that is, $(W^{\varepsilon}, \mathbb{W}^{\varepsilon}) \to (W, \mathbb{W})$ weakly in $C([0, T]; \mathbb{R}^m \times \mathbb{R}^{m \times m})$, where

$$(108) \qquad \mathbb{W}_t = \int_0^t W_s \otimes \circ dW_s + Dt,$$

with area correction $D \in \mathbb{R}^{m \times m}_{\text{skew}}$. We refer to [19], Theorem 4.4, for the corresponding statement in p-variation rough path topology.

In what follows, we consider $(Z_t^{\varepsilon})_{t \geq 0}$ to be driven by $(L_t^{(1:2), \varepsilon})_{t \geq 0} := (L_t^{(1), \varepsilon}, L_t^{(2), \varepsilon})_{t \geq 0}$,

$$(109) \qquad dZ_t^{\varepsilon} = \theta f(Z_t^{\varepsilon}) \, dt + \frac{\lambda}{\varepsilon} dL_t^{(1:2), \varepsilon},$$

with the parameter $\theta \in \mathbb{R}$ to be inferred from noisy observations

$$(110) \qquad dY_t^{\varepsilon} = dZ_t^{\varepsilon} + R^{1/2} \, dV_t,$$

and $\lambda > 0$ mediating the strength of the chaotic perturbation. Since $\int_{\mathbb{R}^d} (L^{(1)}, L^{(2)})^{\top} \, d\mu = 0$ according to [97], Section 11.7.2, we expect (109) to be well approximated by

$$(111) \qquad dZ_t^0 = \theta f(Z_t^0) \, dt + G^{1/2} \, dW_t,$$

in the regime $\varepsilon \ll 1$, with standard two-dimensional Brownian motion $(W_t)_{t \geq 0}$ and appropriate covariance $G \in \mathbb{R}^{2 \times 2}$. Replacing (109) by (111) is often a desirable simplification both computationally and conceptually, see, for instance, [58], [97], Section 11.7.2, and [108], Section 3.1. To apply our methodology, we need to presuppose G; estimates can be obtained using the approaches suggested in [58], Example 6.2, or [69], for instance. Here we use the value $G_{11}^{1/2} = 0.13$ reported in [69], Section 3.2.6, for $\lambda = \frac{2}{45}$ and note that $W^1 = W^2$ almost surely by a short calculation using (105). To test the RP-EnKF, we simulate data according to (105), (109) and (110) using an Euler–Maruyama discretisation with time step $\Delta t = 10^{-5}$, an observation noise level of $R = 0.01$ and a parameter value of $\theta_{\text{true}} = 0.5$ to be recovered. Furthermore, we set $\varepsilon = 0.05$ and $f(z_1, z_2) = (z_1 - z_2, z_1 + z_2)^{\top}$. The RP-EnKF is set up according to the model (111) with observations (110), using the scheme (14) with $N = 10$ particles. As an illustration for the challenges that are posed by the attempt to incorporate data from (109)–(110) into a model of the form (111), we plot the output of the EnKF scheme (18) in Figure 5a, noting that it fails to recover the correct parameter value θ_{true}. Figures 5b and 5c show the output of the RP-EnKF scheme (14), with $\tau = 1$ (implying $\Delta \mathbb{Y}_k^{\text{skew}} = 0$) and $\tau = 500$, respectively. We see that the area correction obtained through the subsampling procedure is necessary to obtain satisfactory numerical results. The time-lag $\tau = 500$ has been determined in the same way as in Section 6.2; in particular, plots for the subsampled area processes are qualitatively similar to Figures 4 and 4b but are omitted here for the sake of brevity.

---

[8]In fact, the convergence takes place μ-almost surely, see [63], Theorem 1.1. The covariance Σ is given in terms of suitable long-time ergodic averages or (under certain conditions) Green–Kubo formulae, see [19, 97].

(b) RP-EnKF without area correction.

(a) EnKF according to (18).
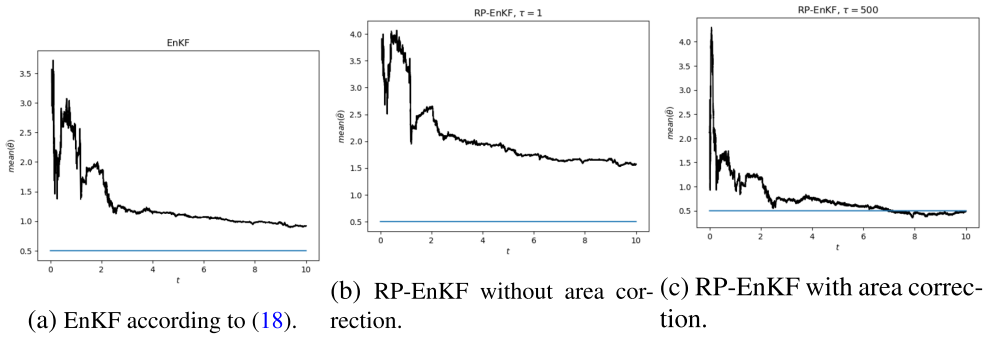
(c) RP-EnKF with area correction.

FIG. 5. *Output (empirical mean of the $\widehat{\theta}$-components) EnKF and RP-EnKF (with and without area correction), for data obtained from the dynamics (109) perturbed by fast chaotic noise (Lorenz-63). The blue line indicates the true value of θ. To suppress sampling error, average results are plotted computed from 5 independent runs.*

6.4. *Homogenisation in a two-scale potential.* In our final example, we consider the motion of a Brownian particle in a (rugged) two-scale potential,

$$(112) \qquad dZ_t^\varepsilon = -\theta \nabla V(Z_t^\varepsilon)dt - \frac{1}{\varepsilon}\nabla p\left(\frac{Z_t^\varepsilon}{\varepsilon}\right)dt + \sqrt{2\sigma}\,dW_t,$$

where $p \in C^\infty(\mathbb{R}^2; \mathbb{R})$, $p(x_1, x_2) = p_1(x_1) + p_2(x_2)$, is an L-periodic function in both directions, that is, $p_i(x + L) = p_i(x)$, for all $x \in \mathbb{R}$ and $i = 1, 2$, modelling small-scale fluctuations around the potential V. It is well known that for $T > 0$, the law of the solution $(Z_t^\varepsilon)_{t \geq 0}$ converges weakly in $C([0, T]; \mathbb{R}^d)$ to the law associated to

$$(113) \qquad dZ_t = -\theta \mathcal{K}\nabla V(Z_t, \theta)\,dt + \sqrt{2\sigma\mathcal{K}}\,dW_t,$$

where $\mathcal{K} = \mathrm{diag}(L^2/(C_1\widehat{C}_1), L^2/(C_2\widehat{C}_2))$, with

$$(114) \qquad C_i = \int_0^L e^{-\frac{p_i(y)}{\sigma}}\,dy, \qquad \widehat{C}_i = \int_0^L e^{\frac{p_i(y)}{\sigma}}\,dy,$$

see [96] and [97], Chapter 11. Similar results in a rough-path context can be found in [78]. The homogenised dynamics (113) encapsulate the rugged landscape described by p in the diffusion-mass matrix $\mathcal{K}$. Like in the previous experiments, we consider the task of estimating the parameter θ from noisy observations of (112), using the EnKF and RP-EnKF based on (113). We choose $V(z) = \frac{1}{2}|z|^2$, $p_1(x) = \cos(x)$, $p_2(x) = \frac{1}{2}\cos(x)$ and $\sigma = 1$. Data from the two-scale dynamics (112) is simulated for $\varepsilon = 10^{-2}$, $\theta_{\text{true}} = 1$ using an Euler–Maruyama discretisation with time step $\Delta t = 10^{-4}$. With the same values for θ_{true}, Δt and σ, we simulate data from the reduced model (113), where $\mathcal{K} \approx \mathrm{diag}(0.62386, 0.884176)$ has been obtained by numerical integration in (114). Both (112) and (113) are perturbed by noise according to (110) with $R = 10^{-2}$ and, as in Sections 6.2 and 6.3, the resulting observation paths are used in the EnKF- and RP-EnKF schemes with $N = 10$ particles (see equations (18) and (14)) to estimate θ_{true}. We display the results for the means of $\widehat{\theta}$ over time using the EnKF and the RP-EnKF in Figures 6 and 7, respectively.

Clearly, the RP-EnKF deals adequately with multiscale data, while the EnKF fails to recover the true parameter θ_{true} in this setting. We would like to stress that the RP-EnKF scheme has been implemented without area correction, that is imposing $\mathbb{Y}_k = \mathbb{Y}_k^{\text{sym}}$ as defined in (87). The choice $\mathbb{Y}_k^{\text{skew}} = 0$ is motivated by a plot analogous and qualitatively similar to Figure 4c (omitted due to space considerations), showing that subsampling does not indicate substantial Lévy area correction terms (intuitively, the dynamics (112) do not contain significant "rotational" contributions in the regime $\varepsilon \to 0$).
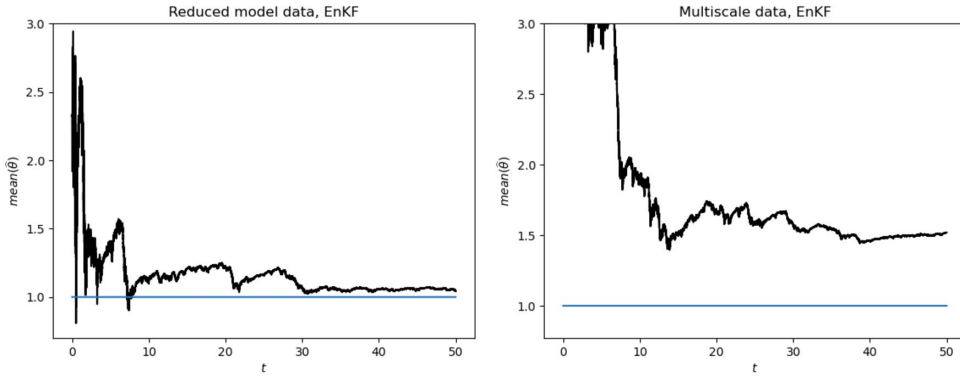
(a) Data obtained from the reduced model (113). (b) Data obtained from the multiscale model (112).

