

THE CONVERGENCE OF STOCHASTIC DIFFERENTIAL EQUATIONS TO THEIR LINEARISATION IN SMALL NOISE LIMITS*

LIAM BLAKE[†], JOHN MACLEAN[‡], AND SANJEEVA BALASURIYA[‡]

Abstract. Prediction via deterministic continuous-time models will always be subject to model error, for example due to unexplainable phenomena, uncertainties in any data driving the model, or discretisation/resolution issues. In this paper, we build upon previous small-noise studies to provide an explicit bound for the error between a general class of stochastic differential equations and corresponding computable linearisations written in terms of a deterministic system. Our framework accounts for non-autonomous coefficients, multiplicative noise, and uncertain initial conditions. We demonstrate the predictive power of our bound on diverse numerical case studies. We confirm that our bound is *sharp*, in that it accurately predicts the error scaling in the moments of the linearised approximation as both the uncertainty in the initial condition and the magnitude of the noise in the differential equation are altered. This paper also provides an extension of stochastic sensitivity, a recently introduced tool for quantifying uncertainty in dynamical systems, to arbitrary dimensions and establishes the link to our characterisation of stochastic differential equation linearisations.

Keywords. Stochastic differential equations; Gaussian approximations; uncertainty quantification; multiplicative noise

AMS subject classifications. 34F05; 60H10; 60H35

1. Introduction Stochastic differential equations (SDEs) are a natural framework for introducing different sources of uncertainty into the continuous time evolution of a variable [19, 27, 36]. Generally, in modelling situations the dynamics are highly nonlinear and one expects the noise to be multiplicative (i.e. vary with state), e.g. in atmospheric [38] and oceanic [20] systems and from experimental and observational considerations. Such SDEs are usually intractable to solve analytically and computationally expensive to approximate accurately. Data-based models—that is, models possessing terms in the equations which are driven by data rather than by explicitly specified functions—and uncertainty in the initial state render additional problems in obtaining a theoretical understanding of the stochastic system.

A common approach to characterising and approximating the solution of nonlinear SDEs with small noise is via a “linearisation” through time about a single deterministic trajectory. A linearised stochastic differential equation, obtained by truncating Taylor expansions of each coefficient [9, 17, e.g.], can be solved analytically and is accordingly used across a diversity of literature and applications [2, 13, 17, 21, 22, 32, 35, 36].

Much is already known about these linearisations; classical results in the context of small-noise series expansions [9] and large deviations theory [14] show that the strong error between SDE solution with a fixed initial condition and that of an appropriate linearisation is bounded. Sanz-Alonso and Stuart [35] establish a strong result, bounding the Kullback-Leibler divergence between the solutions of autonomous SDEs with additive stationary noise and a linearised equivalent. Their result considers both an uncertain initial condition, and the evolving error due to the discrepancy between the models.

*LB acknowledges support from an Australian Government Research Training Program Scholarship. SB acknowledges with thanks partial support from the Australian Research Council via grant DP200101764.

[†]School of Computer and Mathematical Sciences, University of Adelaide, Adelaide 5005, Australia (corresponding author: liam.blake@adelaide.edu.au).

[‡]School of Computer and Mathematical Sciences, University of Adelaide, Adelaide 5005, Australia.

In [Section 2](#), we directly bound the expectation of all moments of the distance between the linearisation and the true solution of the stochastic differential equation. Our result extends the former convergence result of [\[35\]](#) in three ways: we predict the scaling of all moments of the error, in the presence of multiplicative noise terms, and including time-dependent coefficients in the stochastic differential equations. We directly compare our results with those of [\[35\]](#) in [Section 2.1](#). In [Section 4](#), we show numerically that the scaling of the strong error predicted by [Theorem 2.1](#) matches our predictions on three example SDEs in 1- and 2-dimensions.

The second contribution of the paper is to the notion of “stochastic sensitivity” [\[4\]](#), which seeks to explicitly quantify the impact of vector field uncertainty on the solution trajectories of dynamical systems [\[6, 10, 11, 21\]](#). This methodology was originally developed by Balasuriya [\[4\]](#) as a tool for determining Lagrangian coherent structures (LCS) [\[5, 15\]](#) in fluid flows, in that clusters of trajectories which have small uncertainty may be thought of as more “coherent” than others. In [Section 3](#), we generalise the (formerly two-dimensional) stochastic sensitivity to arbitrary dimensions by proving that the stochastic sensitivity can be computed as the largest eigenvalue of the covariance matrix of the solution to the linearised equation from [Section 2](#). Our expressions extend [\[4\]](#) by empowering a rapid computation of stochastic sensitivity, as previously the computation required three integrals of particular rotations of the gradient of the flow map. We demonstrate this computation in [Section 4.4](#).

2. Convergence of a SDE to a linearisation Suppose we are interested in the evolution of a \mathbb{R}^n -valued state variable y_t over a finite time interval $[0, T]$. Our model, accounting for uncertainties arising from a range of sources, for the evolution of this variable is the Itô stochastic differential equation

$$dy_t^{(\varepsilon)} = u\left(y_t^{(\varepsilon)}, t\right) dt + \varepsilon \sigma\left(y_t^{(\varepsilon)}, t\right) dW_t, \quad y_0^{(\varepsilon)} = x \quad (2.1)$$

where $u: \mathbb{R}^n \times [0, T] \rightarrow \mathbb{R}^n$ is the governing reference vector field. The canonical m -dimensional Wiener process W_t is a continuous white-noise stochastic process with independent Gaussian increments. The scale of the ongoing noise is assumed to be small and is parameterised as $0 < \varepsilon \ll 1$. The noise in [\(2.1\)](#) is multiplicative, in that the diffusion matrix $\sigma: \mathbb{R}^n \times [0, T] \rightarrow \mathbb{R}^{n \times m}$ can vary with state x , as well as with time t . We assume that σ is specified *a priori*, or if no such information is known, then $\sigma \equiv I$, the $n \times m$ identity matrix, is a default modelling choice. We consider [\(2.1\)](#) subject to the *general* uncertain initial condition $y_0^{(\varepsilon)} = x$, where x is an n -dimensional random vector with some given distribution. The two sources of randomness, x and W_t , are assumed independent from each other.

In the absence of any uncertainty (i.e. $\varepsilon = 0$ and the initial condition is a known deterministic quantity), [\(2.1\)](#) reduces to the ordinary differential equation

$$\frac{dy_t^{(0)}}{dt} = u\left(y_t^{(0)}, t\right), \quad y_0^{(0)} = x_0. \quad (2.2)$$

where the initial condition $x_0 \in \mathbb{R}^n$ is fixed. The formal convergence of the stochastic solution $y_t^{(\varepsilon)}$ (under certain conditions on the initial condition) to the deterministic $y_t^{(0)}$ in the limit as $\varepsilon \rightarrow 0$ is well-established using the large deviations principle [\[14, e.g.\]](#). We refer to [\(2.2\)](#) as the *reference* deterministic model associated with [\(2.1\)](#). Solutions to the reference deterministic model are more readily available, e.g. in terms of computational efficiency when solving numerically, than those of the stochastic model, but do not account for inevitable uncertainty.

Let the flow map $F_0^t: \mathbb{R}^n \rightarrow \mathbb{R}^n$ be the function which evolves an initial condition from time 0 to time t according to the flow of (2.2), i.e. $F_0^t(x_0) = y_t^{(0)}$.

We assume certain smoothness and boundedness conditions on the various terms outlined, which are stated explicitly in **Hypothesis H**. Throughout this article, we use the norm symbol $\|\cdot\|$ to denote (i) for a vector, the standard Euclidean vector norm, (ii) for a matrix, the spectral norm induced by the Euclidean norm, and (iii) for a 3rd-order tensor, the spectral norm induced by the matrix norm. The gradient symbol ∇ generically refers to derivatives with respect to the state variable.

HYPOTHESIS H. *Let the deterministic functions $u: \mathbb{R}^n \times [0, T] \rightarrow \mathbb{R}^n$ and $\sigma: \mathbb{R}^n \times [0, T] \rightarrow \mathbb{R}^{n \times m}$, and the random initial condition x be such that:*

- (H.1) *For all $t \in [0, T]$ flow map $F_0^t: \mathbb{R}^n \rightarrow \mathbb{R}^n$ is well-defined, and continuously differentiable (with respect to the initial condition) with invertible derivative.*
- (H.2) *For each $t \in [0, T]$, the function $u(\cdot, t): \mathbb{R}^n \rightarrow \mathbb{R}^n$ given by $u(x, t)$ is twice continuously differentiable on \mathbb{R}^n , and each component of the function $\sigma(\cdot, t): \mathbb{R}^n \rightarrow \mathbb{R}^{n \times m}$ given by $\sigma(x, t)$ is differentiable on \mathbb{R}^n .*
- (H.3) *There exists a constant $K_{\nabla u} \geq 0$ such that for any $t \in [0, T]$ and $x \in \mathbb{R}^n$,*

$$\|\nabla u(x, t)\| \leq K_{\nabla u}.$$

Equivalently, for all $t \in [0, T]$, the function $u(\cdot, t)$ is Lipschitz continuous with Lipschitz constant $K_{\nabla u}$.

- (H.4) *For each $x \in \mathbb{R}^n$, the function $u(x, \cdot): [0, T] \rightarrow \mathbb{R}^n$ and each component of the function $\sigma(x, \cdot): [0, T] \rightarrow \mathbb{R}^{n \times m}$ are Borel-measurable on $[0, T]$.*
- (H.5) *There exists a constant K_L such that for any $t \in [0, T]$ and $x \in \mathbb{R}^n$,*

$$\|u(x, t)\| + \|\sigma(x, t)\| \leq K_L (1 + \|x\|).$$

- (H.6) *There exists a constant $K_{\nabla \sigma} \geq 0$ such that for any $t \in [0, T]$ and $x \in \mathbb{R}^n$,*

$$\|\nabla \sigma(x, t)\| \leq K_{\nabla \sigma},$$

and we take $K_{\nabla \sigma} = 0$ if there is no spatial dependence in σ . Equivalently, for all $t \in [0, T]$, the function $\sigma(x, \cdot)$ is Lipschitz continuous with Lipschitz constant $K_{\nabla \sigma}$.

- (H.7) *The initial condition x is defined on the same probability space as W_t , and is independent of W_t for all $t \in [0, T]$.*
- (H.8) *There exists a constant $K_{\nabla \nabla u} \geq 0$ such that for any $t \in [0, T]$ and $x \in \mathbb{R}^n$,*

$$\|\nabla \nabla u(x, t)\| \leq K_{\nabla \nabla u},$$

and we take $K_{\nabla \nabla} = 0$ if the second spatial derivatives of u are all zero.

- (H.9) *There exists a constant $K_\sigma \geq 0$ such that for any $t \in [0, T]$ and $x \in \mathbb{R}^n$,*

$$\|\sigma(x, t)\| \leq K_\sigma.$$

The conditions H.2 to H.7 guarantee that (2.1) with the initial condition $y_0 = x$ has a unique strong solution [19]. The bound $K_{\nabla \nabla u}$ placed on the second derivatives of u in H.8 quantifies exactly when the deterministic dynamics (that is, u) of (2.1) are linear. Similarly, the bound $K_{\nabla \sigma}$ on the spatial derivatives of σ in H.6 allows us to distinguish when the noise in (2.1) is multiplicative.

Our aim is to construct and formally justify a computable linearisation of (2.1) about a trajectory solving the deterministic system (2.2). To that end, we take a

fixed initial condition $x_0 \in \mathbb{R}^n$ to the reference deterministic model (2.2) and consider linearising the SDE (2.1) about the corresponding trajectory $F_0^t(x_0)$. We consider the following linearisation of (2.1):

$$dl_t^{(\varepsilon)} = \left[u(F_0^t(x_0), t) + \nabla u(F_0^t(x_0), t) \left(l_t^{(\varepsilon)} - F_0^t(x_0) \right) \right] dt + \varepsilon \sigma(F_0^t(x_0), t) dW_t, \quad l_0^{(\varepsilon)} = x, \quad (2.3)$$

where the initial condition x is still permitted to be random. Informally, we can arrive at (2.3) by performing a Taylor expansion of the coefficient u up to first-order and σ to zeroth-order about the time-varying trajectory $F_0^t(x_0)$. Such a linearisation is advantageous over the nonlinear SDE (2.1), since (2.3) can be solved analytically. We will later (see Corollary 2.1) provide explicit expressions for computing the distribution of the solution $l_t^{(\varepsilon)}$ solely in terms of the solution dynamics of the deterministic system (2.2), the specified diffusion matrix σ , and the distribution of x .

In order to quantify the error arising from the choice of reference point x_0 , we define

$$\delta_r := \mathbb{E}[\|x - x_0\|^r]^{1/r},$$

i.e. δ_r is the L_r distance between x and the deterministic point x_0 . We can think of δ_r as a scalar measure of the uncertainty in the initial condition, relative to the choice of reference point x_0 . Alternatively, the limit as δ_r approaches zero is equivalent to convergence in r th mean of x to the fixed point x_0 . We can therefore distinguish two sources of uncertainty in our model; that arising from the initial condition, quantified by δ_r , and the ongoing uncertainty driven by the Wiener process W_t as measured by ε .

Our first and primary result, Theorem 2.1, provides an explicit bound on the r th moment of the error between the SDE solution $y_t^{(\varepsilon)}$ and the linearised solution $l_t^{(\varepsilon)}$.

THEOREM 2.1 (Linearisation error is bounded). *Let $y_t^{(\varepsilon)}$ be the strong solution to the SDE (2.1) and $l_t^{(\varepsilon)}$ be the strong solution to the corresponding linearisation (2.3), both driven by the same Wiener process W_t and subject to the same random initial condition $y_0^{(\varepsilon)} = l_0^{(\varepsilon)} = x$. Then, for any $r \geq 1$ such that $\delta_{2r} < \infty$ and $t \in [0, T]$, there exist constants $D_1(r, t, K_{\nabla u}, K_\sigma)$, $D_2(r, t, K_{\nabla u})$, $D_3(r, t, K_{\nabla u}) \in [0, \infty)$ independent of x and x_0 such that for all $\varepsilon > 0$,*

$$\mathbb{E} \left[\left\| y_t^{(\varepsilon)} - l_t^{(\varepsilon)} \right\|^r \right] \leq (K_{\nabla \nabla u}^r + K_{\nabla \sigma}^r) D_1(r, t, K_{\nabla u}, K_\sigma) \varepsilon^{2r} + K_{\nabla \nabla u}^r D_2(r, t, K_{\nabla u}) \delta_{2r}^{2r} + K_{\nabla \sigma}^r D_3(r, t, K_{\nabla u}) \delta_r^r \varepsilon^r. \quad (2.4)$$

Proof. See Appendix B. Our proof employs the Burkholder-Davis-Gundy inequality, Grönwall's inequality, and Taylor's theorem to explicitly construct the bounding coefficients in terms of the conditions on the SDE coefficients set out in Hypothesis H. The bounding coefficients D_1 , D_2 , and D_3 are given explicitly in (B.11). \square

In (2.4), we have an explicit scaling of the error in terms of ε and δ_r . The three terms can be informally considered as: a contribution purely from the ongoing linearisation error, a contribution purely from the initial uncertainty, and a term resulting from the interaction between the initial and ongoing uncertainties. By explicitly identifying the dependence of the bound on $K_{\nabla \nabla u}$ and $K_{\nabla \sigma}$, we note three special cases that are summarised by Remarks 2.1 to 2.3.

