Parameter estimation with increased precision for elliptic and hypo-elliptic diffusions

YUGA IGUCHI[1,a], ALEXANDROS BESKOS[1,b] and MATTHEW GRAHAM[2,c]

[1]*Department of Statistical Science, University College London, London, UK*, [a]*yuga.iguchi.21@ucl.ac.uk*, [b]*a.beskos@ucl.ac.uk*
[2]*Advanced Research Computing Centre, University College London, London, UK*, [c]*m.graham@ucl.ac.uk*

This work aims at making a comprehensive contribution in the general area of parametric inference for discretely observed diffusion processes. Established approaches for likelihood-based estimation invoke a time-discretisation scheme for the approximation of the intractable transition dynamics of the Stochastic Differential Equation (SDE) model over finite time periods. The scheme is applied for a step-size $\delta > 0$, that is either user-selected or determined by the data. Recent research has highlighted the critical effect of the choice of numerical scheme on the behaviour of derived parameter estimates in the setting of *hypo-elliptic* SDEs. In brief, in our work, first, we develop two weak second order *sampling schemes* (to cover both hypo-elliptic and elliptic SDEs) and produce a *small time expansion* for the density of the schemes to form a proxy for the true intractable SDE transition density. Then, we establish a collection of analytic results for likelihood-based parameter estimates obtained via the formed proxies, thus providing a theoretical framework that showcases advantages from the use of the developed methodology for SDE calibration. We present numerical results from carrying out classical or Bayesian inference, for both elliptic and hypo-elliptic SDEs.

Keywords: CLT; data augmentation; hypo-elliptic diffusion; small time density expansion; stochastic differential equation

1. Introduction

Our work is placed within the general framework of parametric inference for diffusion processes. Calibration approaches under broad observation regimes, in both Bayesian and classical settings, necessitate the use of a numerical scheme used as proxy for the underlying, typically intractable, Markovian dynamics of the model over finite time steps. Following the latest contributions in the area, we aim to make connections between approximation schemes (i.e., a research area mainly within the remit of stochastic analysis) and their impact on the accuracy and performance of induced likelihood-based inferential approaches (i.e., the field of statistical calibration for Stochastic Differential Equations (SDEs)).

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ be a filtered probability space and $B = (B_{1,t}, \ldots, B_{d_R,t})_{t\geq 0}$ a standard d_R-dimensional Brownian motion defined thereon, $d_R \geq 1$. We use the convention $B_{0,t} = t$. Consider the following general class of SDEs:

$$dX_t = \begin{bmatrix} dX_{R,t} \\ dX_{S,t} \end{bmatrix} = \begin{bmatrix} V_{R,0}(X_t, \beta) \\ V_{S,0}(X_t, \gamma) \end{bmatrix} dt + \sum_{1\leq k\leq d_R} \begin{bmatrix} V_{R,k}(X_t, \sigma) \\ \mathbf{0}_{d_S} \end{bmatrix} dB_{k,t}, \qquad X_0 = x_0 \in \mathbb{R}^d, \tag{1}$$

with $V_{R,0} : \mathbb{R}^d \times \Theta_\beta \to \mathbb{R}^{d_R}$, $V_{S,0} : \mathbb{R}^d \times \Theta_\gamma \to \mathbb{R}^{d_S}$, $V_{R,k} : \mathbb{R}^d \times \Theta_\sigma \to \mathbb{R}^{d_R}$, $1 \leq k \leq N$, where $d_S \geq 0$ and $d = d_R + d_S$. Also, $\Theta_\beta \subseteq \mathbb{R}^{d_\beta}$, $\Theta_\gamma \subseteq \mathbb{R}^{d_\gamma}$, $\Theta_\sigma \subseteq \mathbb{R}^{d_\sigma}$, $d_\beta \geq 1$, $d_\gamma \geq 0$, $d_\sigma \geq 1$. We set $\theta = (\beta, \gamma, \sigma)$ and $\Theta = \Theta_\beta \times \Theta_\gamma \times \Theta_\sigma$. $X_{R,t}$ and $X_{S,t}$ denote the rough and smooth components, respectively, of process X_t that solves (1). For $d_S > 0$, $\{X_t\}_{t\geq 0}$ is a *hypo-elliptic diffusion* if the law of X_t, $t > 0$, admits a density with respect to (w.r.t.) the Lebesgue measure, while for $d_S = 0$ the process is an elliptic one.

For such a diffusion model, we consider the problem of estimating $\theta \in \Theta$ given observations for state X_t at the discrete-time instances $0 \le t_0 < t_1 < \cdots < t_{n-1} < t_n$, $n \ge 0$. For simplicity, we assume that observation times are equidistant and set $\Delta := t_i - t_{i-1}$, $1 \le i \le n$.

Non-linear SDEs do not permit, in general, analytical solutions, thus one must rely on approximate, time-discretisation schemes to generate diffusion sample paths and obtain closed-form expressions for the SDE transition density. Development of approximation schemes with high accuracy can lead to effective parameter estimation methods, e.g. to Bayesian data augmentation without excessive imputation of latent variables to cover a non-small step-size Δ. Indeed, several works on Markov chain Monte-Carlo (MCMC) methods for diffusion processes suggest use of high order approximations, e.g. the Milstein or a strong 1.5 order scheme (Kloeden and Platen, 1992), but practical application of such schemes is restricted to limited classes of SDEs due to involvement of intractable random variables, such as the *Lévy area* $\int_0^\Delta (B_{k_1,s} dB_{k_2,s} - B_{k_2,s} dB_{k_1,s})$, $k_1 \neq k_2$. Motivated by the above, we develop our work under the following strategy. First, we propose an explicit sampling scheme for the generation of sample paths for model (1), with higher order accuracy in distributional sense compared to classical (conditionally) Gaussian numerical schemes, e.g. the Euler-Maruyama scheme. Then, we derive an Edgeworth-type density expansion of the above sampling scheme. Instances of the expansion will be used to provide proxies for the true intractable transition density and, ultimately, likelihood-based parameter estimates. Finally, we provide analytic results showcasing advantages of the obtained parameter estimates in both high/low-frequency observation regimes, where in the former case one assumes $\Delta \to 0$ and in the latter that Δ is fixed and not small enough.

One set of analytic results that we provide in this work – and in accordance with recent contributions involving hypo-elliptic SDEs by, e.g., Ditlevsen and Samson (2019), Gloter and Yoshida (2021) – will correspond to asymptotic results for the Maximum Likelihood Estimator (MLE), in the high-frequency regime, with $n \to \infty$ and $\Delta = \Delta_n \to 0$, and in a setting of *complete* discrete-time observations. That is, the dataset Y^c is as follows:

$$Y^c := \{X_{t_0}, X_{t_1}, \ldots, X_{t_n}\}. \tag{2}$$

We refer to the dataset in (2) in the high-frequency setting as the 'complete observation regime' in the sequel. We stress that the discretisation schemes developed in the above context are practically relevant for generic observation regimes, e.g. within Bayesian data augmentation methods. Consideration of these latter methods motivates the derivation of further analytic results, now in a low-frequency regime. In particular, for elliptic SDEs, we explicitly connect the weak order of the numerical scheme with the proximity between the true transition density over a period of size Δ and the transition density produced by convolution of $M \ge 1$ proxy transition densities over time intervals of size $\delta = \Delta/M$.

We briefly review recent works in the high-frequency regime. In the hypo-elliptic setting, a first main contribution is the work of Pokern, Stuart and Wiberg (2009) that uses an Itô-Taylor expansion to add a noise term of size $O(\Delta^{3/2})$ in the numerical scheme for the smooth component of SDE (1), thus obtaining a non-degenerate (conditionally) Gaussian approximation of the true transition density. The class of models considered in Pokern, Stuart and Wiberg (2009) is restrictive and no analytical results are provided. Ditlevsen and Samson (2019) make a major contribution by starting with a strong 1.5 order scheme before removing terms that do not affect their asymptotic results. Noise of size $O(\Delta^{3/2})$ is propagated onto the smooth component (as in Pokern, Stuart and Wiberg (2009)) and quantities of size $O(\Delta^2)$ are retained in the mean terms. These latter components remove the bias for the estimates of drift parameters observed experimentally (within a Bayesian data augmentation setting) in Pokern, Stuart and Wiberg (2009). Ditlevsen and Samson (2019) provide analytic asymptotic results for a contrast estimator in the complete observation regime (2), under the condition $\Delta_n = o(n^{-1/2})$; this is a type of condition referred to as 'rapidly increasing experimental design' in early investigations for elliptic SDEs in Prakasa Rao (1988). The estimation procedure in Ditlevsen and Samson (2019) is separated into two contrast functions, one for γ assuming knowledge of the true values of (β, σ), and vice-versa for the contrast function for (β, σ). The analysis provides marginal CLTs for the estimates of the parameters rather than an ideal joint CLT. The class of models covered is restricted to a scalar smooth component and diagonal diffusion coefficient matrix for the rough component. Gloter and Yoshida (2020, 2021) provide the most recent contributions. Closer to our purposes, Gloter and Yoshida (2020) describe a non-adaptive approach (as opposed to one-step adaptive methods in Gloter and Yoshida (2021)) and prove a CLT in the complete observation regime for $\Delta_n = o(n^{-1/2})$, as in Ditlevsen and Samson (2019), but without strong restrictions on the class of models. In the elliptic setting, Kessler (1997) developed contrast functions for the scalar case that, within regime (2), deliver estimates satisfying a CLT for $\Delta_n = o(n^{-1/q})$, for any integer $q \geq 2$. The method in Kessler (1997) is based on the use of a Gaussian density for the approximation of the transition density of the SDE for small Δ, together with high order expansions in Δ for the mean and variance of the SDE transitions. Uchida and Yoshida (2012) extended such results to general model dimension based on multi-step adaptive estimates.

Our main contributions can be summarised as follows:

(a) In the hypo-elliptic case, we propose a new weak second order sampling scheme. The scheme is explicit as one can produce SDE sample paths by generating Gaussian variates. A related weak second order scheme is also put forward for the elliptic case.

(b) The above sampling scheme possess, in general, an intractable transition density due to involvement of polynomials of Gaussian variates. We derive a closed-form *small time density expansion* of the scheme by making use of tools from Malliavin calculus.

(c) In the high-frequency regime, we develop our contrast function by selecting appropriate high order terms from the density expansion. We then prove a joint CLT for the deduced parameter estimates under $\Delta_n = o(n^{-1/3})$, which is the largest step-size permitted in the hypo-elliptic case to the best of our knowledge. In particular, a main improvement by the new estimator – when compared with the existing estimator satisfying the CLT under $\Delta_n = o(n^{-1/2})$ – is observed in the estimation of diffusion parameter σ.

(d) In the low-frequency regime, we choose particular high order terms from the small time density expansion to develop our local weak third order transition density scheme. We then study the use of the developed scheme in the practical setting where its density is iteratively applied with user-specified steps of size $\delta = \Delta/M$, $M \geq 1$, to cover fixed inter-observation times of length Δ. We prove that the induced bias is $\mathcal{O}(M^{-2})$ for elliptic SDEs.

(e) We show numerical examples illustrating the benefits of the new schemes in applications with hypo-elliptic SDEs in the high-frequency regime. In the context of Bayesian data augmentation, in the low-frequency regime, we apply the sampling scheme to an elliptic model and observe that the bias in the induced posterior is reduced when comparing with the Euler-Maruyama scheme.

The remaining part of the paper is organised as follows. Section 2 puts forward our sampling schemes for elliptic and hypo-elliptic SDEs. Section 3 derives the closed-form small time density expansion for the sampling scheme. Section 4 provides a collection of analytical results for parameter inference carried out via appropriate choice of high order terms from the developed density expansion, both in the high and low-frequency regimes. Section 5 provides numerical experiments related to the analytic results. Section 6 concludes our work.