FIG. 6. *Output (empirical mean of the $\widehat{\theta}$-components) of the EnKF-scheme* (18), *for data obtained from the reduced model* (113) *(left) and the multiscale model* (112) *(right). The blue line indicates the true value $\theta_{\text{true}} = 1$. To suppress sampling error, we plot averaged results computed from 5 independent runs.*

APPENDIX A: FROM THE FILTER PROBLEM TO THE MCKEAN–VLASOV EQUATION: PROOFS

In this section we prove the formal connection between the filtering problem and the McKean–Vlasov equation. In order to prove Proposition 2.1, we need a few preparations. Let us define
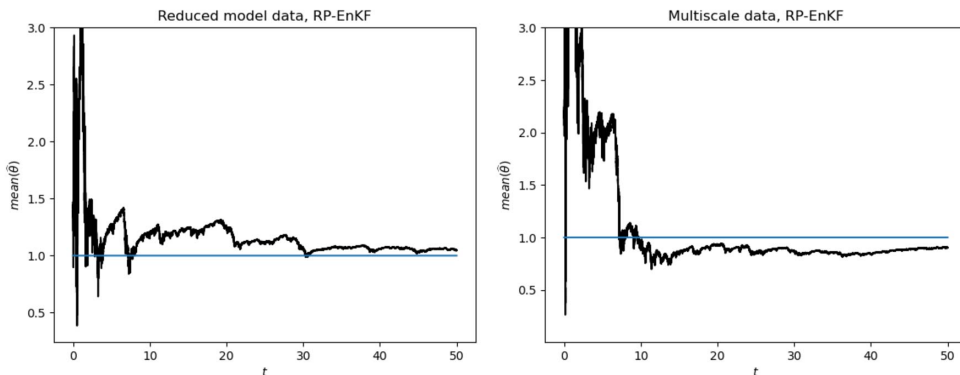
$$M_t = \int_0^t h(X_s) \cdot C^{-1} \, dY_s,$$

and the likelihood

$$l_t = \exp\left(M_t - \frac{1}{2} \langle M \rangle_t \right) = \exp\left(\int_0^t h(X_s) \cdot C^{-1} \, dY_s - \frac{1}{2} \int_0^t h(X_s) \cdot C^{-1} h(X_s) \, ds \right).$$

We can now introduce the unnormalised filtering measures

$$\rho_t[\phi] = \mathbb{E}[\phi(X_t) l_t | \mathcal{Y}_t], \quad t \geq 0.$$

The measures ρ_t satisfy the Zakai equation:

(a) Data obtained from the reduced model (113). (b) Data obtained from the multiscale model (112).

FIG. 7. *Output (empirical mean of the $\widehat{\theta}$-components) of the RP-EnKF-scheme* (14), *for data obtained from the reduced model* (113) *(left) and the multiscale model* (112) *(right). No area correction is used, that is, $\mathbb{Y}_k^{\text{skew}} = 0$. The blue line indicates the true value $\theta_{\text{true}} = 1$. To suppress sampling error, we plot averaged results computed from 5 independent runs.*

PROPOSITION A.1 (Zakai equation). *The evolution of $(\rho_t)_{t\geq 0}$ is given by*

$$\text{(115)} \qquad \rho_t[\phi] = \rho_0[\phi] + \int_0^t \rho_s[\mathcal{L}\phi]\,\mathrm{d}s + \int_0^t \rho_s[\phi h]\cdot C^{-1}\,\mathrm{d}Y_s + \int_0^t \rho_s[\nabla\phi]\cdot B\,\mathrm{d}Y_s,$$

$\mathbb{P}$-*a.s.*, $\forall\phi\in C_b^2(\mathbb{R}^D)$, *where*

$$\mathcal{L}\varphi = \frac{1}{2}\operatorname{Trace}(G D^2\phi) + f\cdot\nabla\phi$$

is the generator associated to (1a). *The filtering measures* $(\pi_t)_{t\geq 0}$ *can be recovered from the Kallianpur–Striebel formula* (*or Bayes theorem*)

$$\text{(116)} \qquad\qquad\qquad \pi_t[\phi] = \frac{\rho_t[\phi]}{\rho_t[\mathbf{1}]}.$$

PROOF. We connect our setting to the one from [10], Section 3.8, by setting $\widetilde{Y}_t := C^{-1/2}Y_t$ and $\widetilde{h}(x) := C^{-1/2}h(x)$, noting that $\widetilde{Y}$ satisfies

$$\text{(117)} \qquad\qquad \mathrm{d}\widetilde{Y}_t = \widetilde{h}(X_t)\,\mathrm{d}t + \mathrm{d}\widetilde{W}_t, \qquad \widetilde{Y}_0 = 0,$$

where $\widetilde{W}_t := C^{-1/2}(U W_t + R^{1/2}V_t)$ is a standard Brownian motion. Within this proof, the tilde indicates that the corresponding object occurs in [10], Section 3.8, (without tilde). We also introduce the vector of operators $\widetilde{B} = (\widetilde{B}_1,\dots,\widetilde{B}_d)^\top$ by $\widetilde{B}_i = (C^{1/2}U G^{1/2}\nabla)_i$, satisfying

$$\text{(118)} \qquad\qquad \langle M^\phi, \widetilde{W}^i\rangle_t = \int_0^t \widetilde{B}_i\phi(X_s)\,\mathrm{d}s, \qquad \phi\in C_b^2(\mathbb{R}^D).$$

Here, M^ϕ is the martingale

$$\text{(119)} \qquad M_t^\phi = \phi(X_t) - \phi(X_0) - \int_0^t \mathcal{L}\phi(X_s)\,\mathrm{d}s, \qquad t\geq 0, \phi\in C_b^2(\mathbb{R}^D),$$

so that (118) corresponds to the equation (3.70) in [10]. The claim therefore follows from [10], Corollary 3.39, by substituting Y and h for $\widetilde{Y}$ and $\widetilde{h}$. $\quad\square$

Next, we write the Zakai equation in its Stratonovich form:

LEMMA A.2. *The Stratonovich-version of the Zakai equation is given by*

$$\text{(120a)} \quad \rho_t[\phi] = \rho_0[\phi] + \int_0^t \rho_s[\mathcal{L}\phi]\,\mathrm{d}s + \int_0^t \rho_s[\phi h]C^{-1}\circ\mathrm{d}Y_s + \int_0^t \rho_s[\nabla\phi]\cdot B\circ\mathrm{d}Y_s$$

$$\text{(120b)} \qquad\qquad - \frac{1}{2}\int_0^t \big(\rho_s[(\phi h)\cdot C^{-1}h] + \rho_s[\operatorname{Trace}(D(\phi h)B) + \nabla\phi\cdot Bh]\big)\,\mathrm{d}s$$

$$\text{(120c)} \qquad\qquad - \frac{1}{2}\int_0^t \rho_s[\operatorname{Trace}(D^2\phi\underbrace{G^{1/2}U^\top C^{-1}U G^{1/2}}_{BCB^\top})]\,\mathrm{d}s.$$

PROOF. Using (115), we see that for $i = 1,\dots,d$

$$\rho_t[\phi h_i] = \sum_{j,k=1}^d \int_0^t \rho_s[\phi h_i h_j](C^{-1})^{jk}\,\mathrm{d}Y_s^k + \sum_{r=1}^D\sum_{k=1}^d \int_0^t \rho_s[\partial_r(\phi h_i)]B^{rk}\,\mathrm{d}Y_s^k + FV,$$

where FV stands for a contribution of finite variation. Recalling $\langle Y^k, Y^{\bar{k}} \rangle_t = C^{k\bar{k}}t$, for $k, \bar{k} = 1, \ldots, d$, we obtain

$$\int_0^t \rho_s[\phi h]C^{-1}\,\mathrm{d}Y_s = \int_0^t \rho_s[\phi h]C^{-1} \circ \mathrm{d}Y_s$$
$$- \frac{1}{2}\int_0^t \sum_{i=1}^d \left(\sum_{j=1}^d \rho_s[\phi h_i h_j](C^{-1})^{ij} + \sum_{j=1}^D \rho_s[\partial_j(\phi h_i)B^{ji}] \right) \mathrm{d}s.$$

Similarly, for $r = 1, \ldots, D$,

$$\rho_t[\partial_r \phi] = \sum_{j,k=1}^d \int_0^t \rho_s[(\partial_r\phi)h_j](C^{-1})^{jk}\,\mathrm{d}Y_s^k + \sum_{p=1}^D \sum_{k=1}^d \int_0^t \rho_s[\partial_r\partial_p\phi]B^{pk}\,\mathrm{d}Y_s^k + FV,$$

implying

$$\int_0^t \rho_s[\nabla\phi]B\,\mathrm{d}Y_s = \int_0^t \rho_s[\nabla\phi]B \circ \mathrm{d}Y_s$$
$$- \frac{1}{2}\int_0^t \sum_{r=1}^D \left(\sum_{i=1}^d \rho_s[(\partial_r\phi)h_i]B^{ri} + \sum_{p=1}^D \rho_s[\partial_r\partial_p\phi](BCB^\top)^{rp} \right) \mathrm{d}s. \qquad \square$$

We now proceed to the proof of Proposition 2.1:

PROOF OF PROPOSITION 2.1.   Let us define

$$\widehat{\rho}_t[\phi] = \mathbb{E}[\phi(\widehat{X}_t)l_t | \mathcal{Y}_t], \quad t \geq 0,$$

which is the unnormalised filtering measure associated to (5). Since X_t and $\widehat{X}_t$ are independent given $\mathcal{Y}_t$, we have that

$$\widehat{\rho}_t[\phi] = \mathbb{E}[\phi(\widehat{X}_t) | \mathcal{Y}_t]\mathbb{E}[l_t | \mathcal{Y}_t] = \widehat{\pi}_t[\phi]\rho_t[\mathbf{1}].$$