REMARK 2.1 (Linear drift). *When the deterministic dynamics are linear, we set $K_{\nabla \nabla u} = 0$ and (2.4) becomes*

$$\mathbb{E} \left[\left\| y_t^{(\varepsilon)} - l_t^{(\varepsilon)} \right\|^r \right] \leq K_{\nabla \sigma}^r D_1(r, t, K_{\nabla u}, K_\sigma) \varepsilon^{2r} + K_{\nabla \sigma}^r D_3(r, t, K_{\nabla u}) \delta_r^r \varepsilon^r.$$

The linearisation of the drift term u is exact, so the error is purely due to the spatial dependency of the diffusion term σ .

REMARK 2.2 (Additive noise). When the noise in (2.1) is additive (non-multiplicative), we set $K_{\nabla\sigma} = 0$ and (2.4) becomes

$$\mathbb{E} \left[\left\| y_t^{(\varepsilon)} - l_t^{(\varepsilon)} \right\|^r \right] \leq K_{\nabla\nabla u}^r D_1(r, t, K_{\nabla u}, K_{\sigma}) \varepsilon^{2r} + K_{\nabla\nabla u}^r D_2(r, t, K_{\nabla u}) \delta_{2r}^{2r}.$$

The error is then purely due to the linearisation of the drift term u , and as expected is of second order in both the initial condition uncertainty δ_{2r} and the ongoing uncertainty ε .

REMARK 2.3 (Exact linearisation). When the deterministic dynamics are linear and the noise in (2.1) is additive (non-multiplicative), the linearisation (2.3) should be exact. Accordingly, we set $K_{\nabla\nabla u} = K_{\nabla\sigma} = 0$ and (2.4) becomes

$$\mathbb{E} \left[\left\| y_t^{(\varepsilon)} - l_t^{(\varepsilon)} \right\|^r \right] = 0.$$

In turn, this implies that $y_t^{(\varepsilon)} = l_t^{(\varepsilon)}$ almost surely, for any choice of reference point x_0 .

We postpone a discussion of an additional special case—where the initial condition is fixed or Gaussian—to a later section. For the general situation, we next explicitly establish the solution to the linearisation (2.3), in terms of the initial condition and the deterministic evolution of (2.2).

THEOREM 2.2 (Solution of the linearised SDE). The strong solution to the linearised SDE (2.3) is

$$l_t^{(\varepsilon)} = \nabla F_0^t(x_0)(x - x_0) + F_0^t(x_0) + \varepsilon \nabla F_0^t(x_0) \int_0^t L(x_0, \tau) dW_{\tau}. \quad (2.5)$$

where the term involving the uncertain initial condition x and the Itô integral are independent, and

$$L(x_0, \tau) := [\nabla F_0^{\tau}(x_0)]^{-1} \sigma(F_0^{\tau}(x_0), \tau). \quad (2.6)$$

Proof. See Appendix C. \square

The representation of the linearised solution as an independent sum in (2.5) can be seen as a decomposition into contributions from the initial uncertainty (the transformation of initial condition x), a deterministic prediction (the flow map $F_0^t(x_0)$) and the ongoing uncertainty in u (the remaining Itô integral term).

We can further show that the Itô integral term follows a Gaussian random variable, which ensures that the independent sum in (2.5) is a convenient expression for both theoretical analysis and numerical computation. We also provide explicit expressions for the mean and covariance matrix of the linearised solution, written in terms of the deterministic dynamics and σ .

COROLLARY 2.1 (Distribution of the linearised solution). The Itô integral term in (2.5) follows a Gaussian distribution independent of x , namely

$$\int_0^t L(x_0, \tau) dW_{\tau} \sim \mathcal{N} \left(0, \int_0^t L(x_0, \tau) L(x_0, \tau)^{\top} d\tau \right).$$

The mean of the linearised solution is

$$\mathbb{E} \left[l_t^{(\varepsilon)} \right] = F_0^t(x_0) + \nabla F_0^t(x_0) \mathbb{E}[x - x_0]. \quad (2.7)$$

The $n \times n$ covariance matrix of the linearised solution is given explicitly by

$$\mathbb{V}\left[l_t^{(\varepsilon)}\right] = \nabla F_0^t(x_0) \left(\mathbb{V}[x] + \varepsilon^2 \int_0^t L(x_0, \tau) L(x_0, \tau)^\top d\tau \right) \left[\nabla F_0^t(x_0) \right]^\top \quad (2.8)$$

where $L(x_0, \tau)$ is as defined in (2.6) and the integral is taken in the elementwise sense.

Proof. See [Appendix D](#). The expressions follow from the representation of the linearised solution as an independent sum in (2.5). \square

In [Theorem 2.2](#), we have provided expressions for the distribution of the solution $l_t^{(\varepsilon)}$ to the linearised SDE (2.3) written solely in terms of the behaviour of the deterministic system (2.2), the specified diffusion matrix σ , and the distribution of the initial condition x . This describes a method for approximating the solution to the nonlinear SDE (2.1), or for characterising the impact of uncertainty in a dynamical system (2.2), that circumvents the need for expensive stochastic simulation.

Thus far, we have stated our results in terms of a general initial condition x , and provided expressions for the linearised solution in terms of this otherwise arbitrary distribution. However, we later consider two special cases for the initial condition x , a fixed (deterministic) initial condition in [Section 2.3](#), and a Gaussian initial condition in [Section 2.2](#). In both these cases, the linearised solution also follows a Gaussian distribution which is characterised entirely by the mean and covariance described in [Corollary 2.1](#), allowing for easy computation. We also relate these results directly to other literature [[4](#), [9](#), [14](#), [17](#), [35](#), [36](#)] which uses linearisation procedures and Gaussian process approximations for nonlinear SDEs in these situations.

Next, we establish the ordinary differential equation satisfied by the covariance matrix, which is an expression consistent with linearisations schemes described elsewhere [[2](#), [17](#), [35](#), [36](#)]. This ODE enables rapid computation of the mean and covariance of the linearised solutions by solving a system of ODEs, i.e. (2.2) and (2.9).

REMARK 2.4. The $n \times n$ covariance matrix $\mathbb{V}\left[l_t^{(\varepsilon)}\right]$ of the linearised solution is the symmetric positive-semidefinite $n \times n$ matrix solution to the ordinary differential equation

$$\frac{d\Pi(t)}{dt} = \nabla u(F_0^t(x_0), t) \Pi(t) + \Pi(t) \left[\nabla u(F_0^t(x_0), t) \right]^\top + \varepsilon^2 \sigma(F_0^t(x_0), t) \sigma(F_0^t(x_0), t)^\top, \quad (2.9)$$

subject to the initial condition $\Pi(0) = \mathbb{V}[x]$. We show that the variance satisfies (2.9) in [Appendix D](#).

2.1. Comparison to existing results In this section we connect our work to the cognate bound derived by Sanz-Alonso and Stuart [[35](#)]. That paper considers the following SDE:

$$dy_t^{(\varepsilon)} = u\left(y_t^{(\varepsilon)}\right) dt + \varepsilon \tilde{\sigma} dW_t, \quad (2.10)$$

where the diffusion coefficient $\tilde{\sigma}$ is a constant matrix, which is a special case of (2.1). In this section, we apply our results to (2.10) to enable both bounds to be compared. Note that ε in our article is written as $\sqrt{\varepsilon}$ in [[35](#)]; we will translate results from [[35](#)] to use our notation, so that all results in this article are directly comparable. In the following, c denotes an arbitrary finite and non-negative constant that can vary between inequalities.

Theorem 2.2 of [[35](#)], summarised, is as follows. Let $\xi_t^{(\varepsilon)}$ be the probability measure associated with $y_t^{(\varepsilon)}$ (as defined in (2.10)), and let $\nu_t^{(\varepsilon)}$ be the probability measure

associated with the corresponding linearisation $l_t^{(\varepsilon)}$ (as defined in [Section 2](#)). Then there exists a constant c such that the Kullback–Leibler (KL) divergence D_{KL} between $\xi_t^{(\varepsilon)}$ and $\nu_t^{(\varepsilon)}$ is bounded;

$$D_{\text{KL}}\left(\xi_t^{(\varepsilon)} \parallel \nu_t^{(\varepsilon)}\right) \leq D_{\text{KL}}\left(\xi_0^{(\varepsilon)} \parallel \nu_0^{(\varepsilon)}\right) + c\varepsilon^2. \quad (2.11)$$

To focus on the scaling with ε , assume a fixed initial condition with $D_{\text{KL}}\left(\xi_0^{(\varepsilon)} \parallel \nu_0^{(\varepsilon)}\right) = 0$ (and $\delta_r = 0$ in our bound [\(2.4\)](#)). Then, employing the Hellinger distance D_{H} , [\(2.11\)](#) implies

$$\left\| \mathbb{E}\left[y_t^{(\varepsilon)} - l_t^{(\varepsilon)}\right] \right\| \leq cD_{\text{H}}\left(\xi_t^{(\varepsilon)}, \nu_t^{(\varepsilon)}\right) \leq c\sqrt{D_{\text{KL}}\left(\xi_t^{(\varepsilon)} \parallel \nu_t^{(\varepsilon)}\right)} \leq c\varepsilon, \quad (2.12)$$

while our result [\(2.4\)](#) and Jensen’s inequality imply

$$\left\| \mathbb{E}\left[y_t^{(\varepsilon)} - l_t^{(\varepsilon)}\right] \right\| \leq \mathbb{E}\left[\left\|y_t^{(\varepsilon)} - l_t^{(\varepsilon)}\right\|\right] \leq c\varepsilon^2.$$

Thus, our bound on the moments is quadratic in ε rather than linear. If our conversion in [\(2.12\)](#) was optimal, then our approach in this article provides a sharper bound on $\left\| \mathbb{E}\left[y_t^{(\varepsilon)} - l_t^{(\varepsilon)}\right] \right\|$ that the results of [\[35\]](#) imply, and do so for a more general σ . The results in [\[35\]](#) on the KL divergence would be more natural in information-theoretic contexts, and our hope is that our explicit bound on the moments would be similarly preferred in other contexts.

2.2. Gaussian initial condition We now briefly consider the case when the initial condition follows a Gaussian distribution, i.e. $x \sim \mathcal{N}(\mu_0, \Sigma_0)$, where $\mu_0 \in \mathbb{R}^n$ and $\Sigma_0 \in \mathbb{R}^{n \times n}$ are fixed and specified. The linearisation then follows a Gaussian distribution itself, which is entirely characterised by the mean and covariance matrix described in [Corollary 2.1](#). Alternatively, these moments can be conveniently computed by simultaneously solving [\(2.2\)](#) for the state variable and [\(2.9\)](#) for the linearised covariance.

A natural choice of reference point x_0 is the mean of the initial Gaussian density, i.e. $x_0 = \mu_0$. The L_r distance between x and the mean μ_0 can be bounded by the trace of Σ_0 ; for example, one such bound is

$$\delta_r^r \leq n^{3r/2-1} M_r \text{tr}(\Sigma_0)^{r/2}, \quad M_r := \frac{2^{r/2} \Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi}}, \quad (2.13)$$

where Γ denotes the Gamma function, with equality when $n = 1$. The initial covariance Σ_0 directly measures the uncertainty in the initial condition, and we see through [\(2.13\)](#) that as the components of Σ_0 approach zero, the contribution of the initial uncertainty to the linearisation error in [\(2.4\)](#) approaches zero also. The linearised solution is then

$$l_t^{(\varepsilon)} \sim \mathcal{N}\left(F_0^t(x_0), \nabla F_0^t(x_0) \Sigma_0 [\nabla F_0^t(x_0)]^\top + \varepsilon^2 \Sigma_0^t(x_0)\right),$$

where $\Sigma_0^t(x_0)$ is given explicitly by

$$\Sigma_0^t(x_0) = \nabla F_0^t(x_0) \left(\int_0^t L(x_0, \tau) L(x_0, \tau)^\top d\tau \right) [\nabla F_0^t(x_0)]^\top, \quad (2.14)$$

and is the solution to the matrix differential equation (2.9) in Remark 2.4, subject to $\Sigma_0^0(x_0) = O$, the $n \times n$ zero matrix. The covariance matrix $\Sigma_0^t(x_0)$ characterises the contribution of the ongoing uncertainty in the stochastic system. The full covariance matrix $\mathbb{V}[l_t^{(\varepsilon)}]$ is also the solution to (2.9) subject to the initial condition $\Pi(0) = \mathbb{V}[x]$.

By jointly solving (2.2) for the deterministic trajectory (the mean of $l_t^{(\varepsilon)}$) and (2.9) for the covariance matrix, one can easily compute the linearised solution, describing exactly the assumed Gaussian approximation presented in Särkkä and Solin [36], and the dynamics linearisation used in the extended Kalman filter [17].

2.3. Fixed initial condition Consider when the initial condition x is itself a fixed and known deterministic value, in which case we take $x = x_0$ and $\delta_r = 0$ for all r . In this situation, the bound (2.4) on the linearisation error reduces to

$$\mathbb{E} \left[\left\| y_t^{(\varepsilon)} - l_t^{(\varepsilon)} \right\|^r \right] \leq (K_{\nabla}^r u + K_{\nabla}^r \sigma) D_1(r, t, K_{\nabla} u, K_{\sigma}) \varepsilon^{2r}. \quad (2.15)$$

We can consider the linearisation as equivalently arising from a first-order power series expansion of $y_t^{(\varepsilon)}$ in the noise-scale parameter ε , i.e.

$$y_t^{(\varepsilon)} = F_0^t(x_0) + \varepsilon z_t^{(\varepsilon)} + R_2(x, t, \varepsilon).$$

where $z_\varepsilon := (l_t^{(\varepsilon)} - F_0^t(x_0)) / \varepsilon$ is the first order term and R_2 is a random quantity capturing the remaining deviation between $y_t^{(\varepsilon)}$ and the linearisation. By rearranging and taking $r = 1$ in (2.15), we therefore have the explicit Taylor-like bound

$$\frac{\mathbb{E} \left[\|R_2(x, t, \varepsilon)\| \right]}{\varepsilon^2} \leq (K_{\nabla} u + K_{\nabla} \sigma) D_1(1, t, K_{\nabla} u, K_{\sigma}),$$

This result is consistent with the formulation of the linearisation error bounds by Blagoveshchenskii [9] and Freidlin and Wentzell [14], for instance. Moreover, the distribution of the linearisation solution (2.5) is Gaussian, which through Corollary 2.1 we can again explicitly characterise in terms of the deterministic system, namely

$$l_t^{(\varepsilon)} \sim \mathcal{N} \left(F_0^t(x_0), \varepsilon^2 \Sigma_0^t(x_0) \right), \quad (2.16)$$

where $\Sigma_0^t(x_0)$ is defined in (2.14). The distribution can be computed *entirely* from the solution behaviour of the deterministic equation (2.2) and prior specification of σ . In Section 3, we demonstrate an application of these results to extend stochastic sensitivity [4] to arbitrary dimension.