Notation. We define:

$$V_0(\cdot, \beta, \gamma) = \left[V_{R,0}(\cdot, \beta)^\top, V_{S,0}(\cdot, \gamma)^\top \right]^\top, \quad V_k(\cdot, \sigma) = \left[V_{R,k}(\cdot, \sigma)^\top, \mathbf{0}_{d_S}^\top \right]^\top, \quad 1 \leq k \leq d_R.$$

We set $V_R = V_R(x, \sigma) = \left[V_{R,1}(x, \sigma), \ldots, V_{R,d_R}(x, \sigma) \right] \in \mathbb{R}^{d_R \times d_R}$, and define:

$$a_R(x, \sigma) = V_R(x, \sigma) V_R(x, \sigma)^\top \in \mathbb{R}^{d_R \times d_R}. \tag{3}$$

Let $C_b^\infty(\mathbb{R}^n;\mathbb{R}^m)$ (resp. $C_p^\infty(\mathbb{R}^n;\mathbb{R}^m)$), $m,n\ge 1$, be the space of smooth (i.e. infinitely differentiable) bounded (resp. of polynomial growth) functions $f:\mathbb{R}^n\to\mathbb{R}^m$ with bounded derivatives (resp. with derivatives of polynomial growth). We write $\partial_u = \big[\frac{\partial}{\partial u^1},\dots,\frac{\partial}{\partial u^n}\big]^\top$, $\partial_u^2 = \partial_u\partial_u^\top \equiv \big(\frac{\partial^2}{\partial u^i \partial u^j}\big)_{i,j=1}^n$ for the standard differential operators acting upon maps $\mathbb{R}^n\to\mathbb{R}$, $n\ge 1$. As above with $u=(u^1,\dots,u^n)^\top\in\mathbb{R}^n$, we use superscripts to specify co-ordinates of a vector when needed. For sufficiently smooth functions $f=f(x)=f(x_R,x_S):\mathbb{R}^d\to\mathbb{R}$, we define:

$$\hat{V}_0 f = \hat{V}_0 f(x,\theta) := \big\langle V_0, \partial_x f\big\rangle + \tfrac{1}{2}\sum_{1\le k\le d_R} V_{R,k}^\top\big(\partial_{x_R}^2 f\big)V_{R,k};$$

$$\hat{V}_k f = \hat{V}_k f(x,\sigma) := \big\langle V_k, \partial_x f\big\rangle = \big\langle V_{R,k}, \partial_{x_R} f\big\rangle, \qquad 1\le k\le d_R.$$

$\hat{V}_0$, $\hat{V}_k$ apply to vector-valued functions by separate consideration of the scalar co-ordinates. Note that the original SDE (1) is equivalently given as the Stratonovitch-type SDE with drift function defined as:

$$\widetilde{V}_0(x,\theta) = V_0(x,\theta) - \tfrac{1}{2}\sum_{1\le k\le d_R}\hat{V}_k V_0(x,\theta). \tag{4}$$

We define the vector-valued function:

$$[V_k,V_l] = \hat{V}_k V_l(x,\theta) - \hat{V}_l V_k(x,\theta), \quad 1\le k,l\le d_R. \tag{5}$$

We write $\theta^\dagger=(\beta^\dagger,\gamma^\dagger,\sigma^\dagger)\in\Theta$ for the (assumed unique) true value of $\theta=(\beta,\gamma,\sigma)$. We denote iterated stochastic integrals w.r.t. Brownian paths as:

$$I_\alpha(t) = \int_0^t\cdots\int_0^{t_2} dB_{\alpha_1,t_1}\cdots dB_{\alpha_l,t_l}, \quad t>0,\ \ \alpha\in\{0,1,\dots,d_R\}^l,\ l\ge 0. \tag{6}$$

We will sometimes write $\mathbb{P}_\theta$, $\mathbb{E}_\theta$ to emphasise the involvement of θ in calculations. Similarly, we will write $\mathbb{P}_{\theta^\dagger}$, $\mathbb{E}_{\theta^\dagger}$ to stress, when needed, that derivations are under the true parameter value, and write $\xrightarrow{\mathbb{P}_{\theta^\dagger}}$, $\xrightarrow{\mathcal{L}_{\theta^\dagger}}$ to express convergence in probability and in distribution, respectively, under $\theta^\dagger$. For a multi-index $\alpha=(\alpha_1,\dots,\alpha_l)\in\{1,\dots,d\}^l$, $l\ge 1$ and a sufficiently smooth $f:\mathbb{R}^d\to\mathbb{R}$, we write $\partial_\alpha^\xi f(\xi) := \partial^l f(\xi)/\partial\xi_{\alpha_1}\cdots\partial\xi_{\alpha_l}$, $\xi\in\mathbb{R}^d$.

2. Sampling schemes for elliptic & hypo-elliptic diffusions

We propose explicit sampling schemes for the SDE (1). We introduce the conditions so that the law of X_t, $t>0$, admits a smooth Lebesgue density in Section 2.1. Then, we present the sampling schemes in Section 2.2. Hereafter, we make use of the notation X_t^x, $t>0$, when needed to emphasise the initial state $X_0=x\in\mathbb{R}^d$.

2.1. Basic assumptions for diffusion class

We introduce basic conditions to characterise the SDE (1) we consider in this work.

(H1) Θ is a compact subset of $\mathbb{R}^{d_\theta}$. For each $x\in\mathbb{R}^d$, and any multi-index $\alpha\in\{1,\dots,d\}^l$, $l\ge 0$, the function $\theta\mapsto\partial_\alpha^x V_j^i(x,\theta)$, $0\le j\le d_R$, $1\le i\le d$, is continuous.

(H2) For each $\theta\in\Theta$, $V_j(\cdot,\theta)\in C_b^\infty(\mathbb{R}^d;\mathbb{R}^d)$, $0\le j\le d_R$.

(H3) The matrix $a_R=a_R(x;\sigma)$ defined in (3) is positive-definite for all $(x,\theta)\in\mathbb{R}^d\times\Theta$.

(H4) For any $\theta \in \Theta$ and any $x \in \mathbb{R}^d$, the set of the $2d_R$ vectors:

$$\left\{V_1(x,\theta), \ldots, V_{d_R}(x,\theta), [\widetilde{V}_0, V_1](x,\theta), \ldots, [\widetilde{V}_0, V_{d_R}](x,\theta)\right\}$$

spans $\mathbb{R}^d$, where $\widetilde{V}_0 = \widetilde{V}_0(x,\theta)$ is defined in (4) and $[\widetilde{V}_0, V_j](x,\theta)$ is defined as (5).

(H4) is relevant for hypo-elliptic SDEs and is stronger than Hörmander's condition (Nualart, 2006), thus the law of X_t^x is absolutely continuous w.r.t. the Lebesgue measure for any $t > 0$, $(x,\theta) \in \mathbb{R}^d \times \Theta$. Hörmander's condition allows for iterated Lie brackets of order larger than one (used here), e.g. of order two $[\widetilde{V}_0, [\widetilde{V}_0, V_k]]$, $1 \le k \le d_R$, or above, to obtain vectors spanning $\mathbb{R}^d$. We require (H4) so that certain discretisation schemes arising in our methodology also have a Lebesgue density. Note that (H2), (H4) combined imply that the density of X_t^x, $t > 0$, is infinitely differentiable (Nualart, 2006, Theorems 2.3.2, 2.3.3).

2.2. Approximate sampling scheme

We propose approximate sampling schemes for elliptic/hypo-elliptic SDE (1) that satisfy the following three Criteria:

(i) The scheme is explicit.

(ii) The scheme has local weak third order accuracy, i.e. for test functions $\varphi \in C_p^\infty(\mathbb{R}^d;\mathbb{R})$ there exists a constant $C = C(x,\theta) > 0$ such that:

$$\left|\mathbb{E}_\theta[\varphi(X_\Delta^x)] - \mathbb{E}_\theta[\varphi(\bar{X}_\Delta^x)]\right| \le C\Delta^3, \quad \Delta \ge 0,$$

where $\bar{X}_\Delta^x$ denotes the approximated SDE position after a single step of size $\Delta > 0$ with an initial value $\bar{X}_0^x = x \in \mathbb{R}^d$.

(iii) The distribution of $\bar{X}_\Delta^x$ admits a Lebesgue density.

Criterion (ii) implies that the sampling scheme is a weak second order approximation of $\{X_t\}$ in the following sense. Let $T > 0$, $n \ge 1$, and consider the partition $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = T$ with $t_i - t_{i-1} = T/n$, $1 \le i \le n$. Then, for $\varphi \in C_p^\infty(\mathbb{R}^d;\mathbb{R})$, there exists a constant $C = C(x,T,\theta) > 0$ such that $\left|\mathbb{E}_\theta[\varphi(X_T^x)] - \mathbb{E}_\theta[\varphi(\bar{X}_T^{(n),x})]\right| \le C/n^2$, where $\bar{X}_T^{(n),x}$ is obtained after n iterations of a single step of the scheme, starting from the initial state $x \in \mathbb{R}^d$. Criterion (iii) is required so that a well-defined contrast function can be obtained and estimates of the parameters be produced.

2.2.1. *Sampling scheme for elliptic diffusions*

Before writing down our scheme satisfying Criteria (i)-(iii) above for elliptic diffusions, we present the ideas underlying the construction of a weak second order discretisation scheme (equivalently, a local weak third order scheme) following the moment matching techniques in (Milstein and Tretyakov, 2021, Section 2.1.2). We apply an Itô-Taylor expansion (Kloeden and Platen, 1992) for SDE model (1) with $d_S = 0$ and obtain that $X_{e,\Delta}^x = \hat{X}_{e,\Delta}^x + \rho_{e,\Delta}(x,\theta)$, where

$$\hat{X}_{e,\Delta}^x = x + V_{R,0}(x,\theta)\Delta + \sum_{1 \le k \le d_R} V_{R,k}(x,\theta)B_{k,\Delta} + \sum_{0 \le k_1,k_2 \le d_R} \hat{V}_{k_1}V_{R,k_2}(x,\theta)I_{(k_1,k_2)}(\Delta), \tag{7}$$

and $\rho_{e,\Delta}(x,\theta)$ is a residual term involving stochastic iterated integrals of order three and above. Then, it can be shown (Milstein and Tretyakov, 2021) that for test functions $\varphi \in C_p^\infty(\mathbb{R}^{d_R};\mathbb{R})$, $\hat{X}_\Delta^x$ has local weak third order accuracy, though it is not an explicit scheme due to the presence of $I_{(k_1,k_2)}(\Delta)$, $1 \le k_1, k_2 \le d_R$, $k_1 \ne k_2$, in its expression. Thus, we aim at replacing $I_{(k_1,k_2)}(\Delta)$ in (7) by some tractable random variables $\{\xi_{k_1 k_2,\Delta}\}_{k_1,k_2}$ so that a scheme using $\xi_{k_1 k_2,\Delta}$ still achieves local weak third order convergence. Lemma 2.1.5 in Milstein and Tretyakov (2021) states the local weak third order accuracy is preserved if the random variables $\{\xi_{k_1 k_2,\Delta}\}_{k_1,k_2}$ have finite moments up to 6th order and satisfy a collection of moment conditions.

Thus, we write the sampling scheme for elliptic diffusions as:

$$\bar{X}^x_{e,\Delta} = x + V_{R,0}(x,\theta)\Delta + \sum_{1\le k\le d_R} V_{R,k}(x,\theta) B_{k,\Delta} + \sum_{0\le k_1,k_2\le d_R} \hat{V}_{k_1} V_{R,k_2}(x,\theta)\, \xi_{k_1 k_2,\Delta}, \tag{8}$$

with the following specification of the random variables $\xi_{k_1 k_2,\Delta}$, $0 \le k_1, k_2 \le d_R$, satisfying the moment conditions given in Lemma 2.1.5 Milstein and Tretyakov (2021): for $1 \le k_1, k_2 \le d_R$,

$$\begin{aligned} &\xi_{00,\Delta} = \tfrac{\Delta^2}{2}, \qquad \xi_{k_1 0,\Delta} = \xi_{0 k_1,\Delta} = \tfrac{1}{2} B_{k_1,\Delta}\Delta, \qquad \xi_{k_1 k_1,\Delta} = \tfrac{1}{2} B_{k_1,\Delta} B_{k_1,\Delta} - \tfrac{1}{2}\Delta; \\ &\xi_{k_1 k_2,t} = \tfrac{1}{2} B_{k_1,\Delta} B_{k_2,\Delta} + \tfrac{1}{2} B_{k_1,\Delta} \widetilde{B}_{k_2,\Delta} \cdot \mathbf{1}_{k_1<k_2} - \tfrac{1}{2} B_{k_2,\Delta} \widetilde{B}_{k_1,\Delta} \cdot \mathbf{1}_{k_1>k_2}. \end{aligned} \tag{9}$$

Here, $\widetilde{B} := (\widetilde{B}_{2,t},\ldots,\widetilde{B}_{d_R,t})_{t\ge 0}$ denotes a (d_R-1)-dimensional standard Brownian motion, independent of B. Our scheme is marginally different from the one appearing in Milstein and Tretyakov (2021), with the former scheme using $(2d_R - 1)$ Gaussian variates, rather than $2d_R$ ones involved in the latter. The scheme is explicit as it only involves Gaussian variates – thus Criterion (i) stated above is satisfied. Criterion (ii) is satisfied, given the above discussion. Criterion (iii) is clearly satisfied under (H3).

Remark 2.1. Scheme (8) under the choice of variates in (9) can also be used in a hypo-elliptic setting as an explicit weak second order scheme. However, such a specification fails to meet Criterion (iii). Some more details are given in Remark 2.5. Hence, in this work, we propose separate sampling schemes for the elliptic/hypo-elliptic classes.

In (8) we introduced the extra subscript 'e' for sampling scheme $\bar{X}^x_{e,\Delta}$ to stress that it involves the elliptic case (not the hypo-elliptic one). We adopt a similar convention for mathematical expressions, when needed, in the rest of the paper.