From (116) it thus follows that $\pi_t = \widehat{\pi}_t$ is equivalent to $\rho_t = \widehat{\rho}_t$. In the following we therefore compute the evolution of $\widehat{\rho}_t$. Notice first that

$$M_t = \int_0^t h(X_s) \cdot C^{-1}\,\mathrm{d}Y_s = \int_0^t h(X_s) \cdot C^{-1} \circ \mathrm{d}Y_s - \frac{1}{2}\langle h(X)C^{-1}, Y \rangle_t,$$

where, using $B = G^{\frac{1}{2}}U^\top C^{-1}$,

$$\langle h(X)C^{-1}, Y \rangle_t = \sum_{i,j=1}^d \langle h^i(X)(C^{-1})_{ij}, Y^j \rangle_t = \sum_{r=1}^D \sum_{i=1}^d \int_0^t \partial_r h^i(X_s)B^{ri}\,\mathrm{d}s$$
$$= \int_0^t \mathrm{Trace}(Dh(X_s) \cdot B)\,\mathrm{d}s.$$

From $\mathrm{d}l_t = l_t\,\mathrm{d}M_t$ we have

$$l_t = 1 + \int_0^t l_s \circ \mathrm{d}M_s - \frac{1}{2}\int_0^t l_s\,\mathrm{d}\langle M \rangle_s$$
$$= 1 + \int_0^t l_s h(X_s) \circ C^{-1}\,\mathrm{d}Y_s - \frac{1}{2}\int_0^t l_s\,\mathrm{Trace}(Dh(X_s) \cdot B)\,\mathrm{d}s$$
$$- \frac{1}{2}\int_0^t l_s h(X_s) \cdot C^{-1}h(X_s)\,\mathrm{d}s,$$

implying

(121a)

$$\phi(\widehat{X}_t)l_t = \phi(\widehat{X}_0) + \int_0^t \phi(\widehat{X}_s)l_s h(X_s) \cdot C^{-1} \circ dY_s$$

$$- \frac{1}{2}\int_0^t \phi(\widehat{X}_s)l_s h(X_s) \cdot C^{-1}h(X_s)\,ds$$

(121b)

$$- \frac{1}{2}\int_0^t \phi(\widehat{X}_s)l_s \operatorname{Trace}(Dh(X_s) \cdot B)\,ds + \int_0^t l_s \nabla\phi(\widehat{X}_s) \circ d\widehat{X}_s.$$

The last term satisfies

$$\int_0^t l_s \nabla\phi(\widehat{X}_s) \circ d\widehat{X}_s$$

$$= \int_0^t l_s \nabla\phi(\widehat{X}_s) \cdot f(\widehat{X}_s)\,ds + \int_0^t l_s \nabla\phi(\widehat{X}_s) \circ G^{1/2}\,d\widehat{W}_s$$

$$+ \int_0^t l_s \nabla\phi(\widehat{X}_s) K_s(\widehat{X}_s)C^{-1} \circ dY_s - \int_0^t l_s \nabla\phi(\widehat{X}_s) \cdot K_s(\widehat{X}_s)C^{-1}h(\widehat{X}_s)\,ds$$

$$- \int_0^t l_s \nabla\phi(\widehat{X}_s) \cdot K_s(\widehat{X}_s)C^{-1} \circ (U\,d\widehat{W}_s + R^{1/2}\,d\widehat{V}_s) + \int_0^t l_s \nabla\phi(\widehat{X}_s) \cdot \Xi_s(\widehat{X}_s)\,ds.$$

Next, we convert the terms involving $\widehat{V}$ and $\widehat{W}$ back to their Itô-form,

$$\int_0^t l_s \nabla\phi(\widehat{X}_s) \circ G^{1/2}\,d\widehat{W}_s$$

$$= \int_0^t l_s \nabla\phi(\widehat{X}_s)G^{1/2}\,d\widehat{W}_s$$

$$+ \frac{1}{2}\int_0^t l_s \operatorname{Trace}(D^2\phi(\widehat{X}_s)G)\,ds - \frac{1}{2}\int_0^t l_s \operatorname{Trace}(D^2\phi(\widehat{X}_s)(K_s(\widehat{X}_s)B^\top))\,ds.$$

Here we used the fact that, for every $i,k = 1,\ldots,D$ we have $\langle l, (G^{1/2})^{ik}\widehat{W}^k\rangle_t = 0$ and

$$\langle \widehat{X}^j, (G^{1/2})^{ik}\widehat{W}^k\rangle_t = G^{ij}t - \int_0^t (K_s(\widehat{X}_s)C^{-1}UG^{1/2})^{ji}\,ds,$$

where $G^{1/2}U^\top C^{-1} = B$. Similarly,

$$\int_0^t l_s \nabla\phi(\widehat{X}_s) \cdot K_s(\widehat{X}_s)C^{-1} \circ (U\,d\widehat{W}_s + R^{1/2}\,d\widehat{V}_s)$$

$$= \int_0^t l_s \nabla\phi(\widehat{X}_s) \cdot K_s(\widehat{X}_s)C^{-1}(U\,d\widehat{W}_s + R^{1/2}\,d\widehat{V}_s)$$

$$+ \frac{1}{2}\int_0^t l_s \operatorname{Trace}\big[D(\nabla\phi \cdot K_s)(\widehat{X}_s) \cdot C^{-1}(G^{1/2}U^\top - K_s(\widehat{X}_s))\big]\,ds,$$

using $C = UU^\top + R$ and, for $k = 1,\ldots,D$ and $l = 1,\ldots,d$,

$$\langle \widehat{X}^k, (U\widehat{W} + R^{1/2}\widehat{V})^l\rangle_t = (G^{1/2}U^\top)^{kl}t - \int_0^t K_s^{kl}(\widehat{X}_s)\,ds.$$

We now take the conditional expectation in (121). Note that the conditional expectation with respect to $\mathcal{Y}$ commutes with the integration in dY and that the conditional expectation of the integrals in dW and dV vanishes because (W, V) are independent from $\mathcal{Y}$, see [25],

Appendix B. We obtain

$$(125a) \quad \widehat{\rho}_t[\phi] = \widehat{\rho}_0[\phi] + \int_0^t \widehat{\rho}_s[\phi]\pi_s[h] \cdot C^{-1} \circ dY_s - \frac{1}{2}\int_0^t \widehat{\rho}_s[\phi]\pi_s[h \cdot C^{-1}h] \, ds$$

$$(125b) \quad + \int_0^t \widehat{\rho}_s[\mathcal{L}\phi] \, ds - \frac{1}{2}\int_0^t \widehat{\rho}_s\big[\text{Trace}(D^2\phi K_s B^\top)\big] \, ds$$

$$+ \int_0^t \widehat{\rho}_s[\nabla\phi \cdot K_s]C^{-1} \circ dY_s$$

$$- \int_0^t \widehat{\rho}_s[\nabla\phi \cdot K_s C^{-1}h] \, ds$$

$$(125c) \quad - \frac{1}{2}\int_0^t \widehat{\rho}_s\big[\text{Trace}(D(\nabla\phi \cdot K_s) \cdot C^{-1}(G^{1/2}U^\top - K_s))\big] \, ds.$$

$$- \frac{1}{2}\int_0^t \widehat{\rho}_s[\phi]\pi_s\big[\text{Trace}(DhB)\big] \, ds$$

$$(125d) \quad + \int_0^t \widehat{\rho}_s[\nabla\phi \cdot \Xi] \, ds.$$

Importantly, we have used the fact that X_t and $\widehat{X}_t$ are independent given $\mathcal{Y}_t$, and also that $\widehat{W}_t$ and $\widehat{V}_t$ are independent from $\mathcal{Y}_t$. The next step is to compare (120) and (125). The dY-contributions agree if and only if

$$(126) \qquad \rho_s[\phi h] + \rho_s[\nabla\phi] \cdot BC = \widehat{\rho}_s[\phi]\pi_s[h] + \widehat{\rho}_s[\nabla\phi \cdot K_s],$$

which we recognise (after identifying $\widehat{\pi} = \pi$ and $\widehat{\rho} = \rho$) to be a weak version of (6).