3. Extending stochastic sensitivity The results of Section 2.3 for a fixed initial condition provide a direct extension of the stochastic sensitivity tools first introduced by Balasuriya [4] for the fluid flow context. Here, the deterministic model (2.2) is seen as a “best-available” model for the evolution of Lagrangian trajectories, and the driving vector field u is the Eulerian velocity of the fluid. Stochastic sensitivity ascribes a scalar value to each deterministic trajectory by computing a maximum variance of projected deviation [4]. The aim is to provide a *single* computable number for each deterministic trajectory quantifying the impact of uncertainty in the velocity, independent of the scale (ε) of the noise. The natural restating of this original definition of stochastic sensitivity [4] in the n -dimensional setting is as follows:

DEFINITION 3.1 (Stochastic sensitivity in \mathbb{R}^n). *The stochastic sensitivity is a scalar field $S^2: \mathbb{R}^n \times [0, T] \rightarrow [0, \infty)$ given by*

$$S^2(x_0, t) := \limsup_{\varepsilon \downarrow 0} \left\{ \frac{1}{\varepsilon} \mathbb{V} \left[p^\top \left(y_t^{(\varepsilon)} - F_0^t(x_0) \right) \right] : p \in \mathbb{R}^n, \|p\| = 1 \right\}.$$

Definition 3.1 is in the spirit of principal components analysis [18], performing a dimension reduction by projecting onto the direction in which the variance is maximised, thus capturing the most uncertainty in the data with a scalar value. The anisotropic uncertainty in two-dimensions [4] is the direction-dependent projection (prior to optimising over all directions in Definition 3.1). Explicit theoretical expressions for both the stochastic sensitivity and the anisotropic sensitivity in two dimensions were obtained by Balasuriya [4]; these allowed for quantifying certainty in eventual trajectory locations without having to perform stochastic simulations. We show here that our results in n -dimensions are a generalisation of the two-dimensional ones in [4], which moreover establish Gaussianity as well as an explicit expression for the uncertainty measure. A theoretically pleasing and computable expression for the stochastic sensitivity is obtainable;

THEOREM 3.1 (Computation of S^2). *For any $x_0 \in \mathbb{R}^n$ and $t \in [0, T]$,*

$$S^2(x_0, t) = \|\Sigma_0^t(x_0)\|, \quad (3.1)$$

where the covariance matrix Σ_0^t is defined in (2.14). Equivalently, $S^2(x_0, t)$ is given by the maximum eigenvalue of $\Sigma_0^t(x_0)$.

Proof. See Appendix E. This result uses Theorem 2.1 to establish the convergence of the covariance matrices, and then the properties of the spectral norm to establish (3.1). \square

Independent of the fluid mechanics context, Theorem 3.1 indicates that even for general systems, the matrix norm of $\Sigma_0^t(x)$, i.e., the stochastic sensitivity $S^2(x, t)$, can be used as *one* number which encapsulates the uncertainty of an initial state x after t time units. The significance of this result is that the stochastic sensitivity has here been recovered as the maximal eigenvalue of a covariance matrix that is ubiquitous in the literature. Stochastic sensitivity was formerly defined in [4] as the maximal value of the anisotropic uncertainty of a particular stochastic flow, and the connection to a linearisation was not apparent.

The stochastic sensitivity field can be calculated given any velocity data u , and through the explicit expression (2.14) for Σ_0^t can even be computed from only flow map data. Computation does not require knowledge of the noise scale ε , so the S^2 field is intrinsic in capturing the impact of the model dynamics on uncertainty, and any specified non-uniform diffusivity. It has already been shown that, in the fluid flow context, stochastic sensitivity can identify coherent regions in two-dimensions [3, 4]. A simple approach is to define robust sets, which are those initial conditions for which the corresponding S^2 value, i.e., the uncertainty in eventual location, are below some specified threshold. This threshold can be defined precisely in terms of a spatial lengthscale of interest and the advective and diffusive characteristics of the flow, as Definition 2.9 of [4]. Such a definition extends to the n -dimensional case as presented here, moreover establishing an easily computable method for determining coherent sets from the covariance matrix Σ_0^t .

4. Numerical validation & examples This section will validate the theory presented in Sections 2 and 3, by considering three example SDEs each leading to a

different form of the strong error bound (2.4). For each example, we first demonstrate heuristically that the solution converges to the limiting distribution described by Theorem 2.2, and then verify the error bound in Theorem 2.1 directly by considering a range of values for the noise scale ε and initial condition uncertainty δ_r . We demonstrate numerically the form of the bound on the linearisation error predicted by Theorem 2.1 is sharp, in the sense that estimates of the error scale exactly with the initial uncertainty δ_r and ongoing uncertainty ε as predicted.

All simulations in this section were generated using the Julia programming language [8], with the implementations of numerical ODE and SDE solvers provided by the DifferentialEquations.jl package [31]. The code is available at github.com/liamblake/explicit-characterisation-sde-linearisation.

4.1. Nonlinear dynamics, additive noise Consider the following SDE in 1D;

$$dy_t^{(\varepsilon)} = \sin(y_t^{(\varepsilon)}) dt + \varepsilon dW_t. \quad (4.1)$$

The deterministic system corresponding to (4.1) has solution

$$F_0^t(x_0) = 2 \arctan \left(e^{-t} \tan \left(\frac{x_0}{2} \right) \right).$$

Further details of this example, including computation of the derivatives required in the linearisation, are provided in the supplementary material.

To explore the impact of initial condition uncertainty, we consider the univariate Gaussian initial condition $y_0 = x \sim \mathcal{N}(\mu, \rho^2)$, where the mean μ is specified and the standard deviation ρ is a non-negative scaling parameter. We linearise (4.1) about the deterministic trajectory $F_0^t(\mu)$ originating from the mean, that is, μ is the chosen reference point. This ensures that for any $r \geq 0$

$$\delta_r^r = \mathbb{E}[|x - \mu|^r] = M_r \rho^r. \quad (4.2)$$

where M_r is as defined in (2.13). This property of the univariate Gaussian distribution allows us to easily control the uncertainty in the initial condition and verify the bounds; by sending the parameter ρ to zero we ensure that δ_r approaches zero also.

The linearised equation is then

$$dl_t^{(\varepsilon)} = \left[F_0^t(\mu) + \cos(F_0^t(\mu)) \left(l_t^{(\varepsilon)} - F_0^t(\mu) \right) \right] dt + \varepsilon dW_t, \quad l_0^{(\varepsilon)} \sim \mathcal{N}(\mu, \rho^2). \quad (4.3)$$

and the solution follows a Gaussian distribution, specifically

$$l_t^{(\varepsilon)} \sim \mathcal{N} \left(F_0^t(\mu), \rho^2 \nabla F_0^t(\mu)^2 + \varepsilon^2 \Sigma_0^t(\mu) \right). \quad (4.4)$$

where $\Sigma_0^t(\mu)$ is computed by solving (2.9) numerically subject to a zero initial condition.

In this example, we take $\mu_0 = 0.5$ and consider the solutions of (4.1) and (4.3) at time $t = 1.5$. We generate accurate samples of (4.1) and (4.3) jointly (i.e. using the same numerical realisations of the Wiener process W_t) using the stochastic Runge-Kutta scheme SRI [33] with an adaptive step size [30].

In Figure 4.1, we show histograms of $N = 10000$ samples of the solution to non-linear SDE (4.1) and the corresponding probability density function of the linearised solution (4.4), for different combinations of ε and ρ . Even when the ongoing noise

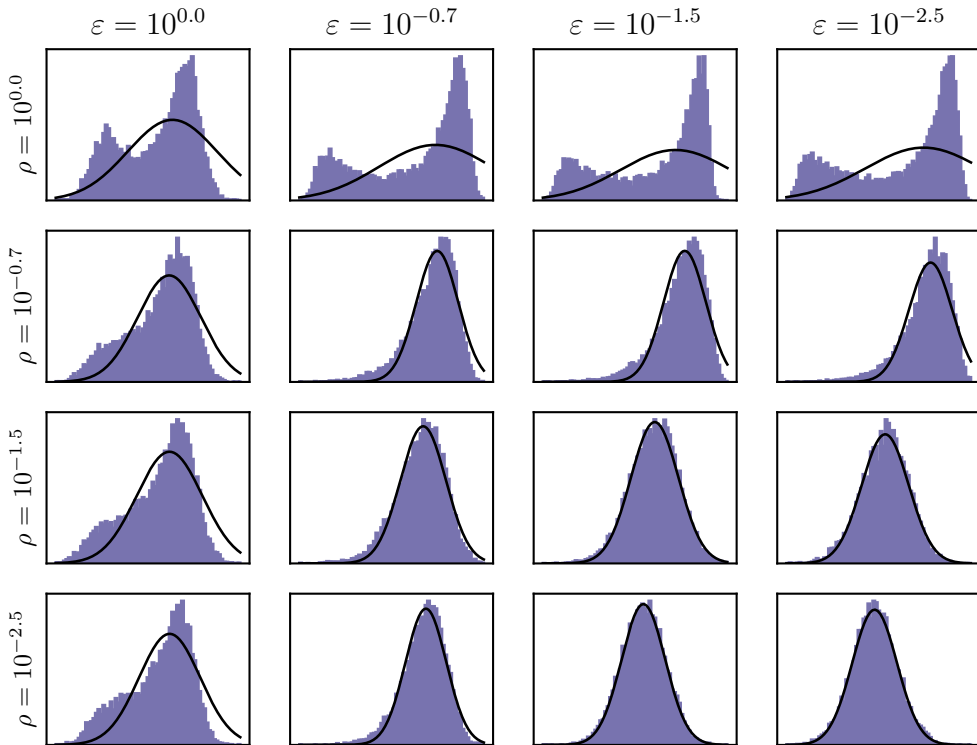


Fig. 4.1: Histograms of stochastic samples of (4.1), subject to the Gaussian initial condition (4.2), for varying initial uncertainty scale ρ and ongoing uncertainty scale ε . The distribution of the corresponding solution (4.4) to the linearised equation is overlaid in black.

is small, the nonlinearity of the drift term means that a large initial uncertainty results in a non-Gaussian distribution. However, in situations where both the initial and ongoing uncertainties are small, the Gaussian solution to the linearised equation provides a reasonable approximation. In the limit of both small initial ($\rho \rightarrow 0$) and small ongoing ($\varepsilon \rightarrow 0$) uncertainty (towards the bottom right), we see that the distribution of the samples approach the Gaussian density of the linearisation solution, matching the understanding that the linearisation approximation is “reasonable” for small noise regimes.

Since the drift term is nonlinear and the noise is additive in (4.1), the bound predicted by Theorem 2.1 has the form

$$\mathbb{E} \left[\left\| y_t^{(\varepsilon)} - l_t^{(\varepsilon)} \right\|^r \right] \leq D_1(r, t, K_{\nabla u}, K_{\sigma}) \varepsilon^{2r} + M_{2r} D_2(r, t, K_{\nabla u}) \rho^{2r}.$$

where we have taken $K_{\nabla \nabla u} = 1$ and $K_{\nabla \sigma} = 0$. To numerically validate this bound under the Gaussian initial condition (4.2), define for $r \geq 1$ the error measure

$$E_r(\varepsilon, \rho) := \frac{1}{N} \sum_{i=1}^N \left\| \hat{y}_i^{(\varepsilon)} - \hat{l}_i^{(\varepsilon)} \right\|^r, \quad (4.5)$$

which is a Monte-Carlo estimator of the right-hand side of (2.4), where $\hat{y}_1^{(\varepsilon)}, \dots, \hat{y}_N^{(\varepsilon)}$ and $\hat{l}_1^{(\varepsilon)}, \dots, \hat{l}_N^{(\varepsilon)}$ are N numerical samples of the solutions to SDE (2.1) and the linearisation (2.3) respectively.

We directly validate the *form* of the error bound (as a function of ε and ρ) in Figure 4.2, by computing E_1 using samples for each pair of ε and ρ values. In Figure 4.2a, we demonstrate the relationship between E_1 and the ongoing uncertainty ε for several different fixed values of ρ , each corresponding to a different colour. A least squares estimate of a line of best fit of the form $E_1 = \beta_0 + \beta_1 \varepsilon^2$, for fixed coefficients β_0 and β_1 , is fitted to the observed errors (in untransformed space) to verify the scaling of our bound in Theorem 2.1. We see that the line of best fit accurately matches the observed values of E_1 , verifying that E_1 is in fact scaling with ε^2 as predicted. Figure 4.2b provides a similar demonstration between E_1 and the initial uncertainty ρ , where now each colour corresponds to a different fixed value of ε . We again fit lines of the form $E_1 = \beta_0 + \beta_1 \rho^2$ to verify the scaling of the bound, and see that the lines match the observed values of E_2 . Thus, we have also validated that E_1 scales with ρ^2 , as expected from Theorem 2.1.

4.2. Linear dynamics, multiplicative noise Now consider the following SDE with multiplicative noise in 1D;

$$dy_t^{(\varepsilon)} = \frac{1}{2} y_t^{(\varepsilon)} dt + \varepsilon \cos\left(y_t^{(\varepsilon)}\right) dW_t. \quad (4.6)$$

The corresponding deterministic system is linear and has solution

$$F_0^t(x_0) = \exp\left(\frac{t}{2}\right) x_0, \quad (4.7)$$

with additional details provided in the supplementary materials. As with the previous example in Section 4.1, we take the Gaussian initial condition (4.2) with variance ρ^2 and linearised (4.6) about the initial mean μ . The linearised equation is then

$$dl_t^{(\varepsilon)} = \frac{1}{2} l_t^{(\varepsilon)} dt + \varepsilon \cos\left(\exp\left(\frac{t}{2}\right) \mu\right) dW_t, \quad l_0^{(\varepsilon)} \sim \mathcal{N}(\mu, \rho^2), \quad (4.8)$$

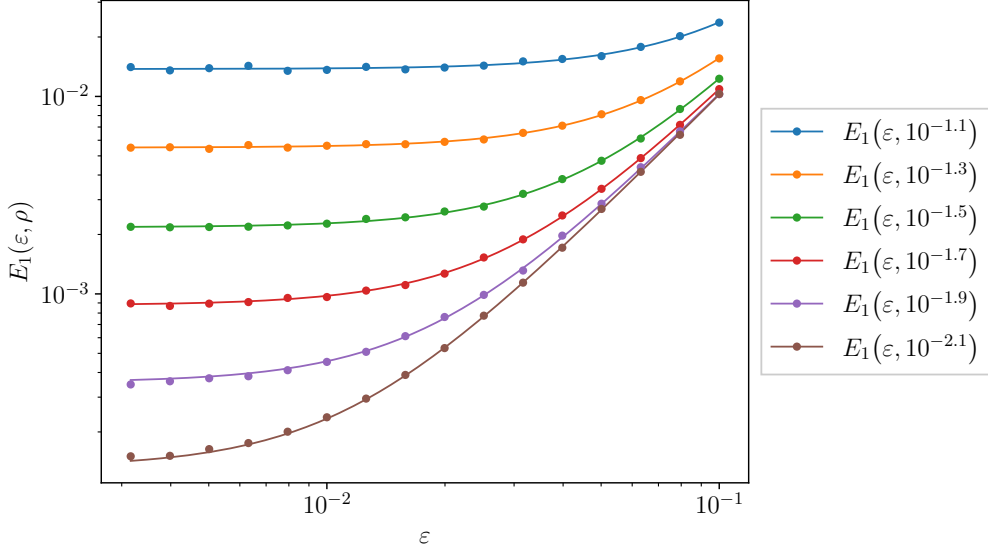
with Gaussian solution (4.4). We take the initial point $\mu = 2$ and consider the solutions at time $t = 1$. To generate numerical realisations of the solutions to (4.6) and (4.8) with the same realisations of W_t , we use the same set-up as in the previous example.

In Figure 4.1, we show histograms of $N = 10000$ samples of the multiplicative noise SDE (4.6) and the corresponding probability density function of the linearised solution, for different combinations of ε and ρ . We again see that in the limit of both small initial and small ongoing uncertainty (towards the bottom right), we see that the distribution of the samples approach the Gaussian density of the linearisation solution.