2.2.2. *Sampling scheme for hypo-elliptic diffusions*

We treat hypo-elliptic SDEs, thus $d_S \ge 1$. We make use of an Itô-Taylor expansion where for the rough component we retain all second order integrals and for the smooth one we retain the second order and third order integrals for which the outside integrator is dt. That is, we have $X^x_\Delta = \hat{X}^x_\Delta + \rho_\Delta(x,\theta)$ where:

$$\begin{aligned} \hat{X}^x_{R,\Delta} &= x_R + V_{R,0}(x,\theta)\Delta + \sum_{1\le k\le d_R} V_{R,k}(x,\theta)\, B_{k,\Delta} + \sum_{0\le k_1,k_2\le d_R} \hat{V}_{k_1} V_{R,k_2}(x,\theta)\, I_{(k_1,k_2)}(\Delta); \\ \hat{X}^x_{S,\Delta} &= x_S + V_{S,0}(x,\theta)\,\Delta + \sum_{0\le k\le d_R} \hat{V}_k V_{S,0}(x,\theta)\, I_{(k,0)}(\Delta) + \sum_{\substack{0\le k_1,k_2\le d_R \\ \text{s.t. } k_1=k_2\neq 0}} \hat{V}_{k_1}\hat{V}_{k_2} V_{S,0}(x,\theta)\, I_{(k_1,k_2,0)}(\Delta). \end{aligned}$$

As with the elliptic case, we next replace the non-explicit iterated integrals $I_{(k_1,k_2)}(\Delta)$, $I_{(k_1,k_2,0)}(\Delta)$, $1 \le k_1, k_2 \le d_R$, $k_1 \ne k_2$, with explicit variates, based on moment conditions that ensure that the resulted scheme remains a weak second order one. The tools in Milstein and Tretyakov (2021) cover only the setting of double integrals, thus we need to carry out an extension of such methodology in the presence of triple integrals. The extension (see the proof of Proposition 2.3 stated below for details) gives rise to the sampling scheme $\bar{X}^x_\Delta = [(\bar{X}^x_{R,\Delta})^\top, (\bar{X}^x_{S,\Delta})^\top]^\top$, $x = [x_R^\top, x_S^\top]^\top \in \mathbb{R}^d$ for hypo-elliptic SDEs determined as follows:

$$
\begin{aligned}
\bar{X}^x_{R,\Delta} &= x_R + V_{R,0}(x,\theta)\Delta + \sum_{1\le k\le d_R} V_{R,k}(x,\theta)\, B_{k,\Delta} + \sum_{0\le k_1,k_2\le d_R} \hat{V}_{k_1} V_{R,k_2}(x,\theta)\, \zeta_{k_1 k_2,\Delta};\\
\bar{X}^x_{S,\Delta} &= x_S + V_{S,0}(x,\theta)\,\Delta + \sum_{0\le k\le d_R} \hat{V}_k V_{S,0}(x,\theta)\, \zeta_{k0,\Delta} + \sum_{\substack{0\le k_1,k_2\le d_R\\ \text{s.t. } k_1=k_2\neq 0}} \hat{V}_{k_1}\hat{V}_{k_2} V_{S,0}(x,\theta)\, \eta_{k_1k_2,\Delta}.
\end{aligned}
\tag{10}
$$

Variables $\zeta_{k_1k_2,\Delta}$ are such that $\zeta_{00,\Delta} = \Delta^2/2$ and for $1 \le k_1, k_2 \le d_R$:

$$
\zeta_{0k_1,\Delta} = I_{(0,k_1)}(\Delta), \quad \zeta_{k_10,\Delta} = I_{(k_1,0)}(\Delta), \quad \zeta_{k_1k_2,\Delta} = \xi_{k_1k_2,\Delta}, \tag{11}
$$

for ξ's as determined earlier in (9) for the elliptic case. Variables $\eta_{k_1k_2,t}$ are required to satisfy the following moment conditions, with $1 \le k_1, k_2, k_3, k_4 \le d_R$:

$$
\begin{aligned}
&\mathbb{E}[\eta_{k_1k_2,\Delta}] = \mathbb{E}[I_{(k_1,k_2,0)}(\Delta)] = 0, \quad \mathbb{E}[\eta_{k_10,\Delta}] = \mathbb{E}[I_{(k_1,0,0)}(\Delta)] = 0;\\
&\mathbb{E}[\eta_{0k_1,\Delta}] = \mathbb{E}[I_{(0,k_1,0)}(\Delta)] = 0, \quad \mathbb{E}[\eta_{k_1k_2,\Delta}B_{k_3,\Delta}] = \mathbb{E}[I_{(k_1,k_2,0)}(\Delta)B_{k_3,\Delta}] = 0;\\
&\mathbb{E}[\eta_{k_10,\Delta}B_{k_2,\Delta}] = \mathbb{E}[\eta_{0k_1,\Delta}B_{k_2,\Delta}] = \mathbb{E}[I_{(k_1,0,0)}(\Delta)B_{k_2,\Delta}] = \mathbb{E}[I_{(0,k_1,0)}(\Delta)B_{k_2,\Delta}] = \tfrac{\Delta^3}{6} \times \mathbf{1}_{k_1=k_2};\\
&\mathbb{E}[\eta_{k_10,\Delta}\zeta_{k_20,\Delta}] = \mathbb{E}[I_{(k_1,0,0)}(\Delta)I_{(k_2,0)}(\Delta)] = \tfrac{\Delta^4}{8} \times \mathbf{1}_{k_1=k_2};\\
&\mathbb{E}[\eta_{0k_1,\Delta}\zeta_{k_20,\Delta}] = \mathbb{E}[I_{(0,k_1,0)}(\Delta)I_{(k_2,0)}(\Delta)] = \tfrac{\Delta^4}{6} \times \mathbf{1}_{k_1=k_2};\\
&\mathbb{E}[\eta_{k_1k_2,\Delta}\zeta_{k_3k_4,\Delta}] = \mathbb{E}[I_{(k_1,k_2,0)}(\Delta)I_{(k_3,k_4)}(\Delta)] = \tfrac{\Delta^3}{6} \times \mathbf{1}_{k_1=k_3,k_2=k_4};\\
&\mathbb{E}[\eta_{k_1k_2,\Delta}\eta_{k_3k_4,\Delta}] = \mathbb{E}[I_{(k_1,k_2,0)}(\Delta)I_{(k_3,k_4,0)}(\Delta)] = \tfrac{\Delta^4}{12} \times \mathbf{1}_{k_1=k_3,k_2=k_4}.
\end{aligned}
$$

A particular choice for the η's that we adopt for the rest of the paper is the one below:

$$
\eta_{k_10,\Delta} = \tfrac{\Delta}{2}\zeta_{k_10,\Delta} - \tfrac{\Delta^2}{12}B_{k_1,\Delta}, \;\; \eta_{0k_1,\Delta} = \Delta\zeta_{k_10,\Delta} - \tfrac{\Delta^2}{3}B_{k_1,\Delta}, \;\; \eta_{k_1k_2,\Delta} = \tfrac{1}{3}\zeta_{k_1k_2,\Delta}\Delta - \widetilde{\eta}_{k_1k_2,\Delta}\Delta.
$$

We have set:

$$
\widetilde{\eta}_{k_1k_2,\Delta} = \begin{cases} \frac{1}{6\sqrt{2}}(\widetilde{B}_{k,\Delta}\widetilde{B}_{k,\Delta} - \Delta), & k_1 = k_2 = k;\\ \frac{1}{6\sqrt{2}}(\widetilde{B}_{k_1,\Delta}\widetilde{B}_{k_2,\Delta} + \widetilde{B}_{k_1,\Delta}W_{k_2,\Delta}\cdot \mathbf{1}_{k_1<k_2} - \widetilde{B}_{k_2,\Delta}W_{k_1,\Delta}\cdot\mathbf{1}_{k_1>k_2}) & k_1 \neq k_2; \end{cases}
$$

where $\widetilde{B} := (\widetilde{B}_{1,t},\ldots,\widetilde{B}_{d_R,t})_{t\ge 0}$ and $W = (W_{2,t},\ldots,W_{d_R,t})_{t\ge0}$ are standard Brownian motions, mutually independent and independent of B.

Remark 2.2. Generation of the Gaussian variables $\zeta_{0k,\Delta}$ and $\zeta_{k0,\Delta} = B_{k,\Delta}\Delta - \zeta_{0k,\Delta}$ in (11) must take under consideration the dependency structure amongst $\{B_{k,\Delta}\}$, $\{\zeta_{0k,\Delta}\}$. That is, we have: $\mathbb{E}[B_{k_1,\Delta}\zeta_{0k_2,\Delta}] = \Delta^2/2\cdot \mathbf{1}_{k_1=k_2}$, $\mathbb{E}[\zeta_{0k_1,\Delta}\zeta_{0k_2,\Delta}] = \Delta^3/3\cdot\mathbf{1}_{k_1=k_2}$ for $1 \le k_1, k_2 \le d_R$. Thus, one can generate $\{B_{k,\Delta}\}$, $\{\zeta_{0k,\Delta}\}$ and $\{\zeta_{k0,\Delta}\}$ as follows:

$$
B_{k,\Delta} = \Delta^{1/2} Z_k, \quad \zeta_{0k,\Delta} = \tfrac{\Delta^{3/2}}{2}\big(Z_k + \tfrac{1}{\sqrt{3}}\widetilde{Z}_k\big), \quad \zeta_{k0,\Delta} = \Delta^{3/2}Z_k - \zeta_{0k,\Delta}, \qquad 1 \le k \le d_R,
$$

where $Z_k, \widetilde{Z}_k \sim \mathscr{N}(0,1)$, $1 \le k \le d_R$, are i.i.d. random variables.

We have developed scheme (10) that is explicit (Criterion (i)) and, as stated in the proposition below, it has local weak third order accuracy (Criterion (ii)), with the proof given in Section D of Supplementary Material (Iguchi, Beskos and Graham, 2024). Criterion (iii) is discussed after the following proposition.

Proposition 2.3. *Let $d_S \geq 1$ and consider $\varphi \in C_p^{\infty}(\mathbb{R}^d;\mathbb{R})$. Under conditions (H1)–(H2), for any $(x,\theta) \in \mathbb{R}^d \times \Theta$, there exist constants $C > 0$, $q \geq 1$ such that:*

$$\left|\mathbb{E}_\theta[\varphi(X_\Delta^x)] - \mathbb{E}_\theta[\varphi(\bar{X}_\Delta^x)]\right| \leq C(1+|x|^q)\Delta^3.$$

Finally, w.r.t Criterion (iii) we notice that scheme (10) contains in its specification (i.e. once some terms are removed from (10)) the *local Gaussian scheme* $\widetilde{X}_\Delta^{\mathrm{LG},x} = \left[(\widetilde{X}_{R,\Delta}^x)^\top, (\widetilde{X}_{S,\Delta}^x)^\top\right]^\top$ given as:

$$\begin{aligned}
\widetilde{X}_{R,\Delta}^x &= x_R + V_{R,0}(x,\beta)\Delta + \sum_{1\leq k\leq d_R} V_{R,k}(x,\sigma)B_{k,\Delta};\\
\widetilde{X}_{S,\Delta}^x &= x_S + V_{S,0}(x,\gamma)\Delta + \hat{V}_0 V_{S,0}(x,\theta)\tfrac{\Delta^2}{2} + \sum_{1\leq k\leq d_R} \hat{V}_k V_{S,0}(x,\theta) I_{(k,0)}(\Delta),
\end{aligned} \tag{12}$$

where the covariance matrix is given as:

$$\Sigma(\Delta,x;\theta) = \begin{bmatrix} \sum_{1\leq k\leq d_R} V_{R,k}(x,\sigma)V_{R,k}(x,\sigma)^\top \Delta & \sum_{1\leq k\leq d_R} V_{R,k}(x,\sigma)\hat{V}_k V_{S,0}(x,\theta)^\top \frac{\Delta^2}{2} \\ \sum_{1\leq k\leq d_R} \hat{V}_k V_{S,0}(x,\theta)V_{R,k}(x,\sigma)^\top \frac{\Delta^2}{2} & \sum_{1\leq k\leq d_R} \hat{V}_k V_{S,0}(x,\theta)\hat{V}_k V_{S,0}(x,\theta)^\top \frac{\Delta^3}{3} \end{bmatrix}. \tag{13}$$

$\Sigma(\Delta,x;\theta)$ is positive definite for any $(\Delta,x,\theta) \in (0,\infty)\times\mathbb{R}^d\times\Theta$ under Assumptions (H3)–(H4), thus Criterion (iii) is satisfied.

Remark 2.4. Proposition 2.3 holds for scheme (10) with or without retainment of the triple integral term in the smooth component $I_3 := \sum_{(k_1,k_2)\in\{0,\ldots,d_R\}^2\setminus\{0,0\}} \hat{V}_{k_1}\hat{V}_{k_2}V_{S,0}(x,\theta)\, I_{(k_1,k_2,0)}(\Delta)$. However, our analysis of statistical methodology based on the above scheme will illustrate that inclusion of term I_3 (or, more precisely, of an explicit substitute for I_3) is necessary for obtaining a CLT for the MLE in the complete observation regime under the improved rate $\Delta_n = o(n^{-1/3})$.

Remark 2.5. A difference versus the scheme for elliptic SDEs in (8) is that in (10) variables $I_{(k,0)}(\Delta)$, $I_{(0,k)}(\Delta)$, $1\leq k\leq d_R$, are kept and are not replaced by $\frac{1}{2}B_{k,\Delta}\Delta$. A replacement of $I_{(k,0)}(\Delta)$ by $\frac{1}{2}B_{k,\Delta}\Delta$ would lead to a degenerate covariance matrix in place of $\Sigma(\Delta,x;\theta)$ and a violation of Criterion (iii).

3. Small time density expansion

We denote the transition densities of the given SDE (1) and the sampling scheme $\bar{X}_\Delta^x$ by

$$y \mapsto p_\Delta^X(x,y;\theta) = \mathbb{P}_\theta[X_\Delta^x \in dy]/dy, \qquad y \mapsto p_\Delta^{\bar{X}}(x,y;\theta) = \mathbb{P}_\theta[\bar{X}_\Delta^x \in dy]/dy$$

respectively, for $(\Delta,x,\theta) \in (0,\infty)\times\mathbb{R}^d\times\Theta$. The (one step) sampling scheme $\bar{X}_\Delta^x$ in (8) or (10), for elliptic or hypo-elliptic SDEs respectively, does not admit, in general, a closed-form density function. To obtain a proxy for the density, we work with a small time expansion of the density of $\bar{X}_\Delta^x$ given below in Lemma 3.1. The expansion will be called upon in Section 4 to provide a method for statistical inference that will possess advantageous characteristics compared with existing methods that use constrast functions induced by conditionally Gaussian discretisation schemes, e.g. the local Gaussian scheme (12) in the hypo-elliptic case or the Euler-Maruyama scheme in the elliptic case. We proceed by assuming a hypo-elliptic setting with $d_S \geq 1$, but our analysis also covers the elliptic case $d_S = 0$.

3.1. Background material

Before giving the small time expansion formula for $p^{\bar{X}}_{\Delta}(x, y; \theta)$ we sketch its derivation via use of Malliavin calculus, a differential calculus on Wiener space. Full rigorous arguments are given in Section E of Supplementary Material (Iguchi, Beskos and Graham, 2024). We refer interested readers to, e.g., Ikeda and Watanabe (2014), Watanabe (1987).