To compare the ds-contributions, let us first manipulate the second term in (125c) using (6),

$$\widehat{\rho}_s\big[\partial_m(\partial_i\phi K_s^{ij})(C^{-1})^{jk}(G^{1/2}U^\top - K_s)^{mk}\big]$$

$$= -\int (\partial_i\phi K_s^{ij})(C^{-1})^{jk}\partial_m\big(\widehat{\rho}_s(G^{1/2}U^\top - K_s)^{mk}\big) \, dx$$

$$= -\int (\partial_i\phi K_s^{ij})(C^{-1})^{jk}(h^k - \widehat{\pi}_s[h^k]) \, d\widehat{\rho}_s = -\widehat{\rho}_s[\nabla\phi \cdot K_s C^{-1}(h - \widehat{\pi}_s[h])].$$

Similarly, the middle term in (125b) satisfies

$$\widehat{\rho}_s\big[(\partial_i\partial_j\phi)K_s^{il}\big]B^{jl} = -B^{jl}\int (\partial_j\phi)\partial_i\big(\widehat{\rho}_s K_s^{il}\big) \, dx$$

$$= B^{jl}\widehat{\rho}_s\big[(\partial_j\phi)(h^l - \widehat{\pi}_s[h^j])\big] - B^{jl}\int (\partial_j\phi)\partial_i\big(\widehat{\rho}_s(BC)^{il}\big) \, dx$$

$$= B^{jl}\widehat{\rho}_s\big[(\partial_j\phi)(h^l - \widehat{\pi}_s[h^l])\big] + \widehat{\rho}_s[\partial_i\partial_j\phi](BCB^\top)^{ij}$$

$$= \widehat{\rho}_s\big[\nabla\phi \cdot B(h - \widehat{\pi}_s[h])\big] + \widehat{\rho}_s\big[\text{Trace}(D^2\phi(BCB^\top))\big].$$

We now collect terms and compare the ds-contributions in (120) and (125), arriving at

$$(128a) \quad -\frac{1}{2}\big(\rho_s[\phi h \cdot C^{-1}h] + \rho_s[\text{Trace}(D(\phi h)B) + \nabla\phi \cdot Bh]\big)$$

$$(128b) \quad = -\frac{1}{2}\widehat{\rho}_s[\phi]\pi_s[h \cdot C^{-1}h] - \frac{1}{2}\widehat{\rho}_s[\nabla\phi \cdot B(h - \widehat{\pi}_s[h])]$$

$$(128c) \quad - \frac{1}{2}\widehat{\rho}_s[\nabla\phi \cdot K_s C^{-1}(h + \widehat{\pi}_s[h])] + \widehat{\rho}_s[\nabla\phi \cdot \Xi_s] - \frac{1}{2}\widehat{\rho}_s[\phi]\pi_s[\text{Trace}(DhB)].$$

Next, we work on the first term in (128c),

$$
\widehat{\rho}_s[(\partial_i \phi K_s^{ij})(C^{-1})^{jk}(h^k + \widehat{\pi}_s[h^k])]
$$

$$
= -\int \phi \partial_i (\widehat{\rho}_s K_s^{ij})(C^{-1})^{jk}(h^k + \widehat{\pi}_s[h^k])\,\mathrm{d}x
$$

$$
\quad - \widehat{\rho}_s[\phi K_s^{ij}(C^{-1})^{jk}\partial_i h^k]
$$

$$
= \int \phi(h^j - \pi[h^j])(C^{-1})^{jk}(h^k + \pi[h^k])\,\mathrm{d}\widehat{\rho}_s
$$

$$
\quad - \int \phi \partial_i (\widehat{\rho}_s(BC)^{ij})(C^{-1})^{jk}(h^k + \widehat{\pi}_s[h^k])\,\mathrm{d}x - \widehat{\rho}_s[\phi K_s^{ij}(C^{-1})^{jk}\partial_i h^k]
$$

$$
= \widehat{\rho}_s[\phi(h^j - \pi[h^j])(C^{-1})^{jk}(h^k + \widehat{\pi}_s[h^k])] + \widehat{\rho}_s[\partial_i \phi(h^k + \widehat{\pi}_s[h^k])]B^{jk}
$$

$$
\quad + \widehat{\rho}_s[\phi B^{ik}\partial_i h^k] - \widehat{\rho}_s[\phi K_s^{ij}(C^{-1})^{jk}\partial_i h^k]
$$

$$
= \widehat{\rho}_s[\phi h^j h^k](C^{-1})^{jk} - \widehat{\rho}_s[\phi]\pi_s[h^j]\pi_s[h^k](C^{-1})^{jk}
$$

$$
\quad + \widehat{\rho}_s[\partial_i \phi(h^k + \widehat{\pi}_s[h^k])]B^{jk} + \widehat{\rho}_s[\phi B^{ik}\partial_i h^k] - \widehat{\rho}_s[\phi K_s^{ij}(C^{-1})^{jk}\partial_i h^k]
$$

$$
= \widehat{\rho}_s[\phi h C^{-1} h] - \widehat{\rho}_s[\phi]\pi_s[h]C^{-1}\pi_s[h] + \widehat{\rho}_s[\nabla\phi \cdot B(h + \widehat{\pi}_s[h])]
$$

$$
\quad + \widehat{\rho}_s[\phi \operatorname{Trace}(Dh B)] - \widehat{\rho}_s[\phi \operatorname{Trace}(K_s C^{-1} Dh)].
$$

Plugging this into (128) we see that after a great number of cancellations that (128) reduces to

$$
\rho_s[\nabla\phi \cdot \Xi_s] = \frac{1}{2}\rho_s[\phi](\pi_s[\operatorname{Trace}(Dh \cdot B) - hC^{-1}h] - \pi_s[h]C^{-1}\pi_s[h])
$$

$$
\quad - \frac{1}{2}\rho_s[\phi \operatorname{Trace}(K_s C^{-1} Dh)].
$$

Using (6) in its weak formulation tested against h on the first term in the right-hand side we obtain

$$
(129) \qquad \pi_s[\nabla\phi \cdot \Xi_s] = \frac{1}{2}(\pi_s[\phi]\pi_s[\operatorname{Trace}(K_s C^{-1} Dh)] - \pi_s[\phi \operatorname{Trace}(K_s C^{-1} Dh)]),
$$

which is a weak version of (7).

We have thus obtained constraints on the coefficients of K and Ξ of the McKean–Vlasov equation (5) such that $\hat{\rho}$ solves the Zakai equation satisfied by the conditional density ρ of the signal of the filtering problem. We can conclude because the solution to the Zakai equation is assumed to be unique. $\square$

We conclude this section with the proof of Lemma 2.2

PROOF OF LEMMA 2.2.   The variance (36) is minimised when $\widetilde{K} = \pi[K]$. The weak formulation of (6) is given as

$$
(130) \qquad \pi[\nabla\phi \cdot (K - BC)] = \pi[(h - \pi[h])^\top \phi],
$$

where $\phi \in C_b^2(\mathbb{R}^D)$, we derive this equation in (126).

To compute the expectation of K, we would like to test equation (130) on $\phi(x) = x^\gamma$, $\gamma = 1, \ldots, D$, so that $\nabla\phi(x) = e_\gamma$, the γth elemental vector. However, we only assume that equation (130) is satisfied for bounded function. We take a sequence $\phi^n \in C_b^2$ that coincides with x^γ on $B(0, n)$, the ball centered in the origin with radius n, and satisfies $|\nabla\phi^n| \leq |\nabla\phi| =$

1. Now we evaluate (130) in ϕ^n and then we pass to the limit on both sides. We use the Cauchy–Schwartz inequality and the integrability assumption on K to obtain

$$\left|\pi\left[(\nabla\phi^n - \nabla\phi)\cdot(K - BC)\right]\right| \leq \pi\left[1_{B(0,n)^c}\|K - BC\|_F\right] \leq \pi\left[1_{B(0,n)^c}\right]^{\frac{1}{2}}\pi\left[\|K - BC\|_F^2\right]^{\frac{1}{2}}$$

$$\lesssim \pi\left[1_{B(0,n)^c}\right]^{\frac{1}{2}}.$$

The last term goes to 0 by tightness of π. Notice that BC is a constant, so it is always integral with respect to π. Similarly, for the right hand side of (130) we obtain

$$\left|\pi\left[(h - \pi[h])^\top(\phi^n - x^\gamma)\right]\right| \lesssim \|h\|_{C_b}\pi\left[1_{B(0,n)^c}|x|\right],$$

the last term goes to 0 as $n \to \infty$ as we assumed $\pi \in \mathcal{P}_1(\mathbb{R}^D)$.

We proved that (130) is satisfied for $\phi(x) = x^\gamma$, this gives us an equation for the γth row of $\pi[K]$

$$\pi[K^\gamma] - (BC)^\gamma = \int_{\mathbb{R}^D} x^\gamma\left(h(x) - \pi[h]\right)^\top d\pi(x), \quad \gamma = 1, \ldots, D.$$

Hence, $\tilde{K} = \pi[K] = \mathrm{Cov}_\pi(x, h) + BC$.

We now show the corresponding statement for $\tilde{\Xi}$ when K is replaced by $\tilde{K}$ in (7). Again $\tilde{\Xi} = \pi[\Xi]$, since the expectation minimises the variance. For $1 \leq \gamma \leq D$, we input the test function $\phi(x) = x^\gamma$ into the weak formulation

$$\pi[\nabla\phi\cdot\Xi_s] = \frac{1}{2}\left(\pi[\phi]\pi\left[\mathrm{Trace}(\tilde{K}C^{-1}Dh)\right] - \pi\left[\phi\,\mathrm{Trace}(\tilde{K}C^{-1}Dh)\right]\right),$$

of (7) to obtain the desired expression for $\tilde{\Xi}$. The weak formulation is derived rigorously in (129). $\square$

APPENDIX B: ON THE FORMAL CONNECTION BETWEEN MCKEAN–VLASOV FILTERING AND MAXIMUM LIKELIHOOD ESTIMATION

Here we discuss the relationship between parameter estimation based on the ensemble Kalman filter and the maximum likelihood approach considered in [40]. In a nutshell, the two methods essentially agree in the noiseless case $R = 0$ with Gaussian initial condition, as relevant McKean–Vlasov dynamics can be solved explicitly in this case. This connection has already been pointed out in [88].

Notice that for the particular choice of the coefficients f and h (see equation (132) below) Assumption 3 and 4 required by our main theorem are not satisfied. Because of this, the connection of our work with maximum likelihood estimation is only established at a formal level.

We are given a d-dimensional process Z, which depends on an m-dimensional parameter θ. We observe Z through the observation Y, which is possibly noisy.

(131a) $$dZ_t = g(Z_t)\theta\,dt + \tilde{G}^{\frac{1}{2}}\,dW_t,$$

(131b) $$d\theta_t = 0,$$

(131c) $$dY_t = dZ_t + R^{\frac{1}{2}}\,dV_t.$$

Here W and V are independent d-dimensional Brownian motions, $g : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ is a given C_b^2 function and $\tilde{G} \in \mathbb{R}^{d\times d}$ is a given positive semidefinite deterministic matrix.