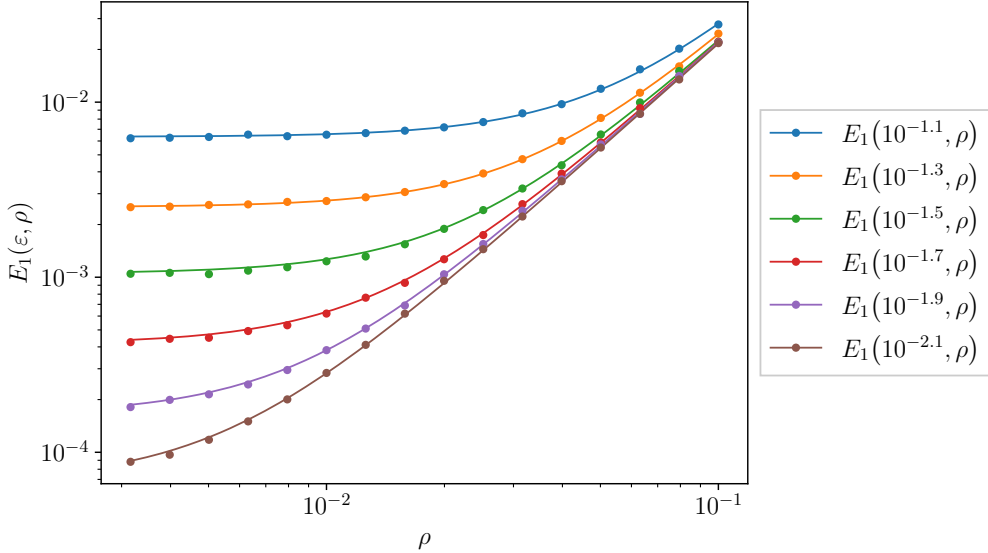
Since the drift term is linear and the noise multiplicative in (4.6), the bound predicted by Theorem 2.1 has the form

$$\mathbb{E} \left[\left\| y_t^{(\varepsilon)} - l_t^{(\varepsilon)} \right\|^r \right] \leq D_1(r, t, K_{\nabla u}, K_{\sigma}) \varepsilon^{2r} + M_r D_3(r, t, K_{\nabla u}) \varepsilon^r \rho^r,$$

where we have $K_{\nabla \nabla u} = 0$ and $K_{\nabla \sigma} = 1$. In Figure 4.4, we again validate the form of this bound (for $r = 1$; results for additional values of r are provided in the supplementary material) by approximating the left-hand side with E_1 computed from realisations of



(a) Estimates of the strong error (with $r = 1$) in linearising (4.1) with (4.3), for varying ongoing uncertainty parameter ε . Each colour corresponds to a different value of the initial uncertainty parameter ρ . A (least squares) line of best fit of the form $\beta_0 + \beta_1 \varepsilon^2$ is included in the corresponding colour.



(b) Estimates of the strong error (with $r = 1$) for varying initial uncertainty parameter ρ . Each colour corresponds to a different value of the ongoing uncertainty parameter ε . A (least squares) line of best fit of the form $\beta_0 + \beta_1 \rho^2$ is included in the corresponding colour.

Fig. 4.2: Validation of the theoretical bound predicted by [Theorem 2.1](#), when $r = 1$, on numerical realisations of the solution to the 1D example (4.1).

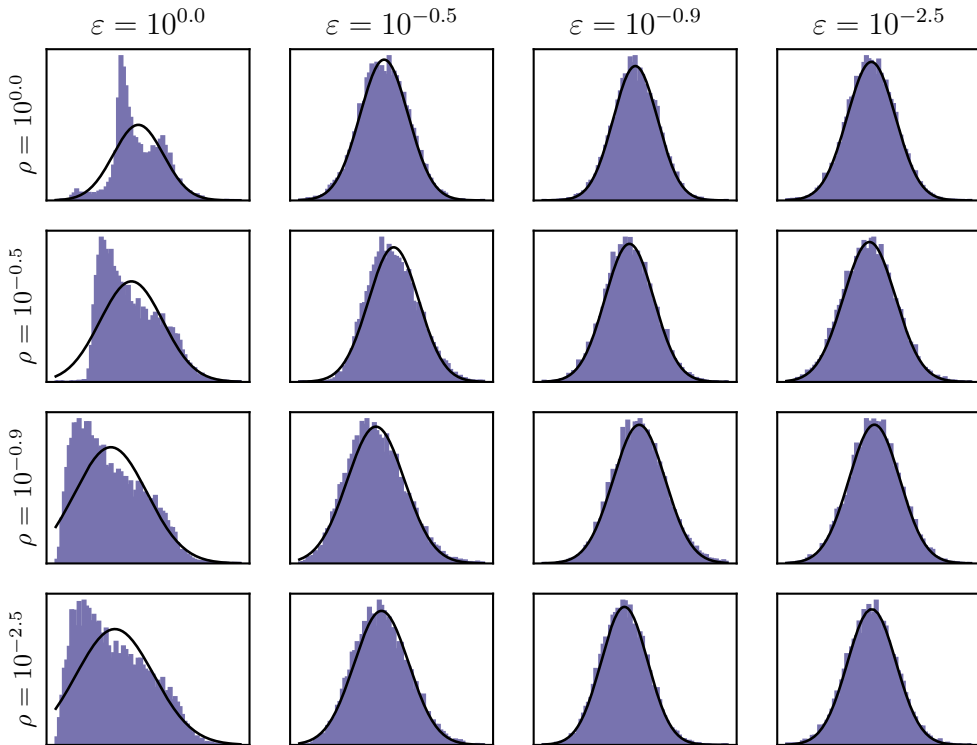


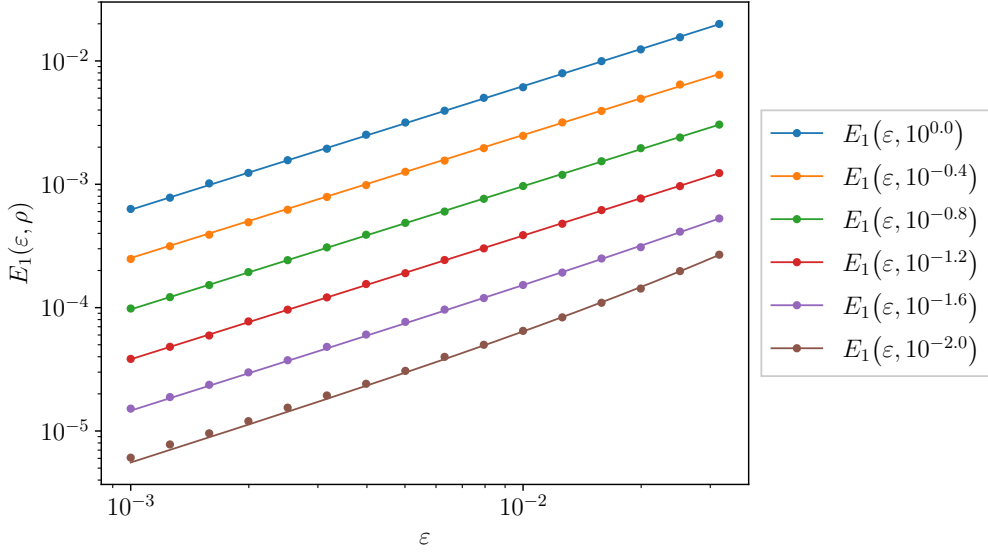
Fig. 4.3: The same arrangement as [Figure 4.1](#), but for the 1D multiplicative noise SDE [\(4.6\)](#).

the solution to [\(4.6\)](#) and the linearisation [\(4.8\)](#). For each fixed value of the initial uncertainty ρ , in [Figure 4.4a](#), we fit a line of best fit of the form $\beta_1 \varepsilon + \beta_2 \varepsilon^2$ to validate that the strong error scales as predicted. Similarly, in [Figure 4.4a](#) we fit a line of best fit of the form $\beta_0 + \beta_1 \rho$ and confirm that the linearisation error follows this scaling.

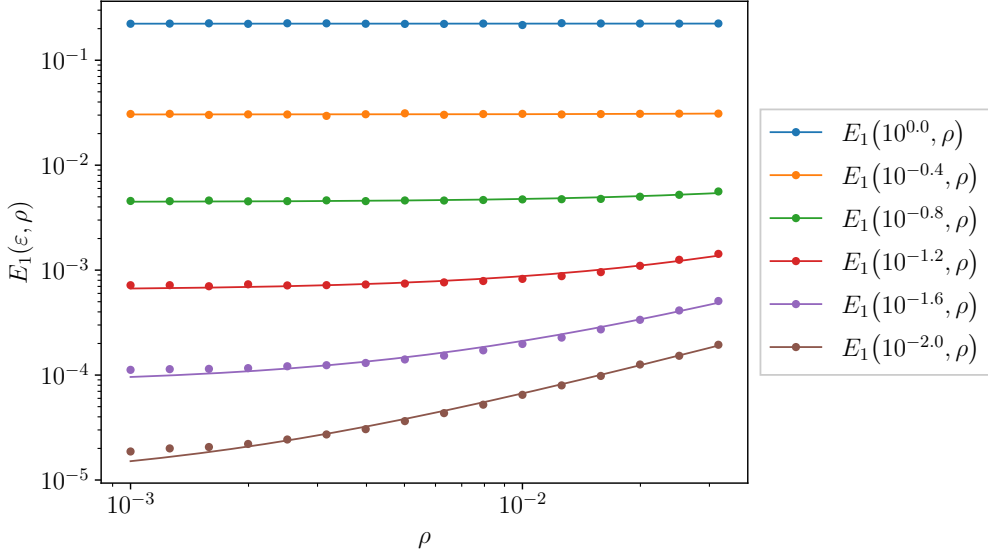
4.3. Fixed initial condition In this example, we consider a two-dimensional model and a fixed initial condition, to validate the results presented in [Section 2.3](#). Following the example in Chapter 5 of [\[34\]](#), we consider an unsteady meandering jet in two dimensions, which may serve as an idealised model of geophysical Rossby waves [\[29\]](#). The velocity field for $y \equiv (y_1, y_2)^\top$ is given by

$$u(y, t) = \begin{bmatrix} c - A \sin(K y_1) \cos(y_2) + \epsilon_{\text{mj}} l_1 \sin(k_1 (y_1 - c_1 t)) \cos(l_1 y_2) \\ AK \cos(K y_1) \sin(y_2) + \epsilon_{\text{mj}} k_1 \cos(k_1 (y_1 - c_1 t)) \sin(l_1 y_2) \end{bmatrix}. \quad (4.9)$$

The velocity field describes a kinematic travelling wave with deterministic oscillatory perturbations in a co-moving frame. Here, A is the amplitude and c is the phase speed of the primary wave, and K is the wavenumber in the y_1 -direction. The oscillatory perturbation has amplitude ϵ_{mj} , phase speed c_1 (in the co-moving frame), and wavenumbers k_1 and l_1 in the y_1 - and y_2 -directions respectively. Throughout, we take the parameter values $c = 0.5$, $A = 1$, $K = 4$, $l_1 = 2$, $k_1 = 1$, $c_1 = \pi$, and $\epsilon_{\text{mj}} = 0.3$. For these values, the



(a) Estimates of the strong order (with $r=1$) for varying ongoing uncertainty parameter ε . Each colour corresponds to a different value of the initial uncertainty parameter ρ . A (least squares) line of best fit of the form $\beta_1\varepsilon + \beta_2\varepsilon^2$ is included in the corresponding colour.



(b) Estimates of the strong order (with $r=1$) for varying initial uncertainty parameter ρ . Each colour corresponds to a different value of the ongoing uncertainty parameter ε . A (least squares) line of best fit of the form $\beta_0 + \beta_1\rho$ is included in the corresponding colour.

Fig. 4.4: Validation of the theoretical bound predicted by [Theorem 2.1](#), when $r=1$, on numerical realisations of the solution to the 1D example [\(4.6\)](#).

flow consists of a meandering jet with vortex structures within the meanders, and a chaotic zone which influences the fluid transfer between the jet and the vortices.

We introduce multiplicative noise by considering stochastic perturbations to the phase speed c and the primary amplitude A , which we model with the respective components of a 2-dimensional Wiener process $W_t = \left(W_t^{(1)}, W_t^{(2)}\right)^\top$. Then, we specify the diffusion term as

$$\sigma(y, t) = \begin{bmatrix} 1 & \sin(Ky_1) \cos(y_2) \\ 0 & K \cos(Ky_1) \sin(y_2) \end{bmatrix}. \quad (4.10)$$

We consider the fixed initial condition $x_0 = (0, 1)$ and the prediction of the model at time $t=1$. We then consider a linearisation of the SDE about the deterministic trajectory $F_0^t(x_0)$, where F_0^t is the deterministic flow map corresponding to the vector field (4.9). To compute the Gaussian distribution (2.16) of the linearised solution, we again solve (2.9) numerically with initial condition $\Sigma_0^t(x_0) = O$. Specifically, (2.9) is solved jointly with the deterministic state equation (2.2) using the hybrid method proposed by Mazzoni [25]. This hybrid method combines a Taylor-Heun approximation with a Gauss-Legendre one and ensures that the numerical solution of the covariance equation is symmetric and positive semi-definite while maintaining both accuracy and computational efficiency.

Figure 4.5 shows the resulting simulations of $y_t^{(\varepsilon)}$ for four different values of ε . The realisations are binned as a histogram and bin counts are normalised, to provide an empirical estimate of the probability density function of $y_t^{(\varepsilon)}$. Superimposed (in solid black) are the first, second and third standard-deviation contours of the probability density function of the Gaussian distribution that solves the linearised equation. The first three standard-deviation levels of the 2×2 sample covariance matrix of the realisations of $y_t^{(\varepsilon)}$, are also overlaid (in dashed blue). As ε decreases towards 0, the samples increasingly resemble a Gaussian distribution, and both the mean and covariance coincide with the corresponding limits.

For a fixed initial condition, (2.4) predicts that the expected distance between the original SDE solution and that of a linearisation satisfies

$$\mathbb{E} \left[\left\| y_t^{(\varepsilon)} - l_t^{(\varepsilon)} \right\|^r \right] \leq (K_{\nabla \nabla u} + K_{\nabla \sigma}) D_1(r, t, K_{\nabla u}, K_{\sigma}) \varepsilon^{2r}.$$

To numerically estimate the left-hand side of (2.4), we again use a Monte-Carlo estimator;

$$E_r(\varepsilon) := \frac{1}{N} \sum_{i=1}^N \left\| \hat{y}_i^{(\varepsilon)} - \hat{l}_i^{(\varepsilon)} \right\|^r.$$

For $r=1, 2, 3, 4$, $E_r(\varepsilon)$ is shown (in a logarithmic scale) for decreasing values of ε in Figure 4.6. Theorem 2.1 predicts that $\log_{10}(E_r(\varepsilon))$ should decay linearly with a slope greater than $2r$ as ε decreases to zero. The least squares lines of best fit for each value of r in Figure 4.6 show this behaviour, and are therefore consistent with Theorem 2.1.