We consider Wiener functionals, $F = F(\omega) : \Omega \to \mathbb{R}^d$, where for current Section 3.1, $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ is the probability space giving rise to the $3 \times d_R - 1$ Brownian motions involved in the definition of sampling scheme $\bar{X}_{\Delta}$ in (10). Malliavin calculus allows for $\delta_y(F)$, with δ_y the Dirac measure at $y \in \mathbb{R}^d$, to be well-defined as an element of a Sobolev space of Wiener functionals, provided that F satisfies regularity and non-degeneracy conditions, in which case such a Wiener functional is referred to as *smooth and non-degenerate in the Malliavin sense*. The non-degeneracy is a sufficient condition for the existence of Lebesgue density for the law of F, and in the case of a Gaussian variate, non-degeneracy is equivalent to the positive definiteness of the covariance matrix. It is shown that the Lebesgue density, $p^F(y)$, of F coincides with $\mathbb{E}\,[\,\delta_y(F)\,]$, i.e. the *generalised expectation* of $\delta_y(F)$, if F is smooth and non-generate in the Malliavin sense. We consider an $F = F^{\varepsilon}$, for small $\varepsilon \in (0, 1)$, of the form

$$F^{\varepsilon} = f_0 + \varepsilon f_1 + \varepsilon^2 f_2 + \cdots + \varepsilon^J f_J \tag{14}$$

for Wiener functionals $f_i = f_i(\omega)$, $0 \le i \le J$, $J \ge 1$. We assume that F^{ε} is smooth and non-degenerate in the Malliavin sense for any $\varepsilon \in (0, 1)$. Then, Theorem 9.4 in Ikeda and Watanabe (2014) gives an asymptotic expansion of the density, $p^{F^{\varepsilon}}(y)$, of F^{ε} as follows. For every $n \ge 0$, there exist Wiener functionals $\phi_k = \phi_k(\omega, y)$, $0 \le k \le n-1$, and $\mathbb{R}$-valued $r_n(\cdot)$ such that

$$p^{F^{\varepsilon}}(y) = \mathbb{E}\,[\,\delta_y(F^{\varepsilon})\,] = \sum_{0 \le k \le n-1} \varepsilon^k \cdot \mathbb{E}\,[\,\phi_k(\cdot, y)\,] + \varepsilon^n \cdot r_n(y, \varepsilon), \tag{15}$$

with $|r_n(y, \varepsilon)| \le C$ for constant $C > 0$ independent of $\varepsilon \in (0, 1)$, $y \in \mathbb{R}^d$, and the ϕ_k's are given via a formal Taylor expansion of $\varepsilon \to \delta_y(F^{\varepsilon})$, e.g., $\phi_0 = \delta_y(f_0)$, $\phi_1 = \langle \partial_z \delta_y(z)|_{z=f_0}, f_1 \rangle$.

3.2. Density expansion for sampling scheme

We develop a small time expansion for the density $p^{\bar{X}}_{\Delta}(x, y; \theta)$ of $\bar{X}^x_{\Delta}$ in (10) via reference to (14), (15). We make use of the multi-index notation: $\|\alpha\| := l + (\#\text{ of zeros in } \alpha)$, $|\alpha| := l$, for $\alpha \in \{0, 1, \dots, d\}^l$, $l \ge 0$. Based on the expression for $\bar{X}^x_{\Delta}$ in (10), we define $\bar{X}^{\varepsilon,x}_{\Delta} = \big[\,(\bar{X}^{\varepsilon,x}_{R,\Delta})^{\top}, (\bar{X}^{\varepsilon,x}_{S,\Delta})^{\top}\,\big]^{\top}$ as follows:

$$\begin{aligned}
\bar{X}^{\varepsilon,x}_{R,\Delta} &= x_R + \varepsilon^2 V_{R,0}(x, \beta)\Delta + \varepsilon \sum_{1 \le k \le d_R} V_{R,k}(x, \sigma)\, B_{k,\Delta} + \sum_{0 \le k_1, k_2 \le d_R} \varepsilon^{\|(k_1,k_2)\|} \hat{V}_{k_1} V_{R,k_2}(x, \theta)\, \zeta_{k_1 k_2,\Delta}; \\
\bar{X}^{\varepsilon,x}_{S,\Delta} &= x_S + \varepsilon^2 V_{S,0}(x, \gamma)\,\Delta + \sum_{0 \le k \le d_R} \varepsilon^{\|(k,0)\|} \hat{V}_k V_{S,0}(x, \theta)\, \zeta_{k0,\Delta} \\
&\qquad + \sum_{\substack{0 \le k_1, k_2 \le d_R \\ \text{s.t. } k_1 = k_2 \neq 0}} \varepsilon^{\|(k_1,k_2,0)\|} \hat{V}_{k_1} \hat{V}_{k_2} V_{S,0}(x, \theta)\, \eta_{k_1 k_2,\Delta}.
\end{aligned} \tag{16}$$

The rationale in the consideration of the above process is that, first, the latter is connected with scheme $\bar{X}^x_{\Delta}$ in (10) via the equality in distributions $\mathbb{P}_{\theta}\big[\bar{X}^x_{\Delta} \in dy\big] = \mathbb{P}_{\theta}\big[\bar{X}^{\sqrt{\Delta},x}_1 \in dy\big]$ and, second, it is straightforward to perform the standardisation defined below upon process (16). We cannot apply expansion

(15) on the density of $\bar{X}_1^{\varepsilon,x}$ as $\{\bar{X}_1^{\varepsilon,x}\}_\varepsilon$ is degenerate in the Malliavin sense as $\varepsilon \to 0$, e.g. notice that $\bar{X}_1^{0,x} = x$ is deterministic. Instead, we introduce $\widetilde{Y}^\varepsilon$ via an appropriate standarisation of $\bar{X}_1^{\varepsilon,x}$ based upon the local Gaussian scheme $\widetilde{X}_\Delta^{\mathrm{LG},x} = \big[(\widetilde{X}_{R,\Delta}^x)^\top, (\widetilde{X}_{S,\Delta}^x)^\top\big]^\top$ in (12). Thus, the small time expansion ultimately obtained will be centred around the density of the local Gaussian scheme. In particular, we set

$$m_{x,\theta,\varepsilon}(y) := \begin{bmatrix} m_{R,x,\theta,\varepsilon}(y_R) \\ m_{S,x,\theta,\varepsilon}(y_S) \end{bmatrix} = \begin{bmatrix} \frac{y_R - x_R - \varepsilon^2 V_{R,0}(x,\beta)}{\varepsilon} \\ \frac{y_S - x_S - \varepsilon^2 V_{S,0}(x,\gamma) - \frac{\varepsilon^4}{2}\hat{V}_0 V_{S,0}(x,\theta)}{\varepsilon^3} \end{bmatrix}, \tag{17}$$

and, define

$$\widetilde{Y}^\varepsilon = \big[(\widetilde{Y}_R^\varepsilon)^\top, (\widetilde{Y}_S^\varepsilon)^\top\big]^\top = m_{x,\theta,\varepsilon}(\bar{X}_1^{\varepsilon,x}) = \widetilde{Y}^{(0)} + \sum_{1\le l\le 3} \varepsilon^l \cdot \widetilde{Y}^{(l)}, \tag{18}$$

where one can obtain the $O(1)$-term in (18), $\widetilde{Y}^0 = [(\widetilde{Y}_R^{(0)})^\top, (\widetilde{Y}_S^{(0)})^\top]^\top$, as

$$\widetilde{Y}_R^{(0)} = \sum_{1\le k\le d_R} V_{R,k}(x,\sigma) B_{k,1}, \qquad \widetilde{Y}_S^{(0)} = \sum_{1\le k\le d_R} \hat{V}_k V_{S,0}(x,\theta) I_{(k,0)}(1),$$

with the remaining terms $\widetilde{Y}^{(l)}$, $1 \le l \le 3$, explicitly derived via (16)–(18). $\widetilde{Y}^{(0)}$ follows a Gaussian distribution and is non-degenerate under (H3)–(H4), as the covariance matrix $\Sigma(1,x;\theta)$ in (13) is positive definite. In brief, the above lead to non-degeneracy in the Malliavin sense of $\{\widetilde{Y}^\varepsilon\}_\varepsilon$, uniformly in $\varepsilon \in (0,1)$. Simple change of variables with $\varepsilon = \sqrt{\Delta}$ yields:

$$p_\Delta^{\bar{X}}(x,y;\theta) = \mathbb{P}_\theta\big[\bar{X}_1^{\sqrt{\Delta},x} \in dy\big]/dy = \frac{1}{\sqrt{\Delta^{d_R+3d_S}}}\, p^{\widetilde{Y}^{\sqrt{\Delta}}}(\xi;\theta)|_{\xi = m_{x,\theta,\sqrt{\Delta}}(y)}. \tag{19}$$

Thus, application of expansion (15) for the density of $\widetilde{Y}^\varepsilon$, $p^{\widetilde{Y}^\varepsilon}(\xi;\theta) = \mathbb{E}\,[\,\delta_\xi(\widetilde{Y}^\varepsilon)\,]$, produces the small time expansion of $p_\Delta^{\bar{X}}(x,y;\theta)$. Due to starting off from (17) the rightmost side of (19) expands around the density of the local Gaussian scheme (12), that is:

$$\frac{1}{\sqrt{\Delta^{d_R+3d_S}}} \times p^{\widetilde{Y}^{(0)}}(\xi;\theta)\big|_{\xi = m_{x,\theta,\sqrt{\Delta}}(y)} = p_\Delta^{\widetilde{X}^{\mathrm{LG}}}(x,y;\theta), \tag{20}$$

where $\xi \mapsto p^{\widetilde{Y}^{(0)}}(\xi;\theta)$ denotes the density of the probability law of $\widetilde{Y}^{(0)}$ and $y \mapsto p_\Delta^{\widetilde{X}^{\mathrm{LG}}}(x,y;\theta)$ denotes the density of the local Gaussian scheme $\widetilde{X}_\Delta^{\mathrm{LG},x}$ defined in (12) and given as:

$$p_\Delta^{\widetilde{X}^{\mathrm{LG}}}(x,y;\theta) = \frac{1}{\sqrt{(2\pi)^d |\Sigma(\Delta,x;\theta)|}} \exp\Big(-\tfrac{1}{2}\big(y-\mu(\Delta,x;\theta)\big)^\top \Sigma^{-1}(\Delta,x;\theta)\big(y-\mu(\Delta,x;\theta)\big)\Big), \tag{21}$$

for mean vector:

$$\mu(\Delta,x;\theta) = \begin{bmatrix} \mu_R(\Delta,x;\beta) \\ \mu_S(\Delta,x;\theta) \end{bmatrix} = \begin{bmatrix} x_R + V_{R,0}(x,\beta)\Delta \\ x_S + V_{S,0}(x,\gamma)\Delta + \hat{V}_0 V_{S,0}(x,\theta)\frac{\Delta^2}{2} \end{bmatrix}, \tag{22}$$

and covariance matrix $\Sigma(\Delta,x;\theta)$ given in (13).

To state the asymptotic expansion of $p_\Delta^{\bar{X}}(x,y;\theta)$, we introduce a class of Hermite polynomials based on the density of $\widetilde{Y}^{(0)}$. To simplify the notation we henceforth write $\widetilde{Y} = \widetilde{Y}^{(0)}$. Thus, we define:

$$\mathcal{H}_\alpha^{\widetilde{Y}}(\xi;\theta) := (-1)^{|\alpha|}\, \partial_\alpha^\xi\, p^{\widetilde{Y}}(\xi;\theta)/p^{\widetilde{Y}}(\xi;\theta), \qquad \xi \in \mathbb{R}^d,\ \theta \in \Theta. \tag{23}$$

Also, we write $\mathcal{H}_\alpha(\Delta, x, y; \theta) = \mathcal{H}_\alpha^{\widetilde{Y}}(\xi;\theta)|_{\xi = m_{x,\theta,\sqrt{\Delta}}(y)}$, $(\Delta, x, y, \theta) \in (0,\infty) \times \mathbb{R}^d \times \mathbb{R}^d \times \Theta$. We then have the following result whose proof is provided in Section E.2 of Supplementary Material (Iguchi, Beskos and Graham, 2024).