Let $D = d + m$. We set the D-dimensional signal process as $X := (Z, \theta)$ and replace the equation for dZ into the equation for the observation dY. System (131) becomes equivalent to system (1) with the following drift coefficients:

(132) $$f(x) = f(z, \theta) = (g(z)\theta, 0), \qquad h(x) = h(z, \theta) = g(z)\theta.$$

Moreover, the coeffiecents of the noise are

$$G = \begin{pmatrix} \tilde{G} & 0 \\ 0 & 0 \end{pmatrix}, \qquad U = \begin{pmatrix} \tilde{G}^{\frac{1}{2}} & 0 \end{pmatrix}.$$

Remember $\pi_t[\phi] := \mathbb{E}[\phi(X_t) \mid \mathcal{Y}_t]$ is the conditional law of X given Y.

First we consider noiseless observations, that is, we set $R = 0$. In this case we have that

$$C := UU^\top + R = \tilde{G}, \qquad B := G^{\frac{1}{2}}U^\top C^{-1} = \begin{pmatrix} \mathrm{Id}_{d \times d} \\ 0_{m \times m} \end{pmatrix}.$$

We interpret system (131) as a filtering problem, as described in Section 2.3. Then we can construct the corresponding (approximate) McKean–Vlasov equation. We need to compute P as defined in (8), which can be decomposed as follows:

$$P(\pi) = \begin{pmatrix} \mathrm{Cov}_\pi(z, h) \\ \mathrm{Cov}_\pi(\theta, h) \end{pmatrix} C^{-1} + B := \begin{pmatrix} \pi\big[z(h - \pi[h])^\top\big] \\ \pi\big[\theta(h - \pi[h])^\top\big] \end{pmatrix} C^{-1} + B.$$

In this case with $R = 0$, we have $Y = Z$, which implies

$$\pi_t(dz, d\theta) = \delta(z - Z_t)\pi^\theta(d\theta),$$

$$\pi_t[h] = \int_{\mathbb{R}^D} g(z)\theta \pi_t(dz, d\theta) = g(Z_t) \int_{\mathbb{R}^m} \theta \pi_t^\theta(d\theta) =: g(Z_t)\bar{\theta}_t.$$

We have the following:

$$\mathrm{Cov}_{\pi_t}(z, h) = \int_{\mathbb{R}^{d+m}} z\big(g(z)\theta - g(z)\bar{\theta}_t\big)^\top \pi_t(dz, d\theta)$$

$$= Z_t\left(\int_{\mathbb{R}^m}(\theta - \bar{\theta})\pi_t^\theta(d\theta)\right)^\top g(Z_t)^\top = 0_{d \times d},$$

$$\mathrm{Cov}_{\pi_t}(\theta, h) = \int_{\mathbb{R}^m} \theta(\theta - \bar{\theta}_t)^\top \pi_t^\theta(d\theta) g(Z_t)^\top = \mathrm{Var}(\theta_t)g(Z_t)^\top.$$

We have thus obtained P explicitly

$$P(\pi_t) = \begin{pmatrix} \mathrm{Cov}_\pi(z, h) \\ \mathrm{Cov}_\pi(\theta, h) \end{pmatrix} C^{-1} + B = \begin{pmatrix} 0_{d \times d} \\ \mathrm{Var}(\theta_t)g(Z_t)^\top \end{pmatrix} \tilde{G}^{-1} + \begin{pmatrix} \mathrm{Id}_{d \times d} \\ 0_{m \times m} \end{pmatrix}$$

$$= \begin{pmatrix} \mathrm{Id}_{d \times d} \\ \mathrm{Var}(\theta_t)g(Z_t)^\top \tilde{G}^{-1} \end{pmatrix}.$$

We now compute $\widehat{\Gamma}$, which is defined in (9). We start with the z contribution, for $\gamma = 1, \ldots, d$, we have

$$\pi\big[z^\gamma (Dh - \pi[Dh])^\top\big] = Z_t^\gamma \pi\big[Dh - \pi[Dh]\big]^\top = 0_{D \times d}.$$

Now compute the contribution in θ. For $\gamma = 1, \ldots, m$ and $i = 1, \ldots, d$, we have

$$\pi_t\big[\bar{\theta}^\gamma Dh^\top\big]_{j,i} = \sum_{k=1}^m \partial_{z^j} g^{i,k}(Z_t) \mathrm{Cov}(\theta_t^k, \theta_t^\gamma) 1_{j \le d} = \big(\partial_{z^j} g(Z_t) \mathrm{Var}(\theta_t)\big)^{i,\gamma} 1_{j \le d},$$

$j = 1, \ldots, d + m$. So that we have

$$P(\pi_t)^\top \pi_t\big[\theta^\gamma (Dh - \pi_t[Dh])^\top\big] = \begin{pmatrix} \mathrm{Id}_{d \times d} & (\mathrm{Var}(\theta_t)g(Z_t)^\top \tilde{G}^{-1})^\top \end{pmatrix} \begin{pmatrix} Dg(Z_t)^\top \mathrm{Var}(\theta_t)^{\cdot \gamma} \\ 0_{m \times d} \end{pmatrix}$$

$$= Dg(Z_t)^\top \mathrm{Var}(\theta_t)^{\cdot \gamma}$$

and

$$\Gamma^{\gamma}(\pi_t) = -\frac{1}{2}\operatorname{Trace}(Dg(Z_t)\operatorname{Var}(\theta_t)^{\cdot\gamma}).$$

If $m = 1$, we have

$$\Gamma(\pi_t) = \begin{pmatrix} 0_d \\ -\frac{1}{2}\operatorname{Var}(\theta_t)\operatorname{Trace} Dg(Z_t) \end{pmatrix}.$$

We can now write the McKean–Vlasov dynamics (11),

(133a) $$\mathrm{d}Z_t = \mathrm{d}\mathbf{Y}_t,$$

(133b)
$$\begin{aligned}\mathrm{d}\widehat{\theta}_t &= \operatorname{Var}(\widehat{\theta}_t)g(Z_t)^{\top}\tilde{G}^{-1}\big[\mathrm{d}\mathbf{Y}_t - \tilde{G}^{\frac{1}{2}}\,\mathrm{d}\widehat{W}_t - g(Z_t)\widehat{\theta}_t\,\mathrm{d}t\big] \\ &\quad - \frac{1}{2}\operatorname{Var}(\widehat{\theta}_t)\operatorname{Trace} Dg(Z_t)\,\mathrm{d}t.\end{aligned}$$

We interpret equation (133) in the sense of Definition 4.12. Taking expectation in (133b), and writing this equation in integral form, we obtain the rough path maximum likelihood estimator from [40], up to the rescaling $\operatorname{Var}(\widehat{\theta}_t)$, see also [88].

We would like to obtain an equation for the variance of $\widehat{\theta}$. We start by computing the following equation for the centered random variable $\bar{\theta} := \widehat{\theta} - \pi[\theta]$:

$$\mathrm{d}\bar{\theta}_t = -\operatorname{Var}(\widehat{\theta}_t)g(Z_t)^{\top}\tilde{G}^{-1}\big[\tilde{G}^{\frac{1}{2}}\,\mathrm{d}\widehat{W}_t + g(Z_t)\bar{\theta}_t\,\mathrm{d}t\big].$$

By taking the square and using Itô formula we obtain the following equation for $\operatorname{Var}(\widehat{\theta})$:

$$\mathrm{d}\operatorname{Var}(\widehat{\theta}_t) = -\operatorname{Var}(\widehat{\theta}_t)^2 g(\mathbf{Y}_t)^{\top}\tilde{G}^{-1}g(\mathbf{Y}_t)\,\mathrm{d}t.$$

The solution is

$$\operatorname{Var}(\widehat{\theta}_t) = \left(\int_0^t g(\mathbf{Y}_s)^{\top}\tilde{G}^{-1}g(\mathbf{Y}_s)\,\mathrm{d}s + \frac{1}{\operatorname{Var}(\widehat{\theta}_0)}\right)^{-1}.$$

Assume now that there exists $f : \mathbb{R}^d \to \mathbb{R}$ such that $g = \tilde{G}^{\top}\nabla f$. Since the variance is a path of bounded variation, we have that the integrand in the rough integral in equation (133b) is controlled by $\mathbf{Y}$ in the following way:

$$\begin{aligned}\operatorname{Var}(\widehat{\theta}_t)g(Y_t)^{\top}\tilde{G}^{-1} - \operatorname{Var}(\widehat{\theta}_s)g(Y_s)^{\top}\tilde{G}^{-1} &= \operatorname{Var}(\widehat{\theta}_t)\nabla f(Y_t)^{\top} - \operatorname{Var}(\widehat{\theta}_s)\nabla f(Y_s)^{\top} \\ &= \operatorname{Var}(\widehat{\theta}_t)D^2 f(Y_t)\delta Y_{s,t} + R_{s,t}.\end{aligned}$$

The Gubinelli derivative $\operatorname{Var}(\widehat{\theta}_t)D^2 f(Z_t)$ is symmetric, which means that the rough integral in (133b) does not depend on the area of the geometric rough path $\mathbf{Y}$. Hence, the expansion of the integral is completely determined by the path itself.

APPENDIX C: ITÔ–STRATONOVICH CORRECTION FOR THE ENSEMBLE KALMAN FILTER

In Section 2.4, we derived the system (10) by approximately solving the PDEs (6) and (7). An alternative approach is to convert the ensemble Kalman filter formulation

(134)
$$\begin{cases} \mathrm{d}\widehat{X}_t = f(\widehat{X}_t)\,\mathrm{d}t + G^{1/2}\,\mathrm{d}\widehat{W}_t + P(\widehat{\pi}_t)\,\mathrm{d}I_t, \\ \mathrm{d}I_t = \mathrm{d}Y_t - (h(\widehat{X}_t)\,\mathrm{d}t + U\,\mathrm{d}\widehat{W}_t + R^{1/2}\,\mathrm{d}\widehat{V}_t), \end{cases}$$

from [88] to its Stratonovich form. Here we show that this leads to the same result.