4.4. Computing stochastic sensitivity In this section, we illustrate the computability of stochastic sensitivity as described in Theorem 3.1. We again consider the meandering jet (4.9) with multiplicative noise specified by (4.10). We take the same choice of parameters as in Section 4.3, except for the perturbation amplitude ϵ_{mj} which

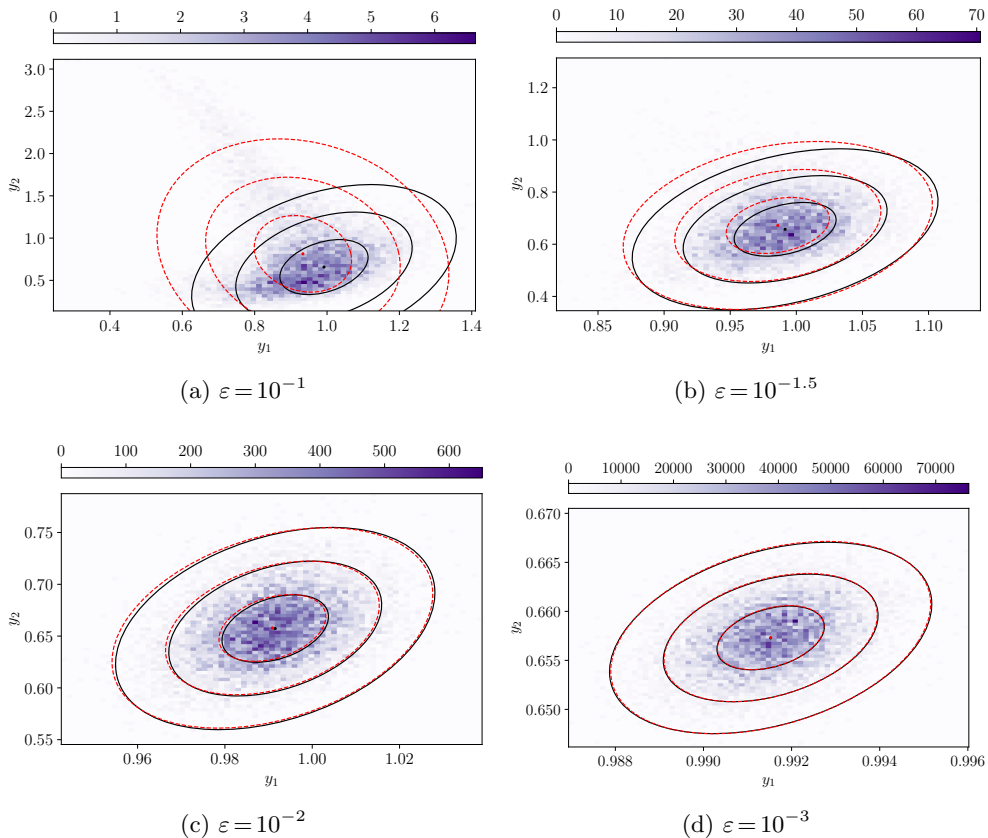


Fig. 4.5: Histograms of $y_t^{(\epsilon)}$ from direct simulation of the SDE with drift (4.9) and diffusivity (4.10) subject to the fixed initial condition, for four different ϵ values. Overlaid in black are contours of the Gaussian solution (2.16) of the linearised SDE (2.3), which correspond to the first three standard deviation levels centred at the mean $F_0^t(x)$. In dashed blue are corresponding contours computed from the sample covariance matrix of the realisations.

is varied to obtain qualitatively different behaviour in the system. For each initial condition in a 400×400 uniform grid on $[0, \pi] \times [0, \pi]$, the S^2 value is calculated using (3.1). Figure 4.7 shows the resulting S^2 field from time 0 to $t=1$, for two different values of ϵ_{mj} . The novel aspect of this section is that stochastic sensitivity is demonstrated to be computable with a straightforward spectral norm (or equivalently eigenvalue) calculation, rather than the computation involving three integrals originally presented in [4]. With the additional extension of stochastic sensitivity to arbitrary dimensions, this computation enables a fast quantification of uncertainty in a dynamical system.

The stochastic sensitivity field can highlight Lagrangian coherent structures within the flow, by identifying regions of the flow with a relatively small uncertainty, as measured by a *single* number for each initial condition. Subsets of the spatial domain corresponding to coherent regions can be extracted by taking a threshold on S^2 as described in the original work [4]; examples of coherent structure extraction with stochastic

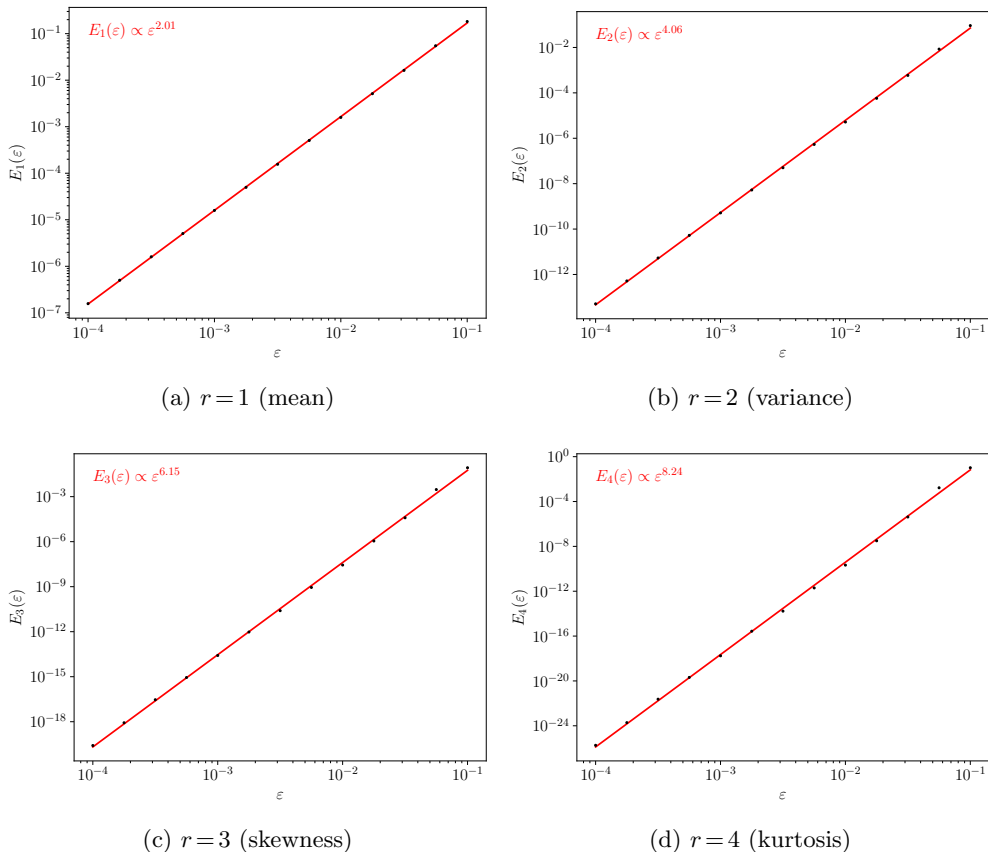


Fig. 4.6: Validation of [Theorem 2.1](#), by plotting the sample r th raw moment distance (the error metric $E_r(\varepsilon)$) between 10000 realisations of the meandering jet SDE and a corresponding linearisation, for decreasing values of ε . A line of best fit (in red) is placed on each, and the resulting slope indicated.

sensitivity on both toy models and real data can be found in [\[4\]](#) and [\[3\]](#).

5. Conclusions This paper has provided an explicit bound on the error between the solution of a nonlinear stochastic differential equation and an easily computable linearisation approximation that has been used across many different applications [\[2, 17, 21, 35, e.g.\]](#), building upon previous studies [\[9, 14, 35\]](#). The theory applies to fully non-autonomous SDEs with multiplicative noise and a random initial condition. Our bound is written in terms of the scale of the ongoing noise (that is, a scaling of the diffusivity coefficient) and a measure of the uncertainty in the initial condition using the L_r -norm. A comparison in [Section 2.1](#) suggests that our bound on the moments is tighter than implied by the gold standard [\[35\]](#) in the literature.

We also provided an explicit characterisation of the distribution of the solution to the linearised SDE, enabling efficient approximation of the original nonlinear SDE using solutions to the corresponding deterministic equation, whereas previously special cases of these computations were dispersed across other literature [\[17, 35, 36\]](#). Finally,

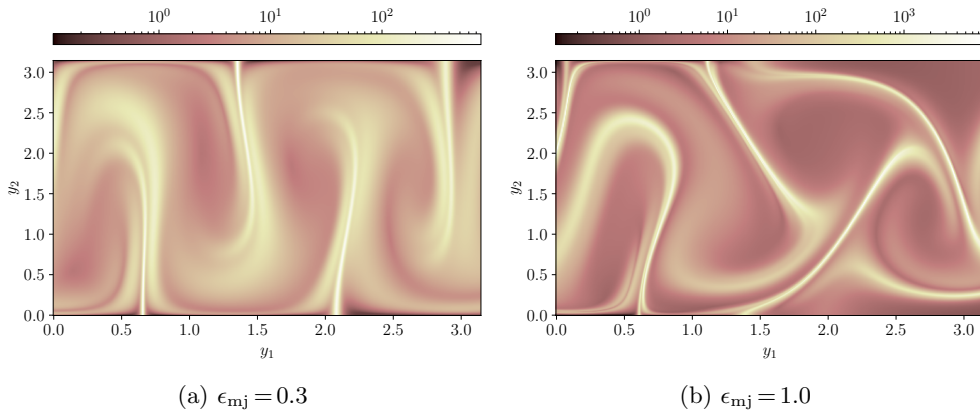


Fig. 4.7: The S^2 field of the meandering jet flow (4.9) over the time interval $[0,1]$, for two different sets of parameters with qualitatively different behaviour. The S^2 value for each initial condition is computed directly as the operator norm of the covariance matrix $\Sigma_0^1(x_0)$, as per (3.1).

we validated the form of the bound numerically on three example stochastic differential equations, and in particular find that the strong error scales with the initial uncertainty and ongoing uncertainty exactly as predicted.

Our error bound may be useful to investigate the validity of SDE linearisations in application domains such as Stochastic Parameterisation [7, 23, 28] or Data Assimilation [13, 22, 32], or to improve such linearisations by monitoring the skew or kurtosis.

This paper also supplied theoretical and computational extension to the “stochastic sensitivity” tools introduced by Balasuriya [4]. Stochastic sensitivity was hitherto derived as the variance of an unknown limiting distribution and could only be computed in two spatial dimensions; we established that stochastic sensitivity, in any number of dimensions, is computable as the operator norm of the covariance matrix of the linearised SDE. We have also established that the limiting distribution in question is Gaussian, which may provide insight into properties of stochastic sensitivity as a means of uncertainty quantification in any model (not just in the fluids context) where an n -dimensional state variable evolves according to a “best available” model.

Our work on stochastic sensitivity computes the exact distribution of a Lagrangian coherent structure measure for the first time, and may be used to employ coherent structures as ‘data’ in a Data Assimilation scheme. This idea is not new, and is explored in [24, 26, 37], albeit that the likelihood of a coherent structure was not computable. Our work fills that gap. Moreover, most traditional LCS measures are completely deterministic measures, not accounting for any uncertainty in the driving velocity field, and the sensitivity of these methods to such uncertainty has not been investigated in detail. The robustness of several LCS methods to stochastic noise has recently been explored in [3], but via stochastic simulation and summary statistics. In this paper we have presented a theoretical result for characterising Lagrangian trajectory uncertainty, which can be used to perform a purely theoretical analysis of such sensitivity in LCS computations.

Appendix A. Preliminaries for proofs.

There are several generic results and inequalities that we use several times throughout our proofs, which we state here for completeness. We write $W_t = \left(W_t^{(1)}, \dots, W_t^{(m)}\right)^\top$ as the components of the canonical m -dimensional Wiener process, where each $W_t^{(i)}$ are mutually independent 1-dimensional Wiener processes. The flow map $F_0^t: \mathbb{R}^n \rightarrow \mathbb{R}^n$ summarises solutions of the deterministic model (2.2), given by

$$F_0^t(x) = x + \int_0^t u(F_0^\tau(x), \tau) d\tau, \quad (\text{A.1})$$

for an initial condition $x \in \mathbb{R}^n$. The spatial gradient (with respect to the initial condition) of the flow map solves the equation of variations associated with (2.2), i.e.

$$\frac{\partial}{\partial t} \nabla F_0^t(x) = \nabla u(F_0^t(x), t) \nabla F_0^t(x). \quad (\text{A.2})$$

For any real numbers $x_1, \dots, x_p \geq 0$ and $r \geq 1$,

$$\left(\sum_{i=1}^p x_i\right)^r \leq p^{r-1} \sum_{i=1}^p x_i^r. \quad (\text{A.3})$$

This results from an application of the finite form of Jensen's inequality. An implication of the equivalence of the L_1 and Euclidean norms and (A.3) is that for any $z \in \mathbb{R}^n$ and $r \geq 1$,

$$\|z\|^r \leq \left(\sum_{i=1}^n |z_i|\right)^r \leq n^{r-1} \sum_{i=1}^n |z_i|^r, \quad (\text{A.4})$$

where z_i denotes the i th component of z . If each component z_i of a vector z is bounded by a constant K , then

$$\|z\| \leq \sqrt{n}K. \quad (\text{A.5})$$

Similarly, if $f: \mathbb{R} \rightarrow \mathbb{R}^n$ is a vector-valued function such that each component of f is integrable over an interval $[0, t]$, then for all $r \geq 1$,

$$\left\| \int_0^t f(\tau) d\tau \right\|^r \leq t^{r-1} \int_0^t \|f(\tau)\|^r d\tau. \quad (\text{A.6})$$

This inequality results from an application of Hölder's inequality.

Appendix B. Proof of Theorem 2.1. To prove the main result, we first require a lemma establishing a bound on the time integral of the expectation of the distance between the SDE solution and the reference deterministic trajectory.

LEMMA B.1. *Let $q \geq 1$ be such that $\delta_q < \infty$, then for all $\varepsilon > 0$ and $\tau \in [0, T]$*

$$\mathbb{E} \left[\int_0^t \left\| y_\tau^{(\varepsilon)} - F_0^t(x_0) \right\|^q d\tau \right] \leq H_1(q, t, K_{\nabla u}, K_\sigma) \varepsilon^q + H_2(q, t, K_{\nabla u}) \delta_q^q,$$

where

$$\begin{aligned} H_1(q, t, K_{\nabla u}, K_\sigma) &:= 3^{q-1} n^{3q/2} K_\sigma^{q/2} G_{q/2} t^{q/2+1} \exp\left(3^{q-1} K_{\nabla u}^q t^q\right), \\ H_2(q, t, K_{\nabla u}) &:= 3^{q-1} t \exp\left(3^{q-1} K_{\nabla u}^q t^q\right). \end{aligned}$$

Proof. Consider the integral form of (2.1),

$$y_t^{(\varepsilon)} = x + \int_0^t u\left(y_\tau^{(\varepsilon)}, \tau\right) d\tau + \varepsilon \int_0^t \sigma\left(y_\tau^{(\varepsilon)}, \tau\right) dW_\tau.$$

Using (A.1),

$$y_t^{(\varepsilon)} - F_0^t(x_0) = x - x_0 + \int_0^t \left(u\left(y_\tau^{(\varepsilon)}, \tau\right) - u\left(F_0^\tau(x_0), \tau\right) \right) d\tau + \varepsilon \int_0^t \sigma\left(y_\tau^{(\varepsilon)}, \tau\right) dW_\tau,$$

and so

$$\begin{aligned} \mathbb{E} \left[\left\| y_t^{(\varepsilon)} - F_0^t(x_0) \right\|^q \right] &\leq 3^{q-1} \mathbb{E} \left[\|x - x_0\|^q \right] \\ &\quad + 3^{q-1} t^{q-1} \mathbb{E} \left[\left\| \int_0^t \left(u\left(y_\tau^{(\varepsilon)}, \tau\right) - u\left(F_0^\tau(x_0), \tau\right) \right) d\tau \right\|^q \right] \\ &\quad + 3^{q-1} \varepsilon^q \mathbb{E} \left[\left\| \int_0^t \sigma\left(y_\tau^{(\varepsilon)}, \tau\right) dW_\tau \right\|^q \right], \end{aligned} \tag{B.1}$$

using (A.3) followed by (A.6), and taking the expectation on both sides.