Lemma 3.1. *Let $x, y \in \mathbb{R}^d$, $\Delta > 0$, $\theta \in \Theta$, and assume that conditions (H1)–(H4) hold. Then, for any integer $J \geq 4$, the transition density $y \mapsto p_\Delta^{\bar{X}}(x, y; \theta)$ admits the following representation:*

$$p_\Delta^{\bar{X}}(x,y;\theta) = p_\Delta^{\widetilde{X}^{\mathrm{LG}}}(x,y;\theta)\Big\{1 + \sum_{1 \le l \le J-1} \Delta^{l/2}\, \Psi_l(\Delta,x,y;\theta)\Big\} + \frac{\Delta^{J/2}}{\sqrt{\Delta^{d_R+3d_S}}} R^J(x,y;\theta), \tag{24}$$

where $R^J(x,y;\theta)$ is the residual term satisfying $\sup_{x,y\in\mathbb{R}^d,\, \theta\in\Theta} |R^J(x,y;\theta)| < C$ for some constant $C > 0$, and $\Psi_l(\Delta,x,y;\theta)$, $l \geq 1$, have the general form:

$$\Psi_l(\Delta,x,y;\theta) = \sum_{1 \le k \le \nu(l)} \sum_{\alpha \in \{1,\dots,d\}^k} \nu_\alpha(x,\theta) \times \mathcal{H}_\alpha(\Delta,x,y;\theta), \tag{25}$$

for some positive integer $\nu(l)$, where $\nu_\alpha : \mathbb{R}^d \times \Theta \to \mathbb{R}$ is explicitly given as a summation of products of $V_j(x,\theta)$, $0 \le j \le d_R$, and their partial derivatives. For the first two Ψ-terms in the expansion, we get:

$$\Psi_1(\Delta,x,y;\theta) = \sum_{\substack{1 \le j_1, j_2 \le d_R \\ 1 \le i_1 \le d}} \Big\{ \hat{V}_{j_1} V_{j_2}^{i_1}(x,\theta) \cdot \mathbf{1}_{1 \le i_1 \le d_R} + \tfrac{1}{3} \hat{V}_{j_1} \hat{V}_{j_2} V_0^{i_1}(x,\theta) \cdot \mathbf{1}_{d_R+1 \le i_1 \le d} \Big\} \cdot \widetilde{\Psi}_{j_1 j_2}^{i_1}(\Delta,x,y;\theta),$$

with

$$\begin{aligned}
\widetilde{\Psi}_{j_1 j_2}^{i_1}(\Delta,x,y;\theta) &= \tfrac{1}{2} \sum_{1 \le i_2, i_3 \le d_R} V_{j_1}^{i_2}(x,\theta) V_{j_2}^{i_3}(x,\theta)\, \mathcal{H}_{(i_1,i_2,i_3)}(\Delta,x,y;\theta) \\
&+ \sum_{\substack{d_R+1 \le i_2 \le d \\ 1 \le i_3 \le d_R}} \Big\{ \tfrac{1}{3} \hat{V}_{j_1} V_0^{i_2}(x,\theta) V_{j_2}^{i_3}(x,\theta) + \tfrac{1}{6} V_{j_1}^{i_3}(x,\theta) \hat{V}_{j_2} V_0^{i_2}(x,\theta) \Big\} \mathcal{H}_{(i_1,i_2,i_3)}(\Delta,x,y;\theta) \\
&+ \tfrac{1}{8} \sum_{d_R+1 \le i_2, i_3 \le d} \hat{V}_{j_1} V_0^{i_2}(x,\theta) \hat{V}_{j_2} V_0^{i_3}(x,\theta) \mathcal{H}_{(i_1,i_2,i_3)}(\Delta,x,y;\theta).
\end{aligned}$$

Also, $\Psi_2(\Delta,x,y;\theta) = \Phi_2(\Delta,x,y;\theta) + \widetilde{\Phi}_2(\Delta,x,y;\theta)$ with

$$\begin{aligned}
\Phi_2(\Delta,x,y;\theta) &= \tfrac{1}{2} \sum_{1 \le i_1, i_2 \le d_R} \sum_{1 \le k \le d_R} \Big(\hat{V}_k V_0^{i_1}(x,\theta) + \hat{V}_0 V_k^{i_1}(x,\theta) \Big) V_k^{i_2}(x,\sigma) \mathcal{H}_{(i_1,i_2)}(\Delta,x,y;\theta) \\
&+ \tfrac{1}{4} \sum_{1 \le i_1, i_2 \le d_R} \sum_{1 \le k_1, k_2 \le d_R} \hat{V}_{k_1} V_{k_2}^{i_1}(x,\theta) \hat{V}_{k_1} V_{k_2}^{i_2}(x,\theta) \mathcal{H}_{(i_1,i_2)}(\Delta,x,y;\theta) \\
&+ \sum_{\substack{1 \le i_1 \le d_R \\ d_R+1 \le i_2 \le d}} \sum_{1 \le k \le d_R} \Big(\tfrac{1}{3} \hat{V}_k V_0^{i_1}(x,\theta) + \tfrac{1}{6} \hat{V}_0 V_k^{i_1}(x,\theta) \Big) \hat{V}_k V_0^{i_2}(x,\theta) \mathcal{H}_{(i_1,i_2)}(\Delta,x,y;\theta) \\
&+ \tfrac{1}{6} \sum_{\substack{1 \le i_1 \le d_R \\ d_R+1 \le i_2 \le d}} \sum_{1 \le k \le d_R} V_k^{i_1}(x,\theta) \Big(\hat{V}_0 \hat{V}_k V_0^{i_2}(x,\theta) + \hat{V}_k \hat{V}_0 V_0^{i_2}(x,\theta) \Big) \mathcal{H}_{(i_1,i_2)}(\Delta,x,y;\theta) \\
&+ \tfrac{1}{6} \sum_{\substack{1 \le i_1 \le d_R \\ d_R+1 \le i_2 \le d}} \sum_{1 \le k_1, k_2 \le d_R} \hat{V}_{k_1} V_{k_2}^{i_1}(x,\theta) \hat{V}_{k_1} \hat{V}_{k_2} V_0^{i_2}(x,\theta) \mathcal{H}_{(i_1,i_2)}(\Delta,x,y;\theta)
\end{aligned}$$

$$+\sum_{d_R+1\le i_1,i_2\le d}\sum_{1\le k\le d_R}\hat V_k V_0^{i_1}(x,\theta)\Big(\tfrac{1}{6}\hat V_0\hat V_k V_0^{i_2}(x,\theta)+\tfrac{1}{8}\hat V_k\hat V_0 V_0^{i_2}(x,\theta)\Big)\mathcal{H}_{(i_1,i_2)}(\Delta,x,y;\theta)$$

$$+\tfrac{1}{24}\sum_{d_R+1\le i_1,i_2\le d}\sum_{1\le k_1,k_2\le d_R}\hat V_{k_1}\hat V_{k_2}V_0^{i_1}(x,\theta)\hat V_{k_1}\hat V_{k_2}V_0^{i_2}(x,\theta)\mathcal{H}_{(i_1,i_2)}(\Delta,x,y;\theta); \tag{26}$$

$\widetilde\Phi_2(\Delta,x,y;\theta)$ *involves higher order Hermite polynomials and is given in the form of*

$$\sum_{l=4,6}\sum_{\alpha\in\{1,\dots,d\}^l} w_\alpha(x,\theta)\times\mathcal{H}_\alpha(\Delta,x,y;\theta),$$

where $w_\alpha:\mathbb{R}^d\times\Theta\to\mathbb{R}$ *is of the same structure as the one described above for* v_α.

Remark 3.2. The Hermite polynomials have explicit expressions. E.g., for $1\le i_1,i_2\le d$:

$$\mathcal{H}^{\widetilde Y}_{(i_1)}(\xi;\theta)|_{\xi=m_{x,\theta,\sqrt\Delta}(y)}=\sum_{1\le i\le d}\Sigma^{-1}_{i_1 i}(1,x;\theta)m^i_{x,\theta,\sqrt\Delta}(y);$$

$$\mathcal{H}^{\widetilde Y}_{(i_1,i_2)}(\xi;\theta)|_{\xi=m_{x,\theta,\sqrt\Delta}(y)}=\big\{\mathcal{H}^{\widetilde Y}_{(i_1)}(\xi;\theta)\mathcal{H}^{\widetilde Y}_{(i_2)}(\xi;\theta)\big\}|_{\xi=m_{x,\theta,\sqrt\Delta}(y)}-\Sigma^{-1}_{i_1 i_2}(1,x;\theta). \tag{27}$$

Remark 3.3. We compare our small time density expansion (24) with some closed-form density expansions that have appeared in the literature.

(i) Aït-Sahalia (2002, 2008) provides Hermite series expansions for the density $p^X_\Delta(x,y;\theta)$ of elliptic diffusions. However, the domains for $x,y\in\mathbb{R}^d$ should be restricted to a compact set so that the related remainder term is bounded and convergence to the true density can be justified.
(ii) Li (2013) develops a closed-form expansion formula for $p^X_\Delta(x,y;\theta)$ in the multivariate elliptic case using Malliavin-Watanabe calculus that enables one to obtain a remainder term that is uniformly bounded without imposing strong restrictions on the domain of $x,y\in\mathbb{R}^d$. Our approach also uses Malliavin-Watanabe calculus to obtain a closed-form expansion for $p^{\bar X}_\Delta(x,y;\theta)$ with controllable remainder terms in a hypo-elliptic setting.

4. Analytic results for statistical inference

Making use of the small time density expansion in Lemma 3.1, we provide analytic results about statistical inference procedures under high and low-frequency observation regimes. We do not require all correction terms $\Delta^{l/2}\Psi_l(\Delta,x,y;\theta)$, $l=1,\dots$, appearing in the density expansion, and aim at obtaining statistical benefits with use of such terms when necessary. Our results below identify the parts of the expansion that lead to parameter estimates of improved performance, for each of the two observation regimes. Interestingly, such parts differ between the two regimes.

4.1. High-frequency observation regime

4.1.1. Contrast estimator

We consider the complete observation regime (2). Such a data setting has been studied in the recent works focused on hypo-elliptic SDEs of Ditlevsen and Samson (2019), Gloter and Yoshida (2021), and in numerous earlier studies for the elliptic case, see e.g. Kessler (1997), Uchida and Yoshida (2012). We introduce the following notation:

$$\Sigma_1(x;\theta) := \Sigma(1,x;\theta), \tag{28}$$

where recall that Σ is the covariance matrix of local Gaussian scheme defined in (13). We consider the likelihood of the complete data in (2). We work with the following proxy of $-2\times$ log-likelihood, with $\mu = \mu(\Delta,x;\theta)$, $\Sigma = \Sigma(\Delta,x;\gamma,\sigma)$, $\Sigma_1 = \Sigma_1(x;\gamma,\sigma)$ defined in (22), (13), (28), respectively:

$$\begin{aligned} \ell_{n,\Delta}(\theta) := \sum_{1\le m\le n} &\big(X_{t_m} - \mu(\Delta, X_{t_{m-1}};\theta)\big)^\top \Sigma^{-1}(\Delta, X_{t_{m-1}};\gamma,\sigma)\big(X_{t_m} - \mu(\Delta, X_{t_{m-1}};\theta)\big) \\ &+ \sum_{1\le m\le n} \log|\Sigma_1(X_{t_{m-1}};\gamma,\sigma)| - 2\Delta \sum_{1\le m\le n} \Phi_2(\Delta, X_{t_{m-1}}, X_{t_m};\theta), \end{aligned} \tag{29}$$

where Φ_2 is given in (26), namely, Φ_2 is a term with second order Hermite polynomials appearing in the expression for Ψ_2 in the density expansion (24). We discuss the effect of the correction term Φ_2 in the obtained CLT in Remark 4.4 later in the paper. In the elliptic case, i.e. $d = d_R$, $d_\gamma = 0$, the corresponding contrast function is as in (29) but with functions μ, Σ_1, Σ^{-1}, Φ_2 replaced by μ_R, a_R, a_R^{-1}/Δ, $\Phi_{e,2}$, respectively. We define the contrast estimator $\hat{\theta}_n := (\hat{\beta}_n, \hat{\gamma}_n, \hat{\sigma}_n) = \arg\min_{\theta\in\Theta} \ell_{n,\Delta}(\theta)$ and write the estimator as $\hat{\theta}_{e,n} = (\hat{\beta}_n, \hat{\sigma}_n)$ in the setting of elliptic diffusions.

Remark 4.1. We state a property of term Φ_2 defined in (26). For simplicity, we write $\Phi_2(\Delta,x,y;\theta) = \sum_{1\le i_1,i_2\le d} G_{i_1 i_2}(x;\theta) \times \mathcal{H}_{(i_1,i_2)}(\Delta,x,y;\theta)$, $x,y\in\mathbb{R}^d$ for appropriate functions $G_{i_1i_2}$ obtain via (26). Terms $G_{i_1i_2}(x;\theta)$ also arise when considering the covariance of the normalised vector $m(\Delta,x,y;\theta) = m_{x,\theta,\sqrt{\Delta}}(y)$. That is, one obtains: for $1\le i_1,i_2\le d$,

$$\mathbb{E}_\theta\left[m^{i_1}(\Delta,x,X^x_\Delta;\theta)\, m^{i_2}(\Delta,x,X^x_\Delta;\theta)\right] = \Sigma_{1,i_1i_2}(x;\theta) + 2\Delta\times G_{i_1i_2}(x;\theta) + \widetilde{G}_{i_1i_2}(\Delta^2,x;\theta), \tag{30}$$

for $\widetilde{G}_{i_1i_2}(\cdot,\cdot;\theta) : [0,\infty)\times\mathbb{R}^d \to \mathbb{R}$ such that $|\widetilde{G}_{i_1i_2}(h,x;\theta)| \le Ch$ for some $C>0$, and $\widetilde{G}_{i_1i_2}(h,\cdot;\theta) \in C_b^\infty(\mathbb{R}^d;\mathbb{R})$ for all $(h,\theta)\in[0,\infty)\times\Theta$ under (H1)–(H2). The above connection between Φ_2 and the covariance expression (30) plays a key role to obtain a CLT under the weaker condition $\Delta_n = o(n^{-1/3})$.

4.1.2. Conditions for high-frequency regime

To state the main result, we introduce the following additional conditions.

(H5) For each $x\in\mathbb{R}^d$, $\theta\mapsto V_j(x,\theta)$ is three times differentiable. For any $\alpha\in\{1,\ldots,d_\theta\}^l$, with $l\in\{1,2\}$, the functions $x\mapsto \partial_\alpha^\theta V_j^i(x,\theta)$, $0\le j\le d_R$, $1\le i\le d$, have bounded derivatives of every order uniformly in $\theta\in\Theta$.

(H6) For $1\le i\le d_S$, there exists a function $B^i\in C_b^\infty(\mathbb{R}^d;\mathbb{R})$ such that for any $x\in\mathbb{R}^d$ and $\gamma',\gamma\in\Theta_\gamma$,

$$|V^i_{S,0}(x,\gamma') - V^i_{S,0}(x,\gamma)| \le B^i(x)|\gamma'-\gamma|.$$

(H7) The diffusion process $\{X_t\}_{t\ge0}$ in (1) is ergodic under $\theta=\theta^\dagger$, with invariant distribution $\nu_{\theta^\dagger}$ on $\mathbb{R}^d$. Furthermore, all moments of $\nu_{\theta^\dagger}$ are finite.

(H8) It holds that for all $p\ge1$, $\sup_{t>0}\mathbb{E}_{\theta^\dagger}[|X_t|^p] < \infty$.