LEMMA C.1.   *For the solution to* (134), *we have that*

$$(135) \qquad \int_0^t P(\widehat{\pi}_s)\,\mathrm{d}I_t = \int_0^t P(\widehat{\pi}_s)\circ\mathrm{d}I_t + \int_0^t \Gamma(\widehat{\pi}_s)\,\mathrm{d}s,$$

thus recovering (10) *from* (134).

PROOF.   Direct calculation (using Itô's formula for the expression (8) and applying the conditional expectation) shows that

$$(136) \quad \mathrm{d}P^{ij}(\widehat{\pi}_t) = \big(\widehat{\pi}_t[x^i \partial_{x_l}h^k P^{lm}(\widehat{\pi}_t)] - \widehat{\pi}_t[x^i]\widehat{\pi}_t[\partial_{x_l}h^k P^{lm}(\widehat{\pi}_t)]\big)(C^{-1})^{kj}\,\mathrm{d}Y^m + FV,$$

where FV stands for a process of finite variation. The claim therefore follows from the conversion formula

$$(137) \qquad \int_0^t P(\widehat{\pi}_s)\,\mathrm{d}I_t = \int_0^t P(\widehat{\pi}_s)\circ\mathrm{d}I_t - \frac{1}{2}\langle P(\widehat{\pi}.),I.\rangle = \int_0^t P(\widehat{\pi}_s)\circ\mathrm{d}I_t - \frac{1}{2}\langle P(\widehat{\pi}.),Y.\rangle,$$

together with the fact that $\langle Y,Y\rangle_t = Ct$.   □

REFERENCES

[1] ABDULLE, A., GAREGNANI, G., PAVLIOTIS, G. A., STUART, A. M. and ZANONI, A. (2023). Drift estimation of multiscale diffusions based on filtered data. *Found. Comput. Math.* **23** 33–84. MR4546144 https://doi.org/10.1007/s10208-021-09541-9

[2] AIT-SAHALIA, Y., MYKLAND, P. A. and ZHANG, L. (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *Rev. Financ. Stud.* **18** 351–416.

[3] ALLAN, A. L. (2021). Robust filtering and propagation of uncertainty in hidden Markov models. *Electron. J. Probab.* **26** Paper No. 73, 37 pp. MR4269203 https://doi.org/10.1214/21-ejp633

[4] ALLAN, A. L. and COHEN, S. N. (2020). Pathwise stochastic control with applications to robust filtering. *Ann. Appl. Probab.* **30** 2274–2310. MR4149529 https://doi.org/10.1214/19-AAP1558

[5] AZENCOTT, R., BERI, A., JAIN, A. and TIMOFEYEV, I. (2013). Sub-sampling and parametric estimation for multiscale dynamics. *Commun. Math. Sci.* **11** 939–970. MR3061147 https://doi.org/10.4310/CMS.2013.v11.n4.a3

[6] AZENCOTT, R., BERI, A. and TIMOFEYEV, I. (2010). Adaptive sub-sampling for parametric estimation of Gaussian diffusions. *J. Stat. Phys.* **139** 1066–1089. MR2646501 https://doi.org/10.1007/s10955-010-9975-y

[7] BAILLEUL, I., CATELLIER, R. and DELARUE, F. (2020). Solving mean field rough differential equations. *Electron. J. Probab.* **25** Paper No. 21, 51 pp. MR4073682 https://doi.org/10.1214/19-ejp409

[8] BAILLEUL, I., CATELLIER, R. and DELARUE, F. (2021). Propagation of chaos for mean field rough differential equations. *Ann. Probab.* **49** 944–996. MR4255135 https://doi.org/10.1214/20-aop1465

[9] BAILLEUL, I. and DIEHL, J. (2015). The inverse problem for rough controlled differential equations. *SIAM J. Control Optim.* **53** 2762–2780. MR3392481 https://doi.org/10.1137/140995982

[10] BAIN, A. and CRISAN, D. (2009). *Fundamentals of Stochastic Filtering. Stochastic Modelling and Applied Probability* **60**. Springer, New York. MR2454694 https://doi.org/10.1007/978-0-387-76896-0

[11] BÁLINT, P. and MELBOURNE, I. (2018). Statistical properties for flows with unbounded roof function, including the Lorenz attractor. *J. Stat. Phys.* **172** 1101–1126. MR3830300 https://doi.org/10.1007/s10955-018-2093-y

[12] BEESON, R., NAMACHCHIVAYA, N. and PERKOWSKI, N. (2020). Quantitative convergence of the filter solution for multiple timescale nonlinear systems with coarse-grain correlated noise. Preprint. Available at arXiv:2011.12801.

[13] BEESON, R., YEONG, H. C., NAMACHCHIVAYA, N. S. and PERKOWSKI, N. (2018). Reduced order nonlinear filters for multi-scale systems with correlated sensor noise. In 2018 21*st International Conference on Information Fusion* (*FUSION*) 131–141. IEEE.

[14] BERRY, T. and SAUER, T. (2018). Correlation between system and observation errors in data assimilation. *Mon. Weather Rev.* **146** 2913–2931.

[15] BISHOP, A. N. and DEL MORAL, P. (2023). On the mathematical theory of ensemble (linear-Gaussian) Kalman–Bucy filtering. *Math. Control Signals Systems* **35** 835–903. MR4645157

[16] BURGERS, G., JAN VAN LEEUWEN, P. and EVENSEN, G. (1998). Analysis scheme in the ensemble Kalman filter. *Mon. Weather Rev.* **126** 1719–1724.

[17] CARMONA, R. and DELARUE, F. (2018). *Probabilistic Theory of Mean Field Games with Applications. II*: *Mean Field Games with Common Noise and Master Equations. Probability Theory and Stochastic Modelling* **84**. Springer, Cham. . MR3753660

[18] CASS, T. and LYONS, T. (2015). Evolving communities with individual preferences. *Proc. Lond. Math. Soc.* (3) **110** 83–107. MR3299600 https://doi.org/10.1112/plms/pdu040

[19] CHEVYREV, I., FRIZ, P. K., KOREPANOV, A., MELBOURNE, I. and ZHANG, H. (2019). Multiscale systems, homogenization, and rough paths. In *Probability and Analysis in Interacting Physical Systems*. *Springer Proc. Math. Stat.* **283** 17–48. Springer, Cham. MR3968507 https://doi.org/10.1007/978-3-030-15338-0_2

[20] CHEVYREV, I. and KORMILITZIN, A. (2016). A primer on the signature method in machine learning. Preprint. Available at arXiv:1603.03788.

[21] CLARK, J. M. C. (1978). The design of robust approximations to the stochastic differential equations of nonlinear filtering. In *Communication Systems and Random Process Theory* (*Proc*. 2*nd NATO Advanced Study Inst*., *Darlington*, 1977). *NATO Adv. Study Inst. Ser. E*: *Appl. Sci.* **25** 721–734. Sijthoff & Noordhoff, Alphen aan den Rijn. MR0529130

[22] CLARK, J. M. C. and CRISAN, D. (2005). On a robust version of the integral representation formula of nonlinear filtering. *Probab. Theory Related Fields* **133** 43–56. MR2197136 https://doi.org/10.1007/s00440-004-0412-5

[23] COGHI, M., DEUSCHEL, J.-D., FRIZ, P. K. and MAURELLI, M. (2020). Pathwise McKean–Vlasov theory with additive noise. *Ann. Appl. Probab.* **30** 2355–2392. MR4149531 https://doi.org/10.1214/20-AAP1560

[24] COGHI, M. and FLANDOLI, F. (2016). Propagation of chaos for interacting particles subject to environmental noise. *Ann. Appl. Probab.* **26** 1407–1442. MR3513594 https://doi.org/10.1214/15-AAP1120

[25] COGHI, M. and GESS, B. (2019). Stochastic nonlinear Fokker–Planck equations. *Nonlinear Anal.* **187** 259–278. MR3954095 https://doi.org/10.1016/j.na.2019.05.003

[26] COGHI, M. and NILSSEN, T. (2021). Rough nonlocal diffusions. *Stochastic Process. Appl.* **141** 1–56. MR4293767 https://doi.org/10.1016/j.spa.2021.07.002

[27] COTTER, C. J. and PAVLIOTIS, G. A. (2009). Estimating eddy diffusivities from noisy Lagrangian observations. *Commun. Math. Sci.* **7** 805–838. MR2604621

[28] COUTIN, L. and LEJAY, A. (2005). Semi-martingales and rough paths theory. *Electron. J. Probab.* **10** 761–785. MR2164030 https://doi.org/10.1214/EJP.v10-162

[29] CRISAN, D., DIEHL, J., FRIZ, P. K. and OBERHAUSER, H. (2013). Robust filtering: Correlated noise and multidimensional observation. *Ann. Appl. Probab.* **23** 2139–2160. MR3134732 https://doi.org/10.1214/12-AAP896

[30] CRISAN, D., LOBBE, A. and ORTIZ-LATORRE, S. (2022). Pathwise approximations for the solution of the non-linear filtering problem. In *Stochastic Analysis*, *Filtering*, *and Stochastic Optimization* 79–99. Springer, Cham. MR4433810 https://doi.org/10.1007/978-3-030-98519-6_4

[31] CRISAN, D. and XIONG, J. (2010). Approximate McKean–Vlasov representations for a class of SPDEs. *Stochastics* **82** 53–68. MR2677539 https://doi.org/10.1080/17442500902723575

[32] DAVIE, A. M. (2008). Differential equations driven by rough paths: An approach via discrete approximation. *Appl. Math. Res. Express. AMRX* **2008** Art. ID abm009, 40 pp. MR2387018

[33] DAVIS, M. H. A. (1982). A pathwise solution of the equations of nonlinear filtering. *Theory Probab. Appl.* **27** 167–175.