Next, we establish a bound on the Itô integral term in (B.1). For $i \in \{1, \dots, n\}$, let σ_i denote the i th row of σ . Define the stochastic process

$$M_\tau^{(i)} := \sigma_i\left(y_\tau^{(\varepsilon)}, \tau\right)$$

for $\tau \in [0, t]$, so that

$$\left[\int_0^t \sigma\left(y_\tau^{(\varepsilon)}, \tau\right) dW_\tau \right]_i = \int_0^t M_\tau^{(i)} dW_\tau.$$

Since $y_t^{(\varepsilon)}$ is a strong solution to (2.1), we have that (e.g. see Definition 6.1.1 of [19])

$$\int_0^t \left\| M_\tau^{(i)} \right\|^2 d\tau \leq \int_0^t n K_\sigma^2 d\tau < \infty, \quad \text{almost surely,}$$

so we can apply the Burkholder-Davis-Gundy inequality (e.g. [19, Thm. 5.6.3]) to M_τ , which asserts that there exists a constant $G_{q/2} > 0$ depending only on q such that

$$\begin{aligned} \mathbb{E} \left[\left| \int_0^t M_\tau^{(i)} dW_\tau \right|^q \right] &\leq G_{q/2} \mathbb{E} \left[\left(\int_0^t \left\| \sigma_i\left(y_\tau^{(\varepsilon)}, \tau\right) \right\|^2 d\tau \right)^{q/2} \right] \\ &\leq G_{q/2} n^p K_\sigma^{q/2} t^{q/2}, \end{aligned}$$

where the second inequality uses [H.9](#). Then,

$$\mathbb{E} \left[\left\| \int_0^t \sigma \left(y_\tau^{(\varepsilon)}, \tau \right) dW_\tau \right\|^q \right] \leq n^{3q/2} K_\sigma^{q/2} G_{q/2} t^{q/2}, \quad (\text{B.2})$$

using [\(A.4\)](#).

Applying the bound [\(B.2\)](#) to [\(B.1\)](#), we have

$$\begin{aligned} \mathbb{E} \left[\left\| y_t^{(\varepsilon)} - F_0^t(x_0) \right\|^q \right] &\leq 3^{q-1} \delta_q^q + 3^{q-1} \varepsilon^q n^{3q/2} K_\sigma^{q/2} G_{q/2} t^{q/2} \\ &\quad + 3^{q-1} t^{q-1} \mathbb{E} \left[\int_0^t \left\| u \left(y_\tau^{(\varepsilon)}, \tau \right) - u \left(F_0^\tau(x_0), \tau \right) \right\|^q d\tau \right]. \end{aligned} \quad (\text{B.3})$$

We note that $\mathbb{E} \left[\left\| y_t^{(\varepsilon)} - F_0^t(x) \right\|^q \right] < \infty$ from [H.3](#), so by Tonelli's theorem (e.g. [\[12, Thm. 2.3.9\]](#)),

$$\mathbb{E} \left[\int_0^t \left\| y_\tau^{(\varepsilon)} - F_0^\tau(x_0) \right\|^q d\tau \right] = \int_0^t \mathbb{E} \left[\left\| y_\tau^{(\varepsilon)} - F_0^\tau(x_0) \right\|^q \right] d\tau.$$

Now, using the Lipschitz continuity of u from [H.6](#) on [\(B.3\)](#) and interchanging the expectation and integral,

$$\begin{aligned} \mathbb{E} \left[\left\| y_t^{(\varepsilon)} - F_0^t(x_0) \right\|^q \right] &\leq 3^{q-1} \delta_q^q + 3^{q-1} K_{\nabla u}^q t^{q-1} \int_0^t \mathbb{E} \left[\left\| y_\tau^{(\varepsilon)} - F_0^\tau(x_0) \right\|^q \right] d\tau \\ &\quad + 3^{q-1} \varepsilon^q n^{3q/2} K_\sigma^{q/2} G_{q/2} t^{q/2}. \end{aligned}$$

Applying Grönwall's inequality and using the monotonicity of the resulting bound in t , we have that for any $\tau \in [0, t]$,

$$\begin{aligned} \mathbb{E} \left[\left\| y_\tau^{(\varepsilon)} - F_0^\tau(x_0) \right\|^q \right] &\leq 3^{q-1} \varepsilon^q n^{3q/2} S^{q/2} G_{q/2} t^{q/2} \exp \left(3^{q-1} K_{\nabla u}^q t^q \right) \\ &\quad + 3^{q-1} \exp \left(3^{q-1} K_{\nabla u}^q t^q \right) \delta_q^q. \end{aligned}$$

Integrating both sides with respect to time and again using Tonelli's theorem, we have

$$\begin{aligned} \mathbb{E} \left[\int_0^t \left\| y_\tau^{(\varepsilon)} - F_0^\tau(x_0) \right\|^q d\tau \right] &\leq 3^{q-1} n^{3q/2} S^{q/2} G_{q/2} t^{q/2+1} \exp \left(3^{q-1} K_{\nabla u}^q t^q \right) \varepsilon^q \\ &\quad + 3^{q-1} t \exp \left(3^{q-1} K_{\nabla u}^q t^q \right) \delta_q^q, \end{aligned}$$

as desired. \square

With these bounds established, we can now prove [Theorem 2.1](#). Subtracting the integral forms of [\(2.3\)](#) and [\(2.1\)](#) gives

$$\begin{aligned} y_t^{(\varepsilon)} - l_t^{(\varepsilon)} &= \int_0^t \left[u \left(y_\tau^{(\varepsilon)}, \tau \right) - u \left(F_0^\tau(x_0), \tau \right) - \nabla u \left(F_0^\tau(x_0) \right) \left(l_\tau^{(\varepsilon)} - F_0^\tau(x_0) \right) \right] d\tau \\ &\quad + \int_0^t \left[\varepsilon \sigma \left(y_\tau^{(\varepsilon)}, \tau \right) - \varepsilon \sigma \left(F_0^\tau(x_0), \tau \right) \right] dW_\tau \end{aligned}$$

$$\begin{aligned}
&= \int_0^t \left[u(y_\tau^{(\varepsilon)}, \tau) - \left(u(F_0^\tau(x_0), \tau) + \nabla u(F_0^\tau(x_0)) (y_\tau^{(\varepsilon)} - F_0^\tau(x_0)) \right) \right] d\tau \\
&\quad + \int_0^t \nabla u(F_0^\tau(x_0), \tau) [y_\tau^{(\varepsilon)} - l_\tau^{(\varepsilon)}] d\tau \\
&\quad + \varepsilon \int_0^t \left[\sigma(y_\tau^{(\varepsilon)}, \tau) - \sigma(F_0^\tau(x_0), \tau) \right] dW_\tau \\
&= A(t) + B(t) + \varepsilon C(t),
\end{aligned}$$

where

$$\begin{aligned}
A(t) &:= \int_0^t \left[u(y_\tau^{(\varepsilon)}, \tau) - \left(u(F_0^\tau(x_0), \tau) + \nabla u(F_0^\tau(x_0)) (y_\tau^{(\varepsilon)} - F_0^\tau(x_0)) \right) \right] d\tau \\
B(t) &:= \int_0^t \nabla u(F_0^\tau(x_0), \tau) [y_\tau^{(\varepsilon)} - l_\tau^{(\varepsilon)}] d\tau \\
C(t) &:= \int_0^t \left[\sigma(y_\tau^{(\varepsilon)}, \tau) - \sigma(F_0^\tau(x_0), \tau) \right] dW_\tau.
\end{aligned}$$

Then, using (A.3) and taking expectation,

$$\mathbb{E} \left[\|y_t^{(\varepsilon)} - l_t^{(\varepsilon)}\|^r \right] \leq 3^{r-1} \left(\mathbb{E} [\|A(t)\|^r] + \mathbb{E} [\|B(t)\|^r] + \varepsilon^r \mathbb{E} [\|C(t)\|^r] \right). \quad (\text{B.4})$$

First consider $A(t)$, for which the integrand is the expected difference between the drift u evaluated along SDE solution and the first-order Taylor expansion of u about the reference deterministic trajectory. Since for any $t \in [0, T]$, $u(\cdot, t)$ is twice continuously differentiable under H.2, for each $i = 1, \dots, n$ there exists by Taylor's theorem (e.g. see [16, Cor. A9.3.]) a function $R_i : \mathbb{R}^n \times [0, T] \rightarrow \mathbb{R}$ such that

$$u_i(z, \tau) = u_i(F_0^\tau(x_0), \tau) + \left[\nabla u_i(F_0^\tau(x_0), \tau) \right] (z - F_0^\tau(x_0)) + R_i(z, \tau) \quad (\text{B.5})$$

for any $z \in \mathbb{R}^n$, where u_i denotes the i th component of u . The function R_i satisfies

$$|R_i(z, \tau)| \leq \frac{1}{2} \left\| \nabla \nabla u_i(F_0^\tau(x_0), \tau) \right\| \|z - F_0^\tau(x_0)\|^2 \leq \frac{K_{\nabla \nabla u}}{2} \|z - F_0^\tau(x_0)\|^2. \quad (\text{B.6})$$

Let $R(z, \tau) := (R_1(z, \tau), \dots, R_n(z, \tau))^\top$, then

$$A(t) = \int_0^t R(y_\tau^{(\varepsilon)}, \tau) d\tau,$$

and since each component of R is bounded as in (B.6), using (A.5)

$$\left\| R(y_t^{(\varepsilon)}, \tau) \right\| \leq \frac{\sqrt{n} K_{\nabla \nabla u}}{2} \|y_t^{(\varepsilon)} - F_0^\tau(x_0)\|^2.$$

Taking the norm and expectation then gives

$$\mathbb{E} [\|A(t)\|^r] = \mathbb{E} \left[\left\| \int_0^t R(y_\tau^{(\varepsilon)}, \tau) d\tau \right\|^r \right]$$

$$\begin{aligned}
&\leq t^{r-1} \mathbb{E} \left[\int_0^t \left\| R(y_\tau^{(\varepsilon)}, \tau) \right\|^r d\tau \right] \\
&\leq \frac{t^{r-1} n^{r/2} K_{\nabla \nabla u}^r}{2^r} \mathbb{E} \left[\int_0^t \left\| y_\tau^{(\varepsilon)} - F_0^\tau(x_0) \right\|^{2r} d\tau \right] \\
&\leq \frac{t^{r-1} n^{r/2} K_{\nabla \nabla u}^r H_1(2r, t, K_{\nabla u}, K_\sigma)}{2^r} \varepsilon^{2r} \\
&\quad + \frac{t^{r-1} n^{r/2} K_{\nabla \nabla u}^r H_2(2r, t, K_{\nabla u})}{2^r} \delta_{2r}^{2r},
\end{aligned} \tag{B.7}$$

where the first inequality uses (A.6), and H_1 and H_2 are obtained from Lemma B.1.

Next, consider $B(t)$, for which

$$\mathbb{E} \left[\|B(t)\|^r \right] \leq \int_0^t t^{r-1} K_{\nabla u}^r \mathbb{E} \left[\left\| y_\tau^{(\varepsilon)} - l_t^{(\varepsilon)} \right\|^r \right] d\tau. \tag{B.8}$$

using (A.6) and then H.3, and interchanging the expectation and the integral uses the fact that $\mathbb{E} \left[\left\| y_\tau^{(\varepsilon)} \right\|^r \right] < \infty$ and $\mathbb{E} \left[\left\| l_\tau^{(\varepsilon)} \right\|^r \right] < \infty$.

Finally, consider $C(t)$. For each $i \in \{1, \dots, n\}$, define the stochastic process

$$N_\tau^{(i)} := \sigma_{i \cdot} \left(y_\tau^{(\varepsilon)}, \tau \right) - \sigma_{i \cdot} \left(F_0^\tau(x_0), \tau \right).$$

Then, the i th component of $C(t)$ is

$$[C(t)]_i = \int_0^t N_\tau^{(i)} dW_\tau.$$

From H.9 and using (A.5),

$$\int_0^t \left\| N_\tau^{(i)} \right\|^2 d\tau \leq \int_0^t 4n K_\sigma^2 d\tau < \infty,$$

so we can apply the Burkholder-Davis-Gundy inequality on $N_\tau^{(i)}$ to write

$$\begin{aligned}
\mathbb{E} \left[\left\| [C(t)]_i \right\|^r \right] &\leq G_{r/2} \mathbb{E} \left[\left(\int_0^t \left\| \sigma_{i \cdot} \left(y_\tau^{(\varepsilon)}, \tau \right) - \sigma_{i \cdot} \left(F_0^\tau(x_0), \tau \right) \right\|^2 d\tau \right)^{r/2} \right] \\
&\leq G_{r/2} \mathbb{E} \left[\left(\int_0^t K_{\nabla \sigma}^2 \left\| y_\tau^{(\varepsilon)} - F_0^\tau(x_0) \right\|^2 d\tau \right)^{r/2} \right] \\
&\leq G_{r/2} K_{\nabla \sigma}^r t^{r/2-1} \mathbb{E} \left[\int_0^t \left\| y_\tau^{(\varepsilon)} - F_0^\tau(x_0) \right\|^r d\tau \right] \\
&\leq G_{r/2} K_{\nabla \sigma}^r t^{r/2-1} H_1(r, t, K_{\nabla u}, K_\sigma) \varepsilon^r \\
&\quad + G_{r/2} K_{\nabla \sigma}^r t^{r/2-1} H_2(r, t, K_{\nabla u}) \delta_r^r,
\end{aligned} \tag{B.9}$$

where the second inequality uses the Lipschitz condition on σ in H.6, the third inequality uses (A.6), and the fourth inequality uses Lemma B.1 with $q=r$. Then, we have

$$\begin{aligned}
\mathbb{E} \left[\|C(t)\|^r \right] &\leq n^r G_{r/2} K_{\nabla \sigma}^r t^{r/2-1} H_1(r, t, K_{\nabla u}, K_\sigma) \varepsilon^r \\
&\quad + n^r G_{r/2} K_{\nabla \sigma}^r t^{r/2-1} H_2(r, t, K_{\nabla u}) \delta_r^r,
\end{aligned} \tag{B.10}$$

using (A.4), and then (B.9).