(H9) The true parameters lie in the interior of Θ. If it holds

$$V_{R,0}(x,\beta) = V_{R,0}(x,\beta^\dagger),\ \ V_{S,0}(x,\gamma) = V_{S,0}(x,\gamma^\dagger),\ \ V(x,\sigma) = V(x,\sigma^\dagger),$$

for x in a set of probability 1 under $\nu_{\theta^\dagger}$, then $\beta=\beta^\dagger$, $\gamma=\gamma^\dagger$, $\sigma=\sigma^\dagger$.

4.1.3. Asymptotic properties of the contrast estimator

We can now prove that the estimator $\hat{\theta}_n$ has the following asymptotic properties.

Theorem 4.2 (Consistency). *Under conditions (H1)–(H9), it holds that if $n \to \infty$, $\Delta_n \to 0$ and $n\Delta_n \to \infty$, then $\hat{\theta}_n \xrightarrow{\mathbb{P}_{\theta^\dagger}} \theta^\dagger$.*

Theorem 4.3 (Asymptotic normality). *Under conditions (H1)–(H9), it holds that if $n \to \infty$, $\Delta_n \to 0$, $n\Delta_n \to \infty$ and $\Delta_n = o(n^{-1/3})$, then:*

(a) *if $d = d_R$, i.e. in the setting of elliptic diffusions,*

$$\left[\sqrt{n\Delta}\, (\hat{\beta}_n - \beta^\dagger)^\top, \sqrt{n}\, (\hat{\sigma}_n - \sigma^\dagger)^\top \right]^\top \xrightarrow{\mathcal{L}_{\theta^\dagger}} \mathcal{N}\left(0, I_e(\theta^\dagger)^{-1}\right),$$

(b) *if $d > d_R$, i.e. in the setting of hypo-elliptic diffusions,*

$$\left[\sqrt{n\Delta}\, (\hat{\beta}_n - \beta^\dagger)^\top, \sqrt{\tfrac{n}{\Delta}}\, (\hat{\gamma}_n - \gamma^\dagger)^\top, \sqrt{n}\, (\hat{\sigma}_n - \sigma^\dagger)^\top \right]^\top \xrightarrow{\mathcal{L}_{\theta^\dagger}} \mathcal{N}\left(0, I(\theta^\dagger)^{-1}\right).$$

$I_e(\theta^\dagger) = \mathrm{diag}\left[\left(I^{\beta}_{e,ij}(\theta^\dagger)\right)_{1 \le i,j \le d_\beta}, \left(I^{\sigma}_{e,ij}(\theta^\dagger)\right)_{1 \le i,j \le d_\sigma} \right]$ is the asymptotic precision matrix with block matrix elements:

$$I^{\beta}_{e,ij}(\theta^\dagger) = \int \left(\partial_{\beta_i} V_{R,0}(x,\beta^\dagger)^\top a_R^{-1}(x,\sigma^\dagger)\, \partial_{\beta_j} V_{R,0}(x,\beta^\dagger) \right) \nu_{\theta^\dagger}(dx);$$
$$I^{\sigma}_{e,ij}(\theta^\dagger) = \frac{1}{2} \int \mathrm{tr}\left(\partial_{\sigma_i} a_R(x,\sigma^\dagger)\, a_R^{-1}(x,\sigma^\dagger)\, \partial_{\sigma_j} a_R(x,\sigma^\dagger)\, a_R^{-1}(x,\sigma^\dagger) \right) \nu_{\theta^\dagger}(dx).$$

Similarly, the asymptotic precision matrix $I(\theta^\dagger)$ has the block-diagonal structure:

$$I(\theta^\dagger) = \mathrm{diag}\left[\left(I^{\beta}_{ij}(\theta^\dagger)\right)_{1 \le i,j \le d_\beta}, \left(I^{\gamma}_{ij}(\theta^\dagger)\right)_{1 \le i,j \le d_\gamma}, \left(I^{\sigma}_{ij}(\theta^\dagger)\right)_{1 \le i,j \le d_\sigma} \right], \tag{31}$$

with block matrix elements:

$$I^{\beta}_{ij}(\theta^\dagger) = \int \left(\partial_{\beta_i} V_{R,0}(x,\beta^\dagger)^\top \Sigma^{-1}_{1,RR}(x;\sigma^\dagger)\, \partial_{\beta_j} V_{R,0}(x,\beta^\dagger) \right) \nu_{\theta^\dagger}(dx);$$
$$I^{\gamma}_{ij}(\theta^\dagger) = 4 \int \left(\partial_{\gamma_i} V_{S,0}(x,\gamma^\dagger)^\top\, \Sigma^{-1}_{1,SS}(x;\gamma^\dagger,\sigma^\dagger)\, \partial_{\gamma_j} V_{S,0}(x,\gamma^\dagger) \right) \nu_{\theta^\dagger}(dx);$$
$$I^{\sigma}_{ij}(\theta^\dagger) = \frac{1}{2} \int \mathrm{tr}\left(\partial_{\sigma_i} \Sigma_1(x;\gamma^\dagger,\sigma^\dagger)\, \Sigma_1^{-1}(x;\gamma^\dagger,\sigma^\dagger)\, \partial_{\sigma_j} \Sigma_1(x;\gamma^\dagger,\sigma^\dagger)\, \Sigma_1^{-1}(x;\gamma^\dagger,\sigma^\dagger) \right) \nu_{\theta^\dagger}(dx).$$

The proofs of Theorems 4.2 and 4.3 are given in Sections B.1 and B.2, respectively, in Supplementary Material (Iguchi, Beskos and Graham, 2024).

Remark 4.4. Gloter and Yoshida (2021) obtain an MLE using the transition density of the local Gaussian scheme in (12), that is $\widetilde{\theta}_n = \mathrm{argmin}_{\theta \in \Theta} \widetilde{\ell}_{n,\Delta}(\theta)$ where $\widetilde{\ell}_{n,\Delta}(\theta)$ is given by (29) without the last term $-2\Delta \sum_{m=1}^{n} \Phi_2(\Delta, X_{t_{m-1}}, X_{t_m};\theta)$. They prove asymptotic normality for $\widetilde{\theta}_n$ under the conditions $n \to \infty$, $\Delta_n \to 0$, $n\Delta_n \to \infty$ and $\Delta_n = o(n^{-1/2})$. The asymptotic precision they obtain is identical to the one we find here in (31). Thus, addition of the term $-2\Delta \sum_{m=1}^{n} \Phi_2(\Delta, X_{t_{m-1}}, X_{t_m};\theta)$ allows for a weaker experimental design condition under which a CLT holds, even if it does not alter the asymptotic variance.

Remark 4.5. We explain that the proposed estimator $\hat{\theta}_n$ contributes to an improvement mainly in the estimation of diffusion parameter σ when compared with the existing estimator $\widetilde{\theta}_n$. This is clarified by observing the role of the weaker condition $\Delta_n = o(n^{-1/3})$ for the CLT. To prove the CLT, we show:

$$\sum_{1 \le m \le n} \mathbb{E}_{\theta^\dagger} [s(\Delta_n, X_{t_{m-1}}, X_{t_m}; \theta^\dagger) | \mathcal{F}_{t_{m-1}}] \xrightarrow{\mathbb{P}_{\theta^\dagger}} 0, \tag{32}$$

as $n \to \infty$, $\Delta_n \to 0$ and $n\Delta_n \to \infty$, with the aid of $\Delta_n = o(n^{-1/3})$, where $s = s(\cdot;\cdot)$ is determined from the scaled score function as:

$$\Big[\tfrac{1}{\sqrt{n\Delta_n}} \partial_\beta^\top \ell_{n,\Delta}(\theta^\dagger), \ \sqrt{\tfrac{\Delta_n}{n}} \partial_\gamma^\top \ell_{n,\Delta}(\theta^\dagger), \ \tfrac{1}{\sqrt{n}} \partial_\sigma \ell_{n,\Delta}(\theta^\dagger) \Big]^\top \equiv \sum_{1 \le m \le n} s(\Delta_n, X_{t_{m-1}}, X_{t_m}; \theta^\dagger).$$

Since the contrast function (29) splits into a log-Gaussian part (related to the local Gaussian scheme) and a correction part (involving Φ_2), we write the score function as: $s(\Delta_n, X_{t_{m-1}}, X_{t_{m-1}}; \theta^\dagger) \equiv s_{\mathrm{LG}}(\Delta_n, X_{t_{m-1}}, X_{t_{m-1}}; \theta^\dagger) + s_{\Phi_2}(\Delta_n, X_{t_{m-1}}, X_{t_{m-1}}; \theta^\dagger)$. Then, the expansion formula (30) yields:

$$\sum_{1 \le m \le n} \mathbb{E}_{\theta^\dagger} [s_{\mathrm{LG}}^{i_1}(\Delta_n, X_{t_{m-1}}, X_{t_m}; \theta^\dagger) | \mathcal{F}_{t_{m-1}}] = \tfrac{1}{n} \sum_{1 \le m \le n} g^{i_1}\big(\sqrt{n\Delta_n^3}, X_{t_{m-1}}; \theta^\dagger\big), \quad 1 \le i_1 \le d_\beta + d_\gamma;$$

$$\begin{aligned} \sum_{1 \le m \le n} \mathbb{E}_{\theta^\dagger} [s_{\mathrm{LG}}^{i_2}(\Delta_n, X_{t_{m-1}}, X_{t_m}; \theta^\dagger) | \mathcal{F}_{t_{m-1}}] &= \tfrac{\sqrt{n\Delta_n^2}}{n} \sum_{1 \le m \le n} \mathrm{tr}\big(2 \cdot G(X_{t_{m-1}}; \theta^\dagger) \partial_{\theta_{i_2}} \Sigma_1^{-1}(X_{t_{m-1}}; \gamma^\dagger, \sigma^\dagger)\big) \\ &+ \tfrac{1}{n} \sum_{1 \le m \le n} g^{i_2}\big(\sqrt{n\Delta_n^4}, X_{t_{m-1}}; \theta^\dagger\big), \quad d_\beta + d_\gamma + 1 \le i_2 \le d_\theta, \end{aligned} \tag{33}$$

where $g^i(\cdot,\cdot;\theta^\dagger) : [0,\infty) \times \mathbb{R}^d \to \mathbb{R}$, $1 \le i \le d$, has the same regularity property as $\widetilde{G}_{i_1 i_2}(\cdot,\cdot;\theta^\dagger)$ in (30). The above two equations correspond to the score functions of the drift parameters $(\beta,\gamma) \in \Theta_\beta \times \Theta_\gamma$ and the diffusion parameter $\sigma \in \Theta_\sigma$, respectively. Also for the correction part, we show that

$$\sum_{1 \le m \le n} \mathbb{E}_{\theta^\dagger} [s_{\Phi_2}^{i_1}(\Delta_n, X_{t_{m-1}}, X_{t_m}; \theta^\dagger) | \mathcal{F}_{t_{m-1}}] = \tfrac{1}{n} \sum_{1 \le m \le n} \widetilde{g}^{i_1}\big(\sqrt{n\Delta_n^3}, X_{t_{m-1}}; \theta^\dagger\big), \quad 1 \le i_1 \le d_\beta + d_\gamma;$$

$$\begin{aligned} \sum_{1 \le m \le n} \mathbb{E}_{\theta^\dagger} [s_{\Phi_2}^{i_2}(\Delta_n, X_{t_{m-1}}, X_{t_m}; \theta^\dagger) | \mathcal{F}_{t_{m-1}}] &= -\tfrac{\sqrt{n\Delta_n^2}}{n} \sum_{1 \le m \le n} \mathrm{tr}\big(2 \cdot G(X_{t_{m-1}}; \theta^\dagger) \partial_{\theta_{i_2}} \Sigma_1^{-1}(X_{t_{m-1}}; \gamma^\dagger, \sigma^\dagger)\big) \\ &+ \tfrac{1}{n} \sum_{1 \le m \le n} \widetilde{g}^{i_2}\big(\sqrt{n\Delta_n^4}, X_{t_{m-1}}; \theta^\dagger\big), \quad d_\beta + d_\gamma + 1 \le i_2 \le d_\theta, \end{aligned}$$

where $\widetilde{g}^i$ has the same structure as g^i. Thus, inclusion of the Φ_2-term in our contrast function $\ell_{n,\Delta}(\theta)$ in (29), results in the cancellation of a quantity corresponding to the 1st term of size $O(\sqrt{n\Delta_n^2})$ on the right side of (33), and then convergence (32) holds under $\Delta_n = o(n^{-1/3})$. For details, see the proof of Lemma B.5 in Section C.3.1 in Supplementary Material (Iguchi, Beskos and Graham, 2024). Note that the required condition for Δ_n is weakened for the score w.r.t. diffusion parameter, while the condition for the score w.r.t. drift parameters remains the same after inclusion of the correction part (see g^{i_1} and $\widetilde{g}^{i_1}$). Thus, it is expected that the new estimator performs better than the existing estimator $\widetilde{\theta}_n$ in the estimation of diffusion parameter.

4.1.4. Contrast estimator for stochastic damping Hamiltonian systems

We focus on an important sub-class of hypo-elliptic diffusions used in applications, namely the *stochastic damping Hamiltonian systems*, and write down in detail the form of the proposed contrast function for such a family of models. An SDE in the above class can be written as follows:

$$\begin{aligned} dX_{R,t} &= -(c_\beta(X_{S,t})X_{R,t} + g_\beta(X_{S,t}))dt + \Xi(\sigma)\, dB_t, \quad X_{R,0} = x_{R,0} \in \mathbb{R}^{\bar{d}}; \\ dX_{S,t} &= X_{R,t}\,dt, \quad X_{S,0} = x_{S,0} \in \mathbb{R}^{\bar{d}}, \end{aligned} \tag{34}$$

where $d_S = d_R = \bar{d} \geq 1$, $g_\beta : \mathbb{R}^{\bar{d}} \to \mathbb{R}^{\bar{d}}$, $c_\beta : \mathbb{R}^{\bar{d}} \to \mathbb{R}^{\bar{d}\times\bar{d}}$ are smooth functions depending on some parameter $\beta \in \Theta_\beta$, and $\Xi(\sigma) = \mathrm{diag}(\sigma)$ for $\sigma = (\sigma_1, \ldots, \sigma_{\bar{d}}) \in \Theta_\sigma$ with $\sigma_j > 0$, $1 \leq j \leq \bar{d}$. System (34) includes many models used in applications, including, e.g., the Langevin equation (Pavliotis, 2014).