[34] DAVIS, M. H. A. (2011). Pathwise nonlinear filtering with correlated noise. In *The Oxford Handbook of Nonlinear Filtering* 403–424. Oxford Univ. Press, Oxford. MR2884603

[35] DAVIS, M. H. A. and SPATHOPOULOS, M. P. (1987). Pathwise nonlinear filtering for nondegenerate diffusions with noise correlation. *SIAM J. Control Optim.* **25** 260–278. MR0877062 https://doi.org/10.1137/0325016

[36] DE WILJES, J., REICH, S. and STANNAT, W. (2018). Long-time stability and accuracy of the ensemble Kalman–Bucy filter for fully observed processes and small measurement noise. *SIAM J. Appl. Dyn. Syst.* **17** 1152–1181. MR3787772 https://doi.org/10.1137/17M1119056

[37] DE WILJES, J. and TONG, X. T. (2020). Analysis of a localised nonlinear ensemble Kalman Bucy filter with complete and accurate observations. *Nonlinearity* **33** 4752–4782. MR4135094 https://doi.org/10.1088/1361-6544/ab8d14

[38] DEL MORAL, P. (2004). Feynman–Kac formulae. In *Feynman–Kac Formulae* 47–93. Springer, New York.

[39] DEYA, A., NEUENKIRCH, A. and TINDEL, S. (2012). A Milstein-type scheme without Lévy area terms for SDEs driven by fractional Brownian motion. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 518–550. MR2954265 https://doi.org/10.1214/10-AIHP392

[40] DIEHL, J., FRIZ, P. and MAI, H. (2016). Pathwise stability of likelihood estimators for diffusions via rough paths. *Ann. Appl. Probab.* **26** 2169–2192. MR3543893 https://doi.org/10.1214/15-AAP1143

[41] DIEHL, J., FRIZ, P. K. and STANNAT, W. (2017). Stochastic partial differential equations: A rough paths view on weak solutions via Feynman-Kac. *Ann. Fac. Sci. Toulouse Math.* (6) **26** 911–947. MR3746646 https://doi.org/10.5802/afst.1556

[42] DING, Z. and LI, Q. (2021). Ensemble Kalman inversion: Mean-field limit and convergence analysis. *Stat. Comput.* **31** Paper No. 9, 21 pp. MR4199469 https://doi.org/10.1007/s11222-020-09976-0

[43] DING, Z., LI, Q. and LU, J. (2021). Ensemble Kalman inversion for nonlinear problems: Weights, consistency, and variance bounds. *Found. Data Sci.* **3** 371–411.

[44] DOBRUŠIN, R. L. (1979). Vlasov equations. *Funktsional. Anal. i Prilozhen.* **13** 48–58, 96. MR0541637

[45] DOUCET, A. and JOHANSEN, A. M. (2011). A tutorial on particle filtering and smoothing: Fifteen years later. In *The Oxford Handbook of Nonlinear Filtering* 656–704. Oxford Univ. Press, Oxford. MR2884612

[46] DUNCAN, A., NÜSKEN, N. and SZPRUCH, L. (2023). On the geometry of Stein variational gradient descent. *J. Mach. Learn. Res.* **24** Paper No. 56, 39 pp. MR4582478

[47] EVENSEN, G. (2003). The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.* **53** 343–367.

[48] EVENSEN, G. (2009). *Data Assimilation*: *The Ensemble Kalman Filter*, 2nd ed. Springer, Berlin. MR2555209 https://doi.org/10.1007/978-3-642-03711-5

[49] FLINT, G., HAMBLY, B. and LYONS, T. (2016). Discretely sampled signals and the rough Hoff process. *Stochastic Process. Appl.* **126** 2593–2614. MR3522294 https://doi.org/10.1016/j.spa.2016.02.011

[50] FOURNIER, N. and GUILLIN, A. (2015). On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields* **162** 707–738. MR3383341 https://doi.org/10.1007/s00440-014-0583-7

[51] FRIZ, P., GASSIAT, P. and LYONS, T. (2015). Physical Brownian motion in a magnetic field as a rough path. *Trans. Amer. Math. Soc.* **367** 7939–7955. MR3391905 https://doi.org/10.1090/S0002-9947-2015-06272-2

[52] FRIZ, P. and RIEDEL, S. (2014). Convergence rates for the full Gaussian rough paths. *Ann. Inst. Henri Poincaré Probab. Stat.* **50** 154–194. MR3161527 https://doi.org/10.1214/12-AIHP507

[53] FRIZ, P. K. and HAIRER, M. (2020). *A Course on Rough Paths*: *With an Introduction to Regularity Structures. Universitext.* Springer, Cham. MR4174393 https://doi.org/10.1007/978-3-030-41556-3

[54] FRIZ, P. K., HOCQUET, A. and LÊ, K. (2021). Rough stochastic differential equations. Preprint. Available at arXiv:2106.10340.

[55] FRIZ, P. K. and VICTOIR, N. B. (2010). *Multidimensional Stochastic Processes as Rough Paths*: *Theory and Applications. Cambridge Studies in Advanced Mathematics* **120**. Cambridge Univ. Press, Cambridge. MR2604669 https://doi.org/10.1017/CBO9780511845079

[56] GAILUS, S. and SPILIOPOULOS, K. (2017). Statistical inference for perturbed multiscale dynamical systems. *Stochastic Process. Appl.* **127** 419–448. MR3583758 https://doi.org/10.1016/j.spa.2016.06.013

[57] GAILUS, S. and SPILIOPOULOS, K. (2018). Discrete-time statistical inference for multiscale diffusions. *Multiscale Model. Simul.* **16** 1824–1858. MR3875799 https://doi.org/10.1137/17M1147408

[58] GIVON, D., KUPFERMAN, R. and STUART, A. (2004). Extracting macroscopic dynamics: Model problems and algorithms. *Nonlinearity* **17** R55–R127. MR2097022 https://doi.org/10.1088/0951-7715/17/6/R01

[59] GOTTWALD, G. A. and REICH, S. (2021). Supervised learning from noisy observations: Combining machine-learning techniques with data assimilation. *Phys. D* **423** Paper No. 132911, 15 pp. MR4249157 https://doi.org/10.1016/j.physd.2021.132911

[60] GYÖNGY, I. (1989). On the approximation of stochastic partial differential equations. II. *Stoch. Stoch. Rep.* **26** 129–164. MR1018542 https://doi.org/10.1080/17442508908833554

[61] HAMMERSLEY, W. R. P., ŠIŠKA, D. and SZPRUCH, Ł. (2021). Weak existence and uniqueness for McKean–Vlasov SDEs with common noise. *Ann. Probab.* **49** 527–555. MR4255126 https://doi.org/10.1214/20-aop1454

[62] HOCQUET, A. and HOFMANOVÁ, M. (2018). An energy method for rough partial differential equations. *J. Differ. Equ.* **265** 1407–1466. MR3797622 https://doi.org/10.1016/j.jde.2018.04.006

[63] HOLLAND, M. and MELBOURNE, I. (2007). Central limit theorems and invariance principles for Lorenz attractors. *J. Lond. Math. Soc.* (2) **76** 345–364. MR2363420 https://doi.org/10.1112/jlms/jdm060

[64] IMKELLER, P., NAMACHCHIVAYA, N. S., PERKOWSKI, N. and YEONG, H. C. (2012). A homogenization approach to multiscale filtering. *Proc. IUTAM* **5** 34–45.

[65] IMKELLER, P., NAMACHCHIVAYA, N. S., PERKOWSKI, N. and YEONG, H. C. (2013). Dimensional reduction in nonlinear filtering: A homogenization approach. *Ann. Appl. Probab.* **23** 2290–2326. MR3127936 https://doi.org/10.1214/12-AAP901

[66] KALLIADASIS, S., KRUMSCHEID, S. and PAVLIOTIS, G. A. (2015). A new framework for extracting coarse-grained models from time series with multiscale structure. *J. Comput. Phys.* **296** 314–328. MR3348504 https://doi.org/10.1016/j.jcp.2015.05.002

[67] KLEIN, R. (2010). Scale-dependent models for atmospheric flows. In *Annual Review of Fluid Mechanics. Vol.* 42. *Annu. Rev. Fluid Mech.* **42** 249–274. Annual Reviews, Palo Alto, CA. MR2647596 https://doi.org/10.1146/annurev-fluid-121108-145537

[68] KLOEDEN, P. E. and PLATEN, E. (1992). Stochastic differential equations. In *Numerical Solution of Stochastic Differential Equations* 103–160. Springer, Berlin.

[69] KRUMSCHEID, S., PAVLIOTIS, G. A. and KALLIADASIS, S. (2013). Semiparametric drift and diffusion estimation for multiscale diffusions. *Multiscale Model. Simul.* **11** 442–473. MR3047437 https://doi.org/10.1137/110854485

[70] KRUMSCHEID, S., PRADAS, M., PAVLIOTIS, G. and KALLIADASIS, S. (2015). Data-driven coarse graining in action: Modeling and prediction of complex systems. *Phys. Rev. B* **92** 042139.

[71] KURTZ, T. G. and XIONG, J. (1999). Particle representations for a class of nonlinear SPDEs. *Stochastic Process. Appl.* **83** 103–126. MR1705602 https://doi.org/10.1016/S0304-4149(99)00024-1

[72] KURTZ, T. G. and XIONG, J. (2001). Numerical solutions for a class of SPDEs with application to filtering. In *Stochastics in Finite and Infinite Dimensions. Trends Math.* 233–258. Birkhäuser, Boston, MA. MR1797090

[73] KUSHNER, H. J. (1979). A robust discrete state approximation to the optimal nonlinear filter for a diffusion. *Stochastics* **3** 75–83. MR0553906

[74] KUTOYANTS, Y. A. (2004). *Statistical Inference for Ergodic Diffusion Processes. Springer Series in Statistics.* Springer, London. MR2144185 https://doi.org/10.1007/978-1-4471-3866-2

[75] LANGE, T. (2022). Derivation of ensemble Kalman–Bucy filters with unbounded nonlinear coefficients. *Nonlinearity* **35** 1061–1092. MR4373995 https://doi.org/10.1088/1361-6544/ac4337

[76] LANGE, T. and STANNAT, W. (2021). On the continuous time limit of the ensemble Kalman filter. *Math. Comp.* **90** 233–265. MR4166460 https://doi.org/10.1090/mcom/3588

[77] LANGE, T. and STANNAT, W. (2021). Mean field limit of ensemble square root filters-discrete and continuous time. *Found. Data Sci.* **3** 563–588.