Combining (B.7), (B.8) and (B.10) into (B.4), we have

$$\begin{aligned} \mathbb{E} \left[\left\| y_t^{(\varepsilon)} - l_t^{(\varepsilon)} \right\|^r \right] &\leq \frac{3^{r-1} t^{r-1} n^{r/2} K_{\nabla \nabla u}^r H_1(2r, t, K_{\nabla u}, K_\sigma)}{2^r} \varepsilon^{2r} \\ &\quad + \frac{3^{r-1} t^{r-1} n^{r/2} K_{\nabla \nabla u}^r H_2(2r, t, K_{\nabla u})}{2^r} \delta_{2r}^{2r} \\ &\quad + \int_0^t 3^{r-1} t^{r-1} K_{\nabla u}^r \mathbb{E} \left[\left\| y_t^{(\varepsilon)} - l_t^{(\varepsilon)} \right\|^r \right] d\tau \\ &\quad + 3^{r-1} G_{r/2} K_{\nabla \sigma}^r t^{r/2-1} H_1(r, t, K_{\nabla u}, K_\sigma) \varepsilon^{2r} \\ &\quad + 3^{r-1} G_{r/2} K_{\nabla \sigma}^r t^{r/2-1} H_2(r, t, K_{\nabla u}) \varepsilon^r \delta_r^r. \end{aligned}$$

Applying Grönwall's inequality, noting that H_1 and H_2 are non-decreasing in t , we have

$$\begin{aligned} \mathbb{E} \left[\left\| y_t^{(\varepsilon)} - l_t^{(\varepsilon)} \right\|^r \right] &\leq \frac{3^{r-1} t^{r-1} n^{r/2} K_{\nabla \nabla u}^r H_1(2r, t, K_{\nabla u}, K_\sigma)}{2^r} \exp \left(3^{r-1} t^r K_{\nabla u}^r \right) \varepsilon^{2r} \\ &\quad + \frac{3^{r-1} t^{r-1} n^{r/2} K_{\nabla \nabla u}^r H_2(2r, t, K_{\nabla u})}{2^r} \exp \left(3^{r-1} t^r K_{\nabla u}^r \right) \delta_{2r}^{2r} \\ &\quad + 3^{r-1} G_{r/2} K_{\nabla \sigma}^r t^{r/2-1} \exp \left(3^{r-1} t^r K_{\nabla u}^r \right) H_1(r, t, K_{\nabla u}, K_\sigma) \varepsilon^{2r} \\ &\quad + 3^{r-1} G_{r/2} K_{\nabla \sigma}^r t^{r/2-1} \exp \left(3^{r-1} t^r K_{\nabla u}^r \right) H_2(r, t, K_{\nabla u}) \varepsilon^r \delta_r^r. \end{aligned}$$

Set

$$D_1(r, t, K_{\nabla u}, K_\sigma) := 3^{r-1} \exp \left(3^{r-1} t^r K_{\nabla u}^r \right) K_M(r, t, K_{\nabla u}, K_\sigma) \quad (\text{B.11a})$$

$$D_2(r, t, K_{\nabla u}) := 3^{r-1} t^{r-1} n^{r/2} H_2(2r, t, K_{\nabla u}) \exp \left(3^{r-1} t^r K_{\nabla u}^r \right) \quad (\text{B.11b})$$

$$D_3(r, t, K_{\nabla u}) := 3^{r-1} G_{r/2} t^{r/2-1} H_2(r, t, K_{\nabla u}) \exp \left(3^{r-1} t^r K_{\nabla u}^r \right), \quad (\text{B.11c})$$

where

$$K_M(r, t, K_{\nabla u}, K_\sigma) := \max \left\{ \frac{t^{r-1} n^{r/2} H_1(2r, t, K_{\nabla u}, K_\sigma)}{2^r}, G_{r/2} t^{r/2-1} H_1(r, t, K_{\nabla u}, K_\sigma) \right\},$$

then we have shown the desired result.

Appendix C. Proof of Theorem 2.2.

Next, we show that the strong solution to the linearised SDE (2.3) can be written as the independent sum (2.5). Let

$$M_t = h \left(l_t^{(\varepsilon)}, t \right) := \frac{1}{\varepsilon} \left[\nabla F_0^t(x_0) \right]^{-1} \left(l_t^{(\varepsilon)} - F_0^t(x_0) \right),$$

where $l_t^{(\varepsilon)}$ is the strong solution to (2.3). Then, the required derivatives for applying Itô's Lemma (e.g. see Theorem 5.5.1 of [19]) are

$$\begin{aligned} M_0 &= h \left(l_0^{(\varepsilon)}, 0 \right) = \frac{1}{\varepsilon} (x - x_0) \\ \frac{\partial h}{\partial t} &= -\frac{1}{\varepsilon} \left[\nabla F_0^t(x_0) \right]^{-1} \frac{\partial \nabla F_0^t(x_0)}{\partial t} \left[\nabla F_0^t(x_0) \right]^{-1} \left(l_t^{(\varepsilon)} - F_0^t(x_0) \right) \\ &\quad - \frac{1}{\varepsilon} \left[\nabla F_0^t(x_0) \right]^{-1} u \left(F_0^t(x_0), t \right) \end{aligned}$$

$$\begin{aligned}\nabla h\left(l_t^{(\varepsilon)}, t\right) &= \frac{1}{\varepsilon} \left[\nabla F_0^t(x_0)\right]^{-1} \\ \nabla \nabla h\left(l_t^{(\varepsilon)}, t\right) &= O,\end{aligned}$$

where in computing the t derivative we have used the fact that $F_0^t(x_0)$ solves the deterministic ODE (2.2). Thus,

$$\begin{aligned}M_t &= \frac{1}{\varepsilon} (x - x_0) + \frac{1}{\varepsilon} \int_0^t \left(- \left[\nabla F_0^\tau(x_0)\right]^{-1} \frac{\partial \nabla F_0^\tau(x_0)}{\partial \tau} \left[\nabla F_0^\tau(x_0)\right]^{-1} \left(l_\tau^{(\varepsilon)} - F_0^\tau(x_0)\right) \right. \\ &\quad \left. - \left[\nabla F_0^\tau(x_0)\right]^{-1} u\left(F_0^\tau(x_0), \tau\right) \right. \\ &\quad \left. + \left[\nabla F_0^\tau(x_0)\right]^{-1} \left[u\left(F_0^\tau(x_0), \tau\right) + \nabla u\left(F_0^\tau(x_0), \tau\right) \left(l_\tau^{(\varepsilon)} - F_0^\tau(x_0)\right) \right] \right) d\tau \\ &\quad + \int_0^t \left[\nabla F_0^\tau(x_0)\right]^{-1} \sigma\left(F_0^\tau(x_0), \tau\right) dW_\tau \\ &= \frac{1}{\varepsilon} (x - x_0) + \frac{1}{\varepsilon} \int_0^t \left(- \left[\nabla F_0^\tau(x_0)\right]^{-1} \nabla u\left(F_0^\tau(x_0), \tau\right) \left(l_\tau^{(\varepsilon)} - F_0^\tau(x_0)\right) \right. \\ &\quad \left. + \left[\nabla F_0^\tau(x_0)\right]^{-1} \nabla u\left(F_0^\tau(x_0), \tau\right) \left(l_\tau^{(\varepsilon)} - F_0^\tau(x_0)\right) \right) d\tau \\ &\quad + \int_0^t \left[\nabla F_0^\tau(x_0)\right]^{-1} \sigma\left(F_0^\tau(x_0), \tau\right) dW_\tau \\ &= \frac{1}{\varepsilon} (x - x_0) + \int_0^t \left[\nabla F_0^\tau(x_0)\right]^{-1} \sigma\left(F_0^\tau(x_0), \tau\right) dW_\tau,\end{aligned}$$

where we reach the second line by using the equation of variations (A.2) satisfied by $\nabla F_0^\tau(x_0)$. It follows that

$$l_t^{(\varepsilon)} = \nabla F_0^t(x_0)(x - x_0) + F_0^t(x_0) + \varepsilon \int_0^t \nabla F_0^t(x_0) \left[\nabla F_0^\tau(x_0)\right]^{-1} \sigma\left(F_0^\tau(x_0), \tau\right) d\tau$$

is a strong solution to (2.3) to (2.5).

By H.7, x is independent of the Wiener process W_t , and since independence is preserved under limits and linear transformations it follows that the Itô integral $\varepsilon \nabla F_0^t(x_0) \int_0^t L(x_0, \tau) dW_\tau$ and $\nabla F_0^t(x_0)(x - x_0)$ are independent.

Appendix D. Proof of Corollary 2.1. We first establish that the Itô integral of a matrix-valued deterministic function with respect to a multidimensional Wiener process is a multidimensional Gaussian process. This is a well-known result in the scalar case, and the extension to our case is straightforward.

LEMMA D.1. *Let $a, b \in \mathbb{R}$ and let $g: [a, b] \rightarrow \mathbb{R}^{n \times n}$ be a matrix-valued deterministic function such that each element of g is Itô-integrable. Consider the Itô integral*

$$\mathcal{I}[g] := \int_a^b g(t) dW_t,$$

Then, the integral $\mathcal{I}[g]$ is a n -dimensional multivariate Gaussian random variable.

Proof. For $i, j \in \{1, \dots, n\}$, let $g_{ij}: [a, b] \rightarrow \mathbb{R}$ be the (i, j) th element of g . Then, let

$$\mathcal{I}[g_{ij}] := \int_a^b g_{ij}(t) dW_t^{(i)},$$

so that the i th element of $\mathcal{I}[g]$ is

$$\mathcal{I}[g]_i = \sum_{j=1}^n \mathcal{I}[g_{ij}].$$

Each $\mathcal{I}[g_{ij}]$ is an Itô integral of a deterministic, scalar-valued function with respect to a one-dimensional Brownian motion, which is well-known to be a Gaussian process (e.g. see [1, Lem. 4.3.11]). Moreover, each element of $\mathcal{I}[g]$ is the sum of independent Gaussian random variables and is therefore itself Gaussian. Hence, $\mathcal{I}[g]$ follows a multivariate Gaussian distribution. \square

Now, we move onto showing **Corollary 2.1**. Consider the Itô integral

$$\mathcal{I}[L] = \int_0^t L(x_0, \tau) dW_\tau.$$

For any fixed $t \in [0, T]$, the integrand is a deterministic, matrix-valued function, and is therefore follows a n -dimensional Gaussian distribution. Moreover [19],

$$\mathbb{E}[\mathcal{I}[L]] = 0,$$

and

$$\mathbb{V}[\mathcal{I}[L]] = \mathbb{E} \left[\left(\int_0^t L(x_0, \tau) dW_\tau \right) \left(\int_0^t L(x_0, \tau) dW_\tau \right)^\top \right].$$

Let L_{ij} denote the (i, j) th element of L , then the (i, j) th element of the variance is

$$\begin{aligned} [\mathbb{V}[\mathcal{I}[L]]]_{ij} &= \sum_{k=1}^m \sum_{l=1}^m \mathbb{E} \left[\left(\int_0^t L_{ik}(x_0, \tau) dW_\tau^{(k)} \right) \left(\int_0^t L_{jl}(x_0, \tau) dW_\tau^{(l)} \right) \right] \\ &= \sum_{k=1}^m \mathbb{E} \left[\left(\int_0^t L_{ik}(x_0, \tau) dW_\tau^{(k)} \right) \left(\int_0^t L_{jk}(x_0, \tau) dW_\tau^{(k)} \right) \right] \\ &\quad + \sum_{k=1}^m \sum_{\substack{l=1 \\ l \neq k}}^m \mathbb{E} \left[\left(\int_0^t L_{ik}(x_0, \tau) dW_\tau^{(k)} \right) \right] \mathbb{E} \left[\left(\int_0^t L_{jl}(x_0, \tau) dW_\tau^{(l)} \right) \right] \\ &= \sum_{k=1}^m \int_0^t L_{ik}(x_0, \tau) L_{jk}(x_0, \tau) d\tau, \end{aligned}$$

where the second equality the fact that $W_t^{(k)}$ is independent of $W_t^{(l)}$ for $k \neq l$ and the third equality uses Itô's isometry [19]. Hence, we have that

$$\int_0^t L(x_0, \tau) d\tau \sim \mathcal{N} \left(0, \int_0^t L(x_0, \tau) L(x_0, \tau)^\top d\tau \right),$$

completing the proof of **Theorem 2.2**. Next, we show that the mean and covariance of $l_t^{(\varepsilon)}$ are given explicitly by (2.7) and (2.8) respectively. It follows immediately from (2.5) that the mean of $l_t^{(\varepsilon)}$ is

$$\mathbb{E} \left[l_t^{(\varepsilon)} \right] = \mathbb{E} \left[\nabla F_0^t(x_0)(x - x_0) \right] + F_0^t(x_0) + \varepsilon \nabla F_0^t(x_0) \mathbb{E} \left[\int_0^t L(x_0, \tau) d\tau \right]$$

$$= \nabla F_0^t(x) (\mathbb{E}[x] - x_0) + F_0^t(x_0),$$

thus showing (2.7).

Since the two summands in (2.5) are independent, the variance of $l_t^{(\varepsilon)}$ is

$$\begin{aligned} \mathbb{V}[l_t^{(\varepsilon)}] &= \mathbb{V}[\nabla F_0^t(x_0)(x - x_0)] + \mathbb{V}\left[\varepsilon \nabla F_0^t(x_0) \int_0^t L(x_0, \tau) dW_\tau\right] \\ &= \nabla F_0^t(x_0) \left(\mathbb{V}[x] + \varepsilon^2 \int_0^t L(x_0, \tau) L(x_0, \tau)^\top d\tau \right) [\nabla F_0^t(x_0)]^\top \end{aligned}$$

where the variance of the Itô integral was established in Appendix C.

Finally, we show Remark 2.4, i.e. that $\mathbb{V}[l_t^{(\varepsilon)}]$ is the solution to the matrix differential equation (2.9). Directly differentiating the expression (2.8)

$$\begin{aligned} \frac{d\mathbb{V}[l_t^{(\varepsilon)}]}{dt} &= \frac{\partial \nabla F_0^t(x_0)}{\partial t} \left(\mathbb{V}[x] + \varepsilon^2 \int_0^t L(x_0, \tau) L(x_0, \tau)^\top d\tau \right) [\nabla F_0^t(x_0)]^\top \\ &\quad + \nabla F_0^t(x_0) \left(\mathbb{V}[x] + \varepsilon^2 \int_0^t L(x_0, \tau) L(x_0, \tau)^\top d\tau \right) \left[\frac{\partial \nabla F_0^t(x_0)}{\partial t} \right]^\top \\ &\quad + \varepsilon^2 \nabla F_0^t(x_0) L(x_0, t) L(x_0, t)^\top [\nabla F_0^t(x_0)]^\top \\ &= \nabla u(F_0^t(x_0), t) \nabla F_0^t(x_0) \left(\mathbb{V}[x] + \varepsilon^2 \int_0^t L(x_0, \tau) L(x_0, \tau)^\top d\tau \right) [\nabla F_0^t(x_0)]^\top \\ &\quad + \nabla F_0^t(x_0) \left(\mathbb{V}[x] + \varepsilon^2 \int_0^t L(x_0, \tau) L(x_0, \tau)^\top d\tau \right) [\nabla F_0^t(x_0)]^\top [\nabla u(F_0^t(x_0), t)]^\top \\ &\quad + \varepsilon^2 \sigma(F_0^t(x_0), t) \sigma(F_0^t(x_0), t)^\top \\ &= \nabla u(F_0^t(x_0), t) \mathbb{V}[l_t^{(\varepsilon)}] + \mathbb{V}[l_t^{(\varepsilon)}] [\nabla u(F_0^t(x_0), t)]^\top \\ &\quad + \varepsilon^2 \sigma(F_0^t(x_0), t) \sigma(F_0^t(x_0), t)^\top, \end{aligned}$$

where the second inequality has used the equation of variations (A.2).