For $X_t = [X_{R,t}^\top, X_{S,t}^\top]^\top \in \mathbb{R}^{2\bar{d}}$ given by (34), the local Gaussian density (21) is specified via the following mean vector $\mu(\Delta, x; \theta) \in \mathbb{R}^{2\bar{d}}$ and covariance matrix $\Sigma(\Delta, x; \theta) \in \mathbb{R}^{2\bar{d}\times 2\bar{d}}$:

$$\mu(\Delta, x; \theta) = \begin{bmatrix} x_R - (c_\beta(x_S)x_R + g_\beta(x_S)) \cdot \Delta \\ x_S + x_R \cdot \Delta - (c_\beta(x_S)x_R + g_\beta(x_S)) \cdot \Delta^2/2 \end{bmatrix}, \quad \Sigma(\Delta, x; \theta) = \begin{bmatrix} \Delta & \Delta^2/2 \\ \Delta^2/2 & \Delta^3/3 \end{bmatrix} \otimes \Xi(\sigma)\Xi(\sigma)^\top.$$

The diffusion matrix Ξ is independent of the state $x \in \mathbb{R}^{2\bar{d}}$ and the drift function is linear in $X_{R,t}$ given $X_{S,t}$. Thus, most of the terms in the definition of Φ_2 appearing in the contrast (29) are equal to 0, e.g., $\hat{V}_{j_1} V_{k_1}^{i_1}(x, \theta) = 0$, $\hat{V}_{j_2} \hat{V}_{k_2} V_0^{i_2}(x, \theta) = 0$ for $1 \leq i_1, k_1, k_2 \leq \bar{d}$, $0 \leq j_1, j_2 \leq \bar{d}$, $\bar{d} + 1 \leq i_2 \leq 2\bar{d}$. Thus, the correction term Φ_2 in the case of system (34) takes up a simple form. For $x = (x_R, x_S)$, $y \in \mathbb{R}^{2\bar{d}}$, $\Delta > 0$ and $\theta = (\beta, \sigma) \in \Theta$, we get:

$$\begin{aligned} \Phi_2(\Delta, x, y; \theta) = &- \sum_{1 \leq i_1, i_2 \leq \bar{d}} c_{\beta, i_1 i_2}(x_S) \times (\sigma_{i_2})^2 \times \Big\{ \tfrac{1}{2} \mathcal{H}_{(i_1 i_2)}(\Delta, x, y; \theta) + \tfrac{1}{3} \mathcal{H}_{(i_1 i_2 + \bar{d})}(\Delta, x, y; \theta) \\ &+ \tfrac{1}{6} \mathcal{H}_{(i_1 + \bar{d}, i_2)}(\Delta, x, y; \theta) + \tfrac{1}{8} \mathcal{H}_{(i_1 + \bar{d}, i_2 + \bar{d})}(\Delta, x, y; \theta) \Big\}, \end{aligned} \tag{35}$$

where $\mathcal{H}_{(i_1 i_2)}(\Delta, x, y; \theta)$ is the second order Hermite polynomial defined in (27).

4.2. Low-frequency observation regime

We consider the scenario of low-frequency observations, i.e. the time interval among successive observations, $\Delta = t_i - t_{i-1}$, $1 \leq i \leq n$, is now assumed fixed, and large enough so that approximation schemes must be combined with a *data augmentation* approach. Thus, we move onto a Bayesian inference setting. Let $\delta = \delta_M := \Delta/M$, $M \geq 1$, be the user-induced step-size after imputation of $(M-1)$ data points amongst a pair of observations. Given a transition density scheme, say $\bar{p}_\delta(x, y; \theta)$, the true (intractable) transition density is approximated as:

$$p_\Delta^X(X_{t_{i-1}}, X_{t_i}; \theta) \approx \int_{\mathbb{R}^{d\times(M-1)}} \Big\{ \prod_{1 \leq j \leq M} \bar{p}_\delta(x_{j-1}, x_j; \theta) \Big\}\, dx_1 \cdots dx_{M-1}, \quad x_0 = X_{t_{i-1}},\ x_M = X_{t_i}.$$

In the case of elliptic diffusions, i.e., $d_S = 0$, $d_R = d$, Gobet and Labart (2008) showed that the bias induced by the Euler-Maruyama (EM) scheme is of size $O(M^{-1})$. In Section 4.2.1, 4.2.2, we develop two-types of explicit transition density schemes achieving local weak third order convergence via appropriate choice of higher order correction terms in the density expansion formula (24). We then illustrate in Section 4.2.3 that the discretisation bias of the developed transition schemes is of size $O(M^{-2})$, for the class of elliptic SDEs.

4.2.1. Local weak third order transition density scheme – version I

Making use of the small time expansion of the transition density $y \mapsto p_{\Delta}^{\bar{X}}(x, y; \theta)$ in Lemma 3.1, we obtain the following key result whose proof is given in Section F of Supplementary Material (Iguchi, Beskos and Graham, 2024).

Proposition 4.6 (Density expansion for local third order weak approximation). *Let $x, y \in \mathbb{R}^d$, $\Delta > 0$, $\theta \in \Theta$, and assume that conditions (H1)–(H4) hold. Then, for any integer $J \geq 4$, the transition density $y \mapsto p_{\Delta}^{\bar{X}}(x, y; \theta)$, admits the following representation:*

$$p_{\Delta}^{\bar{X}}(x, y; \theta) = p_{\Delta}^{\widetilde{X}^{\mathrm{LG}}}(x, y; \theta)\big\{1 + \Psi^{\mathrm{weak}}(\Delta, x, y; \theta) + R_1^J(\Delta, x, y; \theta)\big\} + \frac{\Delta^{J/2}}{\sqrt{\Delta^{d_R + 3d_S}}} R_2^J(x, y; \theta),$$

with residual terms $R_1^J(\Delta, x, y; \theta)$ and $R_2^J(x, y; \theta)$. In particular:

(i) *We have the expression:*

$$\Psi^{\mathrm{weak}}(\Delta, x, y; \theta) = \sqrt{\Delta} \cdot \Psi_1^{\mathrm{weak}}(\Delta, x, y; \theta) + \Delta \cdot \Psi_2^{\mathrm{weak}}(\Delta, x, y; \theta) + \sqrt{\Delta^3} \cdot \Psi_3^{\mathrm{weak}}(\Delta, x, y; \theta), \tag{36}$$

for the individual terms:

$$\Psi_1^{\mathrm{weak}}(\Delta, x, y; \theta) = \tfrac{1}{2} \sum_{1 \leq i_1, i_2, i_3 \leq d_R} \sum_{1 \leq k_1, k_2 \leq d_R} \hat{V}_{k_1} V_{k_2}^{i_1}(x, \theta) V_{k_1}^{i_2}(x, \theta) V_{k_2}^{i_3}(x, \theta)\, \mathcal{H}_{(i_1, i_2, i_3)}(\Delta, x, y; \theta); \tag{37}$$

$$\begin{aligned} \Psi_2^{\mathrm{weak}}(\Delta, x, y; \theta) = \tfrac{1}{2} &\sum_{1 \leq i_1, i_2 \leq d_R} \sum_{1 \leq k \leq d_R} \Big(\hat{V}_k V_0^{i_1}(x, \theta) + \hat{V}_0 V_k^{i_1}(x, \theta)\Big) V_k^{i_2}(x, \theta) \mathcal{H}_{(i_1, i_2)}(\Delta, x, y; \theta) \\ &+ \tfrac{1}{4} \sum_{1 \leq i_1, i_2 \leq d_R} \sum_{1 \leq k_1, k_2 \leq d_R} \hat{V}_{k_1} V_{k_2}^{i_1}(x, \theta) \hat{V}_{k_1} V_{k_2}^{i_2}(x, \theta) \mathcal{H}_{(i_1, i_2)}(\Delta, x, y; \theta); \end{aligned} \tag{38}$$

$$\Psi_3^{\mathrm{weak}}(\Delta, x, y; \theta) = \tfrac{1}{2} \sum_{1 \leq i \leq d_R} \hat{V}_0 V_0^i(x, \theta) \mathcal{H}_{(i)}(\Delta, x, y; \theta). \tag{39}$$

(ii) *For any $\varphi \in C_p^{\infty}(\mathbb{R}^d; \mathbb{R})$, there exist constants $C > 0$, $q \geq 1$ such that:*

$$\Big| \int_{\mathbb{R}^d} \varphi(y) p_{\Delta}^{\widetilde{X}^{\mathrm{LG}}}(x, y; \theta) R_1^J(\Delta, x, y; \theta) dy \Big| \leq C(1 + |x|^q)\Delta^3. \tag{40}$$

(iii) $\sup_{x, y \in \mathbb{R}^d, \theta \in \Theta} |R_2^J(x, y; \theta)| < C$, *for a constant $C > 0$.*

(iv) $\int_{\mathbb{R}^d} \Psi^{\mathrm{weak}}(\Delta, x, y; \theta) p_{\Delta}^{\widetilde{X}^{\mathrm{LG}}}(x, y; \theta) dy = 0.$

Thus, from (ii) and (iii) with $J \geq 6$, we have that for any $\varphi \in C_p^{\infty}(\mathbb{R}^d; \mathbb{R})$ and $(\Delta, x, \theta) \in (0, \infty) \times \mathbb{R}^d \times \Theta$, there exist constants $C > 0$, $q \geq 1$ such that:

$$\Big| \mathbb{E}_{\theta}[\varphi(\bar{X}_{\Delta}^x)] - \int_{\mathbb{R}^d} \varphi(y) p_{\Delta}^{\widetilde{X}^{\mathrm{LG}}}(x, y; \theta)\big\{1 + \Psi^{\mathrm{weak}}(\Delta, x, y; \theta)\big\} dy \Big| \;\leq\; C(1 + |x|^q)\Delta^3.$$

Using Proposition 2.3, the above gives immediately that, for any $\varphi \in C_p^{\infty}(\mathbb{R}^d; \mathbb{R})$ and $(\Delta, x, \theta) \in (0, \infty) \times \mathbb{R}^d \times \Theta$, there exist constants $C > 0$, $q \geq 1$ so that:

$$\Big| \mathbb{E}_{\theta}[\varphi(X_{\Delta}^x)] - \int_{\mathbb{R}^d} \varphi(y)\, p_{\Delta}^{\widetilde{X}^{\mathrm{LG}}}(x, y; \theta)\big\{1 + \Psi^{\mathrm{weak}}(\Delta, x, y; \theta)\big\} dy \Big| \;\leq\; C(1 + |x|^q)\Delta^3.$$

In the case of elliptic diffusions ($d = d_R$, $d_S = 0$), Proposition 4.6 is interpreted as follows:

$$p^{\bar{X}}_{e,\Delta}(x, y; \theta) = p^{\widetilde{X}^{\mathrm{EM}}}_{e,\Delta}(x, y; \theta) \left\{1 + \Psi^{\mathrm{weak}}_e(\Delta, x, y; \theta) + R^J_{e,1}(\Delta, x, y; \theta)\right\} + \frac{\Delta^{J/2}}{\sqrt{\Delta^{d_R}}} R^J_{e,2}(x, y; \theta),$$

where $y \mapsto p^{\widetilde{X}^{\mathrm{EM}}}_{e,\Delta}(x, y; \theta)$ denotes the density of the one-step Euler-Maruyama scheme given $\theta \in \Theta$ and the start point $x \in \mathbb{R}^{d_R}$, and terms $R^J_{e,1}(\Delta, x, y; \theta)$, $R^J_{e,2}(x, y; \theta)$ have the same properties as the corresponding ones in Proposition 4.6. The correction part $\Psi^{\mathrm{weak}}_e(\Delta, x, y; \theta)$ corresponds to $\Psi^{\mathrm{weak}}(\Delta, x, y; \theta)$ with $d = d_R$ and the Hermite polynomials replaced with $\mathcal{H}^{\widetilde{Y}_e}_\alpha(\xi; \theta)|_{\xi = m_{x,\theta,\sqrt{\Delta}}(y)}$ whose definition has been adjusted in an apparent way to conform to the setting $d_S = 0$.

In light of Proposition 4.6, we propose the use of the following transition density scheme for the hypo-elliptic case:

$$\bar{p}^{\mathrm{I}}_\Delta(x, y; \theta) := p^{\widetilde{X}^{\mathrm{LG}}}_\Delta(x, y; \theta)\left\{1 + \Psi^{\mathrm{weak}}(\Delta, x, y; \theta)\right\}, \tag{41}$$

for $\Psi^{\mathrm{weak}}(\Delta, x, y; \theta)$ as given in (36), with the following corresponding scheme for the elliptic case:

$$\bar{p}^{\mathrm{I}}_{e,\Delta}(x, y; \theta) := p^{\widetilde{X}^{\mathrm{EM}}}_{e,\Delta}(x, y; \theta)\left\{1 + \Psi^{\mathrm{weak}}_e(\Delta, x, y; \theta)\right\}. \tag{42}$$

It is worth mentioning that the quantity $\bar{p}^{\mathrm{I}}_{e,\Delta}(x, y; \theta)$ has also appeared in results by Iguchi and Yamada (2021, 2022), where a weak high order approximation was developed for elliptic SDEs via a different approach. We also note that $\bar{p}^{\mathrm{I}}_\Delta(x, y; \theta)$ has the following properties:

(i) The normalising constant is equal to 1, i.e. it holds that

$$\int_{\mathbb{R}^d} \bar{p}^{\mathrm{I}}_\Delta(x, y; \theta) dy = \int_{\mathbb{R}^d} p^{\widetilde{X}^{\mathrm{LG}}}_\Delta(x, y; \theta)\left\{1 + \Psi^{\mathrm{weak}}(\Delta, x, y; \theta)\right\} dy = 1.$$

(ii) $\Psi^{\mathrm{weak}}(\Delta, x, y; \theta)$, thus, also, $\bar{p}^{\mathrm{I}}_\Delta(x, y; \theta)$, can, in general, take negative values, as do other closed-form expansions proposed by Aït-Sahalia (2008) and Li (2013).