[78] LEJAY, A. and LYONS, T. (2005). On the importance of the Lévy area for studying the limits of functions of converging stochastic processes. Application to homogenization. In *Current Trends in Potential Theory. Theta Ser. Adv. Math.* **4** 63–84. Theta, Bucharest. MR2243956

[79] LELIÈVRE, T. and STOLTZ, G. (2016). Partial differential equations and stochastic methods in molecular dynamics. *Acta Numer.* **25** 681–880. MR3509213 https://doi.org/10.1017/S0962492916000039

[80] LEVIN, D., LYONS, T. and NI, H. (2013). Learning from the past, predicting the statistics for the future, learning an evolving system. Preprint. Available at arXiv:1309.0260.

[81] LINGALA, N., NAMACHCHIVAYA, N. S., PERKOWSKI, N. and YEONG, H. C. (2012). Particle filtering in high-dimensional chaotic systems. *Chaos* **22** 047509, 18 pp. MR3388722 https://doi.org/10.1063/1.4766595

[82] LINGALA, N., PERKOWSKI, N., YEONG, H., NAMACHCHIVAYA, N. S. and RAPTI, Z. (2014). Optimal nudging in particle filters. *Probab. Eng. Mech.* **37** 160–169.

[83] LIPTSER, R. S. and SHIRYAEV, A. N. (2001). *Statistics of Random Processes. II*: *Applications*, expanded ed. *Applications of Mathematics (New York)* **6**. Springer, Berlin. MR1800858

[84] LORENZ, E. N. (1963). Deterministic nonperiodic flow. *J. Atmos. Sci.* **20** 130–141. MR4021434 https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2

[85] LYONS, T. (2014). Rough paths, signatures and the modelling of functions on streams. In *Proceedings of the International Congress of Mathematicians—Seoul* 2014. *Vol. IV* 163–184. Kyung Moon Sa, Seoul. MR3727607

[86] MCKEAN, H. P. JR. (1966). A class of Markov processes associated with nonlinear parabolic equations. *Proc. Natl. Acad. Sci. USA* **56** 1907–1911. MR0221595 https://doi.org/10.1073/pnas.56.6.1907

[87] NOLEN, J., PAVLIOTIS, G. A. and STUART, A. M. (2012). Multiscale modeling and inverse problems. In *Numerical Analysis of Multiscale Problems. Lecture Notes in Computational Science and Engineering* **83** 1–34. Springer, Heidelberg. MR3050909 https://doi.org/10.1007/978-3-642-22061-6_1

[88] NÜSKEN, N., REICH, S. and ROZDEBA, P. J. (2019). State and parameter estimation from observed signal increments. *Entropy* **21** Paper No. 505, 23 pp. MR3975241 https://doi.org/10.3390/e21050505

[89] NÜSKEN, N. and RENGER, D. (2023). Stein variational gradient descent: Many-particle and long-time asymptotics. *Found. Data Sci.* **5** 286–320. https://doi.org/10.3934/fods.2022023

[90] OLHEDE, S. C., SYKULSKI, A. M. and PAVLIOTIS, G. A. (2009/10). Frequency domain estimation of integrated volatility for Itô processes in the presence of market-microstructure noise. *Multiscale Model. Simul.* **8** 393–427. MR2581027 https://doi.org/10.1137/090756363

[91] PAPAVASILIOU, A., PAVLIOTIS, G. A. and STUART, A. M. (2009). Maximum likelihood drift estimation for multiscale diffusions. *Stochastic Process. Appl.* **119** 3173–3210. MR2568270 https://doi.org/10.1016/j.spa.2009.05.003

[92] PATHIRAJA, S., REICH, S. and STANNAT, W. (2021). McKean–Vlasov SDEs in nonlinear filtering. *SIAM J. Control Optim.* **59** 4188–4215. MR4334536 https://doi.org/10.1137/20M1355197

[93] PATHIRAJA, S. and STANNAT, W. (2021). Analysis of the feedback particle filter with diffusion map based approximation of the gain. *Found. Data Sci.* **3** 615–645.

[94] PAVLIOTIS, G. A. (2014). *Stochastic Processes and Applications*: *Diffusion Processes*, *the Fokker– Planck and Langevin Equations*. *Texts in Applied Mathematics* **60**. Springer, New York. MR3288096 https://doi.org/10.1007/978-1-4939-1323-7

[95] PAVLIOTIS, G. A., POKERN, Y. and STUART, A. M. (2012). Parameter estimation for multiscale diffusions: An overview. In *Statistical Methods for Stochastic Differential Equations. Monogr. Statist. Appl. Probab.* **124** 429–472. CRC Press, Boca Raton, FL. MR2976988 https://doi.org/10.1201/b12126-8

[96] PAVLIOTIS, G. A. and STUART, A. M. (2007). Parameter estimation for multiscale diffusions. *J. Stat. Phys.* **127** 741–781. MR2319851 https://doi.org/10.1007/s10955-007-9300-6

[97] PAVLIOTIS, G. A. and STUART, A. M. (2008). *Multiscale Methods*: *Averaging and Homogenization. Texts in Applied Mathematics* **53**. Springer, New York. MR2382139

[98] PIDSTRIGACH, J. and REICH, S. (2023). Affine-invariant ensemble transform methods for logistic regression. *Found. Comput. Math.* **23** 675–708. MR4568201 https://doi.org/10.1007/s10208-022-09550-2

[99] REICH, S. (2019). Data assimilation: The Schrödinger perspective. *Acta Numer.* **28** 635–711. MR3963510 https://doi.org/10.1017/s0962492919000011

[100] REICH, S. (2023). Frequentist perspective on robust parameter estimation using the ensemble Kalman filter. In *Stochastic Transport in Upper Ocean Dynamics* (B. Chapron, D. Crisan, D. Holm, E. Mémin and A. Radomska, eds.) 237–258. Springer, Cham.

[101] REICH, S. and COTTER, C. (2015). *Probabilistic Forecasting and Bayesian Data Assimilation*. Cambridge Univ. Press, New York. MR3242790 https://doi.org/10.1017/CBO9781107706804

[102] SIMON, D. (2006). *Optimal State Estimation*: *Kalman*, *H Infinity*, *and Nonlinear Approaches*. Wiley, New York.

[103] SPARROW, C. (1982). *The Lorenz Equations*: *Bifurcations*, *Chaos*, *and Strange Attractors. Applied Mathematical Sciences* **41**. Springer, New York–Berlin. MR0681294

[104] SZNITMAN, A.-S. (1991). Topics in propagation of chaos. In *École D'Été de Probabilités de Saint-Flour XIX—1989. Lecture Notes in Math.* **1464** 165–251. Springer, Berlin. MR1108185 https://doi.org/10.1007/BFb0085169

[105] TAGHVAEI, A., DE WILJES, J., MEHTA, P. G. and REICH, S. (2018). Kalman filter and its modern extensions for the continuous-time nonlinear filtering problem. *J. Dyn. Syst. Meas. Control* **140** 030904.

[106] TAGHVAEI, A., MEHTA, P. G. and MEYN, S. P. (2020). Diffusion map-based algorithm for gain function approximation in the feedback particle filter. *SIAM/ASA J. Uncertain. Quantificat.* **8** 1090–1117. MR4134370 https://doi.org/10.1137/19M124513X

[107] TANAKA, H. (1984). Limit theorems for certain diffusion processes with interaction. In *Stochastic Analysis* (*Katata/Kyoto*, 1982). *North-Holland Math. Library* **32** 469–488. North-Holland, Amsterdam. MR0780770 https://doi.org/10.1016/S0924-6509(08)70405-7

[108] WOUTERS, J. and GOTTWALD, G. A. (2019). Stochastic model reduction for slow-fast systems with moderate time scale separation. *Multiscale Model. Simul.* **17** 1172–1188. MR4027831 https://doi.org/10.1137/18M1219965

[109] YANG, T., MEHTA, P. G. and MEYN, S. P. (2013). Feedback particle filter. *IEEE Trans. Automat. Control* **58** 2465–2480. MR3106055 https://doi.org/10.1109/TAC.2013.2258825

[110] YEONG, H. C., BEESON, R., NAMACHCHIVAYA, N. S., PERKOWSKI, N. and SAUER, P. W. (2018). Dynamic data-driven adaptive observations in data assimilation for multi-scale systems. In *Handbook of Dynamic Data Driven Applications Systems* 47–73. Springer, Berlin.

[111] YEONG, H. C., BEESON, R. T., NAMACHCHIVAYA, N. S. and PERKOWSKI, N. (2020). Particle filters with nudging in multiscale chaotic systems: With application to the Lorenz '96 atmospheric model. *J. Nonlinear Sci.* **30** 1519–1552. MR4113335 https://doi.org/10.1007/s00332-020-09616-x

[112] YING, Y., MADDISON, J. and VANNESTE, J. (2019). Bayesian inference of ocean diffusivity from Lagrangian trajectory data. *Ocean Model.* **140** 101401.

[113] YOUNG, L.-S. (2002). What are SRB measures, and which dynamical systems have them? *J. Stat. Phys.* **108** 733–754. MR1933431 https://doi.org/10.1023/A:1019762724717

[114] ZHANG, L., MYKLAND, P. A. and AÏT-SAHALIA, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *J. Amer. Statist. Assoc.* **100** 1394–1411. MR2236450 https://doi.org/10.1198/016214505000000169