Appendix E. Proof of Theorem 3.1. Let $t \in [0, T]$ and consider the solutions $y_t^{(\varepsilon)}$ to (2.1) and $l_t^{(\varepsilon)}$ to (2.3) subject to the fixed initial condition $x_0 \in \mathbb{R}^n$. On the vector space of n -dimensional random vectors with each component having finite expectation and variance, define the function ρ as

$$\rho(z) := \|\mathbb{V}[z]\|^\frac{1}{2}.$$

Then, ρ is a semi-norm, which can be verified using properties of the spectral norm and the Cauchy-Schwarz inequality. This proof is provided in the supplementary materials. Then,

$$\left| \left\| \mathbb{V}[y_t^{(\varepsilon)}] \right\|^\frac{1}{2} - \left\| \mathbb{V}[l_t^{(\varepsilon)}] \right\|^\frac{1}{2} \right| = \left| \rho(y_t^{(\varepsilon)}) - \rho(l_t^{(\varepsilon)}) \right|$$

$$\begin{aligned}
&\leq \rho\left(y_t^{(\varepsilon)} - l_t^{(\varepsilon)}\right) \\
&= \left\| \mathbb{E} \left[\left(y_t^{(\varepsilon)} - l_t^{(\varepsilon)} \right) \left(y_t^{(\varepsilon)} - l_t^{(\varepsilon)} \right)^\top \right] - \mathbb{E} \left[y_t^{(\varepsilon)} - l_t^{(\varepsilon)} \right] \mathbb{E} \left[y_t^{(\varepsilon)} - l_t^{(\varepsilon)} \right]^\top \right\|^\frac{1}{2} \\
&\leq \left(\mathbb{E} \left[\left\| y_t^{(\varepsilon)} - l_t^{(\varepsilon)} \right\|^2 \right] + \mathbb{E} \left[\left\| y_t^{(\varepsilon)} - l_t^{(\varepsilon)} \right\| \right]^2 \right)^\frac{1}{2} \\
&\leq \left((K_{\nabla \nabla u} + K_{\nabla \sigma}) D_1(2, t) + (K_{\nabla \nabla u} + K_{\nabla \sigma})^2 D_1(1, t)^2 \right)^{1/2} \varepsilon^2
\end{aligned}$$

where the first inequality uses the reverse triangle inequality, the second inequality uses the Jensen's inequality and properties of the spectral norm, and the third inequality results from [Theorem 2.1](#). Thus,

$$\left| \left\| \frac{1}{\varepsilon^2} \mathbb{V} \left[y_t^{(\varepsilon)} \right] \right\|^\frac{1}{2} - \left\| \frac{1}{\varepsilon^2} \mathbb{V} \left[l_t^{(\varepsilon)} \right] \right\|^\frac{1}{2} \right| \leq \left((K_{\nabla \nabla u} + K_{\nabla \sigma}) D_1(2, t) + (K_{\nabla \nabla u} + K_{\nabla \sigma})^2 D_1(1, t)^2 \right)^\frac{1}{2} \varepsilon,$$

and so taking the limit of ε to zero and squaring both sides,

$$\lim_{\varepsilon \downarrow 0} \left\| \frac{1}{\varepsilon^2} \mathbb{V} \left[y_t^{(\varepsilon)} \right] \right\| = \lim_{\varepsilon \downarrow 0} \left\| \frac{1}{\varepsilon^2} \mathbb{V} \left[l_t^{(\varepsilon)} \right] \right\|. \quad (\text{E.1})$$

Now, for $\varepsilon > 0$, define

$$\begin{aligned}
S_{(\varepsilon)}^2(x_0, t) &:= \sup \left\{ \mathbb{V} \left[\frac{1}{\varepsilon} p^\top \left(y_t^{(\varepsilon)} - F_0^t(x_0) \right) \right] \mid p \in \mathbb{R}^n, \|p\| = 1 \right\} \\
&= \frac{1}{\varepsilon^2} \sup \left\{ p^\top \mathbb{V} \left[y_t^{(\varepsilon)} \right] p \mid p \in \mathbb{R}^n, \|p\| = 1 \right\}
\end{aligned}$$

Since $\mathbb{V} \left[y_t^{(\varepsilon)} \right]$ is symmetric and positive definite, the Cholesky decomposition provides a lower triangular $n \times n$ matrix $\Pi^{(\varepsilon)}$ such that $\mathbb{V} \left[y_t^{(\varepsilon)} \right] = \Pi^{(\varepsilon)} \left[\Pi^{(\varepsilon)} \right]^\top$, allowing us to write

$$\begin{aligned}
S_{(\varepsilon)}^2(x_0, t) &= \frac{1}{\varepsilon^2} \sup \left\{ \left\| \Pi^{(\varepsilon)} p \right\|^2 \mid p \in \mathbb{R}^n, \|p\| = 1 \right\} \\
&= \frac{1}{\varepsilon^2} \left\| \Pi^{(\varepsilon)} \right\|^2 \\
&= \frac{1}{\varepsilon^2} \left\| \mathbb{V} \left[y_t^{(\varepsilon)}(x) \right] \right\|,
\end{aligned}$$

using properties of the spectral norm. Taking the limit as ε approaches zero and using [\(E.1\)](#),

$$S^2(x_0, t) = \lim_{\varepsilon \downarrow 0} S_{(\varepsilon)}^2(x_0, t) = \lim_{\varepsilon \downarrow 0} \left\| \frac{1}{\varepsilon^2} \mathbb{V} \left[y_t^{(\varepsilon)} \right] \right\| = \left\| \Sigma_0^t(x_0) \right\|,$$

where Σ_0^t is defined in [\(2.14\)](#). Since $\Sigma_0^t(x_0)$ is symmetric and positive definite, the operator norm, and therefore $S^2(x_0, t)$, is given by the largest eigenvalue of $\Sigma_0^t(x_0)$.

REFERENCES

- [1] D. Applebaum. *Lévy Processes and Stochastic Calculus*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, United Kingdom, 1st edition, 2004. ISBN 978-0-511-21119-5. [D](#)
- [2] C. Archambeau, D. Cornford, M. Opper, and J. Shawe-Taylor. Gaussian Process Approximations of Stochastic Differential Equations. In *Gaussian Processes in Practice*, pages 1–16. PMLR, Mar. 2007. [1](#), [2](#), [5](#)
- [3] A. Badza, T. W. Mattner, and S. Balasuriya. How sensitive are Lagrangian coherent structures to uncertainties in data? *Physica D: Nonlinear Phenomena*, 444:133580, Feb. 2023. doi: 10.1016/j.physd.2022.133580. [3](#), [4.4](#), [5](#)
- [4] S. Balasuriya. Stochastic Sensitivity: A Computable Lagrangian Uncertainty Measure for Unsteady Flows. *SIAM Review*, 62:781–816, Nov. 2020. doi: 10.1137/18M1222922. [1](#), [2](#), [2.3](#), [3](#), [3](#), [3](#), [4.4](#), [4.4](#), [5](#)
- [5] S. Balasuriya, N. T. Ouellette, and I. I. Rypina. Generalized Lagrangian coherent structures. *Physica D: Nonlinear Phenomena*, 372:31–51, June 2018. doi: 10.1016/j.physd.2018.01.011. [1](#)
- [6] F. Balibrea-Iniesta, C. Lopesino, S. Wiggins, and A. M. Mancho. Lagrangian Descriptors for Stochastic Differential Equations: A Tool for Revealing the Phase Portrait of Stochastic Dynamical Systems. *International Journal of Bifurcation and Chaos*, 26(13):1630036, Dec. 2016. doi: 10.1142/S0218127416300366. [1](#)
- [7] J. Berner, U. Achatz, L. Batté, L. Bengtsson, A. de la Cámara, H. M. Christensen, M. Colangeli, D. R. B. Coleman, D. Crommelin, S. I. Dolaptchiev, C. L. E. Franzke, P. Friederichs, P. Imkeller, H. Järvinen, S. Juricke, V. Kitsios, F. Lott, V. Lucarini, S. Mahajan, T. N. Palmer, C. Penland, M. Sakradzija, J.-S. von Storch, A. Weisheimer, M. Weniger, P. D. Williams, and J.-I. Yano. Stochastic Parameterization: Toward a New View of Weather and Climate Models. *Bulletin of the American Meteorological Society*, 98(3):565–588, Mar. 2017. doi: 10.1175/BAMS-D-15-00268.1. [5](#)
- [8] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah. Julia: A Fresh Approach to Numerical Computing. *SIAM Review*, 59(1):65–98, Jan. 2017. doi: 10.1137/141000671. [4](#)
- [9] Y. N. Blagoveshchenskii. Diffusion Processes Depending on a Small Parameter. *Theory of Probability and its Applications*, 7(2):17, 1962. doi: 10.1137/1107013. [1](#), [2](#), [2.3](#), [5](#)
- [10] M. Branicki and K. Uda. Lagrangian Uncertainty Quantification and Information Inequalities for Stochastic Flows. *SIAM/ASA Journal on Uncertainty Quantification*, 9(3):1242–1313, 2021. [1](#)
- [11] M. Branicki and K. Uda. Path-Based Divergence Rates and Lagrangian Uncertainty in Stochastic Flows. *SIAM Journal on Applied Dynamical Systems*, pages 419–482, Mar. 2023. doi: 10.1137/21M1466530. [1](#)
- [12] P. Brémaud. *Probability Theory and Stochastic Processes*. Universitext. Springer, 1 edition, 2020. ISBN 978-3-030-40182-5. [B](#)
- [13] A. Budhiraja, E. Friedlander, C. Guider, C. K. R. T. Jones, and J. Maclean. Assimilating Data into Models. In *Handbook of Environmental and Ecological Statistics*. Chapman and Hall/CRC, 1st edition edition, 2019. ISBN 978-1-315-15250-9. [1](#), [5](#)
- [14] M. I. Freidlin and A. D. Wentzell. *Random Perturbations of Dynamical Systems*. Grundlehren Der Mathematischen Wissenschaften. Springer, New York, NY, 2 edition, 1998. [1](#), [2](#), [2](#), [2.3](#), [5](#)
- [15] A. Hadjighasem, M. Farazmand, D. Blazeovski, G. Froyland, and G. Haller. A critical comparison of Lagrangian methods for coherent structure detection. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(5):053104, May 2017. doi: 10.1063/1.4982720. [1](#)
- [16] J. H. Hubbard and B. B. Hubbard. *Vector Calculus, Linear Algebra, and Differential Forms: A Unified Approach*. Matrix Editions, Ithaca, NY, 4th ed edition, 2009. ISBN 978-0-9715766-5-0. [B](#)
- [17] A. H. Jazwinski. *Stochastic Processes and Filtering Theory*, volume 64 of *Mathematics in Science and Engineering*. Elsevier Science, Burlington, 2014. ISBN 978-0-08-096090-6. [1](#), [2](#), [2.2](#), [5](#)
- [18] I. T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. Springer, New York, NY, 2 edition, 2002. ISBN 978-0-387-22440-4. [3](#)
- [19] G. Kallianpur and P. Sundar. *Stochastic Analysis and Diffusion Processes*. Number 24 in Oxford Graduate Texts in Mathematics. Oxford University Press, Oxford, United Kingdom, first edition edition, 2014. ISBN 978-0-19-965706-3 978-0-19-965707-0. [1](#), [2](#), [B](#), [C](#), [D](#)
- [20] I. Kamenkovich, I. I. Rypina, and P. Berloff. Properties and Origins of the Anisotropic Eddy-Induced Transport in the North Atlantic. *Journal of Physical Oceanography*, 45(3):778–791, Mar. 2015. doi: 10.1175/JPO-D-14-0164.1. [1](#)
- [21] B. Kaszás and G. Haller. Universal upper estimate for prediction errors under moderate model uncertainty. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 30(11):113144, Nov. 2020. doi: 10.1063/5.0021665. [1](#), [5](#)
- [22] K. Law, A. Stuart, and K. Zygalakis. *Data Assimilation: A Mathematical Introduction*. Number volume 62 in Texts in Applied Mathematics. Springer, Cham Heidelberg New York Dordrecht

- London, 2015. ISBN 978-3-319-20324-9 978-3-319-36687-6. doi: 10.1007/978-3-319-20325-6. [1](#), [5](#)
- [23] M. Leutbecher, S.-J. Lock, P. Ollinaho, S. T. K. Lang, G. Balsamo, P. Bechtold, M. Bonavita, H. M. Christensen, M. Diamantakis, E. Dutra, S. English, M. Fisher, R. M. Forbes, J. Goddard, T. Haiden, R. J. Hogan, S. Juricke, H. Lawrence, D. MacLeod, L. Magnusson, S. Malardel, S. Mas-sart, I. Sandu, P. K. Smolarkiewicz, A. Subramanian, F. Vitart, N. Wedi, and A. Weisheimer. Stochastic representations of model uncertainties at ECMWF: State of the art and future vision. *Quarterly Journal of the Royal Meteorological Society*, 143(707):2315–2339, 2017. doi: 10.1002/qj.3094. [5](#)
- [24] J. Maclean, N. Santitissadeekorn, and C. K. R. T. Jones. A coherent structure approach for parameter estimation in Lagrangian Data Assimilation. *Physica D: Nonlinear Phenomena*, 360: 36–45, Dec. 2017. doi: 10.1016/j.physd.2017.08.007. [5](#)
- [25] T. Mazzoni. Computational aspects of continuous–discrete extended Kalman-filtering. *Computational Statistics*, 23(4):519–539, Oct. 2008. doi: 10.1007/s00180-007-0094-4. [4.3](#)
- [26] M. Morzfeld, J. Adams, S. Lunderman, and R. Orozco. Feature-based data assimilation in geophysics. *Nonlinear Processes in Geophysics*, 25(2):355–374, May 2018. doi: 10.5194/npg-25-355-2018. [5](#)
- [27] B. Øksendal. *Stochastic Differential Equations*. Universitext. Springer, Berlin, Heidelberg, 6th edition, 2003. ISBN 978-3-540-04758-2 978-3-642-14394-6. doi: 10.1007/978-3-642-14394-6. [1](#)
- [28] T. N. Palmer. Stochastic weather and climate models. *Nature Reviews Physics*, 1(7):463–471, July 2019. doi: 10.1038/s42254-019-0062-2. [5](#)
- [29] R. T. Pierrehumbert. Chaotic mixing of tracer and vorticity by modulated travelling Rossby waves. *Geophysical & Astrophysical Fluid Dynamics*, 58(1-4):285–319, July 1991. doi: 10.1080/03091929108227343. [4.3](#)
- [30] C. Rackauckas and Q. Nie. Adaptive methods for stochastic differential equations via natural embeddings and rejection sampling with memory. *Discrete and continuous dynamical systems. Series B*, 22(7):2731–2761, 2017. doi: 10.3934/dcdsb.2017133. [4.1](#)
- [31] C. Rackauckas and Q. Nie. DifferentialEquations.jl – A Performant and Feature-Rich Ecosystem for Solving Differential Equations in Julia. *Journal of Open Research Software*, 5(1):15, May 2017. doi: 10.5334/jors.151. [4](#)
- [32] S. Reich and C. Cotter. *Probabilistic Forecasting and Bayesian Data Assimilation*. Cambridge University Press, Cambridge, 2015. ISBN 978-1-107-06939-8. doi: 10.1017/CBO9781107706804. [1](#), [5](#)
- [33] A. Rößler. Runge-Kutta Methods for the Strong Approximation of Solutions of Stochastic Differential Equations. *SIAM Journal on Numerical Analysis*, 48(3):922–952, 2010. doi: 10.1137/09076636X. [4.1](#)
- [34] R. M. Samelson and S. Wiggins. *Lagrangian Transport in Geophysical Jets and Waves: The Dynamical Systems Approach*, volume 31 of *Interdisciplinary Applied Mathematics*. Springer, New York, NY, 2006. [4.3](#)
- [35] D. Sanz-Alonso and A. M. Stuart. Gaussian Approximations of Small Noise Diffusions in Kullback-Leibler Divergence. *Communications in Mathematical Sciences*, 15(7):2087–2097, Oct. 2017. doi: https://dx.doi.org/10.4310/CMS.2017.v15.n7.a13. [1](#), [2](#), [2.1](#), [2.1](#), [2.1](#), [5](#)
- [36] S. Särkkä and A. Solin. *Applied Stochastic Differential Equations*. Institute of Mathematical Statistics Textbooks. Cambridge University Press, Cambridge, 2019. ISBN 978-1-316-51008-7. doi: 10.1017/9781108186735. [1](#), [2](#), [2.2](#), [5](#)
- [37] K. L. Schlueter-Kuck and J. O. Dabiri. Model parameter estimation using coherent structure colouring. *Journal of Fluid Mechanics*, 861:886–900, Feb. 2019. doi: 10.1017/jfm.2018.898. [5](#)
- [38] P. Sura, M. Newman, C. Penland, and P. Sardeshmukh. Multiplicative Noise and Non-Gaussianity: A Paradigm for Atmospheric Regimes? *Journal of the Atmospheric Sciences*, 62(5):1391–1409, May 2005. doi: 10.1175/JAS3408.1. [1](#)