4.2.2. *Local weak third order transition density scheme – version II*

We construct an alternative transition density scheme that provides a proper, everywhere positive, density function. See also Stramer, Bognar and Schneider (2010) for a related consideration in the context of the density expansion for elliptic SDEs put forward by Aït-Sahalia (2002). We consider the truncated Taylor expansion of $\log(1 + z)$ and introduce, for $z \in \mathbb{R}$,

$$K(z) := \sum_{1 \le l \le 6} (-1)^{l+1} \tfrac{z^l}{l}. \tag{43}$$

For $\Delta > 0$, $x, y \in \mathbb{R}^d$, $\theta \in \Theta$, we define the unnormalised non-negative approximation:

$$\bar{p}^{\mathrm{II}}_\Delta(x, y; \theta) := p^{\widetilde{X}^{\mathrm{LG}}}_\Delta(x, y; \theta) \exp\big(K(\Psi^{\mathrm{weak}}(\Delta, x, y; \theta))\big). \tag{44}$$

The mapping $y \mapsto \bar{p}^{\mathrm{II}}_\Delta(x, y; \theta)$ is integrable as the highest order term of $K(z)$ in (43) is $-z^6$, so $\exp(K(z))$ is bounded. Notice that the normalising constant $\bar{Z}(\Delta, x; \theta) := \int_{\mathbb{R}^d} \bar{p}^{\mathrm{II}}_\Delta(x, y; \theta) dy$ is no longer 1. Nevertheless, the unnormalised non-negative approximation $\bar{p}^{\mathrm{II}}_\Delta(x, y; \theta)$ delivers a local weak third order approximation for X^x_Δ, in the following sense with its proof provided in Section G of Supplementary Material (Iguchi, Beskos and Graham, 2024).

Proposition 4.7. *For any $\varphi \in C_p^\infty(\mathbb{R}^d;\mathbb{R})$ and $(\Delta, x, y) \in (0,\infty) \times \mathbb{R}^d \times \Theta$, there exist constants $C > 0$, $q \geq 1$, such that:*

$$\left| \int_{\mathbb{R}^d} \varphi(y) \bar{p}_\Delta^{\mathrm{I}}(x,y;\theta) dy - \int_{\mathbb{R}^d} \varphi(y) \bar{p}_\Delta^{\mathrm{II}}(x,y;\theta) dy \right| \leq C(1+|x|^q)\Delta^3, \tag{45}$$

and then, from Proposition 4.6 (iv),

$$\left| \mathbb{E}_\theta[\varphi(X_\Delta^x)] - \int_{\mathbb{R}^d} \varphi(y) \bar{p}_\Delta^{\mathrm{II}}(x,y;\theta) dy \right| \leq C(1+|x|^q)\Delta^3,$$

for some constants $C > 0$, $q \geq 1$.

The transition density scheme $\bar{p}_\Delta^{\mathrm{II}}(x,y;\theta)$ in (44) has the properties:

(i) It takes non-negative values for all $x, y \in \mathbb{R}^d$ and $\theta \in \Theta$.

(ii) From Proposition 4.7, it provides a local weak third order approximation for the solution of the SDE (1), for test functions $\varphi \in C_p^\infty(\mathbb{R}^d;\mathbb{R})$. The discrepancy between the normalising constant $\bar{Z}(\Delta, x;\theta)$ and 1 is of size $O(\Delta^3)$, by setting $\varphi(y) = 1$ in (45).

4.2.3. Discretisation based upon the transition density schemes

We finally illustrate the discretisation bias induced from the proposed transition density schemes for elliptic SDEs, i.e., $d_S = 0, d_R = d$. For the first-type transition density scheme in (42), Iguchi and Yamada (2021) showed that the discretisation bias is of size $O(M^{-2})$, i.e., there exist constants $C, c > 0$, $q \geq d_R/2$ and a non-decreasing $h(\cdot)$ such that for any $x, y \in \mathbb{R}^{d_R}$,

$$\left| p_{e,\Delta}^X(x,y;\theta) - \bar{p}_{e,\Delta}^{\mathrm{I},(M)}(x,y;\theta) \right| \leq \frac{C}{M^2} \frac{h(\Delta)}{\Delta^q} e^{-c\frac{|y-x|^2}{\Delta}}, \tag{46}$$

Remark 4.8. The bound (46) was derived without reference to unknown SDE parameters, thus are not uniform in θ. Nevertheless, under compactness of Θ, see (H1), they can be readily adapted to be uniform also in $\theta \in \Theta$.

We have the following estimate for the second-type transition density scheme with a proof given in Section H of Supplementary Material (Iguchi, Beskos and Graham, 2024).

Theorem 4.9. *Let $\Delta > 0$. Assume $d_S = 0$ and that conditions (H1)–(H3) hold. Then, there exist constants $C, c > 0$, $q \geq d_R/2$ and a non-decreasing function $h(\cdot)$ such that for any $x, y \in \mathbb{R}^{d_R}$, $\theta \in \Theta$:*

$$\left| p_{e,\Delta}^X(x,y;\theta) - \bar{p}_{e,\Delta}^{\mathrm{II},(M)}(x,y;\theta) \right| \leq \frac{C}{M^2} \frac{h(\Delta)}{\Delta^q} e^{-c\frac{|y-x|^2}{\Delta}}. \tag{47}$$

5. Numerical experiments

We provide numerical experiments related to the analytic results in Section 4. Note that our software implementation (https://github.com/matt-graham/simsde) of both the proposed contrast function and the sampling schemes is fully general. The user is only required to specify functions for evaluating the drift and diffusion coefficient terms for given a process state and model parameters, with functions for evaluating the contrast estimator or simulating from the sampling scheme then automatically generated. This is achieved by symbolically computing the terms in the relevant estimator or sampling scheme expression using the Python-based computer algebra system SymPy (https://www.sympy.org/). The generated functions can also optionally use the numerical primitives defined in the high-performance numerical computing framework JAX (https://jax.readthedocs.io/); this allows the derivatives of the generated functions to be computed using JAX's automatic differentiation support. We exploit the above to automatically generate efficient functions for computing the gradient of the contrast function in the optimisation-based simulation study in Section 5.1 and the gradient of the posterior density in the Bayesian inference numerical experiments in Section 5.2.

5.1. High-frequency observation regime

We consider the stochastic Jansen–Rit neural mass model that describes the evolution of neural population in a local cortical circuit, that is, the interaction of the main pyramidal cells with the excitatory and inhibitory interneurons. See Ableidinger, Buckwar and Hinterleitner (2017) for more details about the model. The model is defined as the following 6-dimensional hypo-elliptic SDE driven by a 3-dimensional Brownian motion ($d = 6$, $d_S = d_R = 3$):

$$\begin{aligned} dX_{R,t} &= \big(-\Gamma_\theta^2 X_{S,t} - 2\Gamma_\theta X_{R,t} + G(X_{S,t},\theta)\big)dt + \Sigma_\theta dB_t, \quad X_{R,0} = x_{R,0} \in \mathbb{R}^3; \\ dX_{S,t} &= X_{R,t}dt, \quad X_{S,0} = x_{S,0} \in \mathbb{R}^3, \end{aligned} \tag{48}$$

for parameter specified as $\theta = (\beta, \sigma)$ with $\beta = (A, B, C, \mu, \nu_0, a, b, r, \nu_{\max}) \in \mathbb{R}^5 \times (0,\infty)^4$, $\sigma = (\sigma_1, \sigma_2, \sigma_3) \in (0,\infty)^3$. We have set $\Gamma_\theta := \mathrm{diag}(a, a, b)$, $\Sigma_\theta := \mathrm{diag}(\sigma_1, \sigma_2, \sigma_3)$ and

$$G(x_S, \theta) := \big[Aa \times \mathcal{S}_\theta(x_S^2 - x_S^3),\ Aa\big(\mu + C_1 \times \mathcal{S}_\theta(C x_S^1)\big),\ Bb\, C_2 \times \mathcal{S}_\theta(C_3 x_S^1)\big]^\top, \quad x_S = (x_S^1, x_S^2, x_S^3),$$

where $C_1 = 0.8 \times C$, $C_2 = C_3 = 0.25 \times C$ and $\mathcal{S}_\theta : \mathbb{R} \to [0, \nu_{\max}]$ is a sigmoid function defined as $\mathcal{S}_\theta(z) := \nu_{\max}/\{1 + \exp(r(\nu_0 - z))\}$, $z \in \mathbb{R}$. Following the numerical experiment in Buckwar, Tamborrino and Tubikanec (2020), we fix a part of the parameter vector θ to $A = 3.25$, $B = 22$, $\nu_0 = 6$, $a = 100$, $b = 50$, $\nu_{\max} = 5$, $r = 0.56$, $\sigma_1 = 0.01$, $\sigma_3 = 1$. Then, we estimate the parameter (C, μ, σ_2) from observations of all coordinates of model (48) using the new contrast and the local Gaussian estimators. We set the true parameter values to $(C^\dagger, \mu^\dagger, \sigma_2^\dagger) = (135.0, 220.0, 2000.0)$, and generate synthetic datasets Y_{JR} from the local Gaussian scheme (12) with discretisation step 10^{-4} on the time interval $[0, 100]$. Then, we check the performance of the new contrast estimator versus the local Gaussian one in the following three scenarios for n and Δ_n by subsampling from the synthetic datasets Y_{JR}:
JR-1. $(n, \Delta_n) = (1.25 \times 10^4, 0.008)$ with the time interval of observations $T = 100$.
JR-2. $(n, \Delta_n) = (2.5 \times 10^4, 0.004)$ with $T = 100$. **JR-3.** $(n, \Delta_n) = (5 \times 10^4, 0.002)$ with $T = 100$.
Since model (48) belongs in the class of stochastic damping Hamiltonian system (34) and matrix Γ_θ is diagonal, the correction term Φ_2 in the proposed contrast function is given as follows:

$$\begin{aligned} \Phi_2(\Delta, x, y; \theta) = -\sum_{1 \le i \le 3} 2\Gamma_{\theta,ii} (\sigma_i)^2 \times \Big\{ &\tfrac{1}{2}\mathcal{H}_{(i,i)}(\Delta, x, y; \theta) + \tfrac{1}{3}\mathcal{H}_{(i,i+3)}(\Delta, x, y; \theta) \\ &+ \tfrac{1}{6}\mathcal{H}_{(i+3,i)}(\Delta, x, y; \theta) + \tfrac{1}{8}\mathcal{H}_{(i+3,i+3)}(\Delta, x, y; \theta) \Big\}, \end{aligned}$$

with $\Delta > 0$, $x, y \in \mathbb{R}^d$, $\theta \in \Theta$. As we noted in Section 4.1.4, partial derivatives of the non-linear function $G(X_{S,t}, \theta)$ are not required in the computation of the new contrast function. For the minimisation of the contrast, we use the Adam optimiser (Kingma and Ba, 2015) with the following algorithmic specifications: (step-size) = 0.01, (exponential decay rate for the first moment estimates) = 0.9, (exponential decay rate for the second moment estimates) = 0.999, (additive term for numerical stability) =

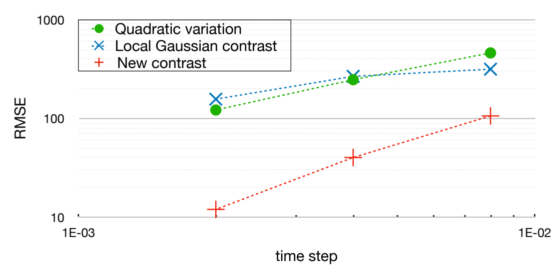
Table 1. Mean and standard error (in brackets) from 50 replicates of parameter estimates in the Jansen-Rit neural mass model (48), under scenario **JR-3**.

Parameter	True value	Local Gaussian contrast	New contrast
C	135.0	134.80 (0.0062)	134.80 (0.0062)
μ	220.0	220.84 (0.6167)	220.84 (0.6176)
σ_2	2000.0	1843.49 (4.3520)	1989.00 (4.750)

Table 2. Mean and standard deviation (in brackets) of (true value of σ_2) - (estimator of σ_2) from 50 replicates of three different estimates in the Jansen-Rit neural mass model (48), under scenarios **JR-1,2,3**.

Scenario	Quadratic variation	Local Gaussian contrast	New contrast
JR-1	460.34 (11.327)	314.99 (8.913)	-105.16 (12.419)
JR-2	246.26 (8.677)	266.46 (6.695)	39.46 (7.450)
JR-3	121.42 (6.588)	156.51 (4.352)	11.00 (4.750)

Figure 1: Root mean squared errors (RMSE) for estimators for σ_2 (50 replicates) in the Jansen-Rit neural mass model (48) under scenarios **JR-1,2,3**.



1×10^{-8} and (number of iterations) = 20,000. In Table 1, we summarise the mean and standard error from 50 replicates of parameter estimates under scenario **JR-3**. The table indicates that both the local Gaussian and the new contrast estimate accurately the drift parameters (C, μ), but the new contrast produces better results for the diffusion parameter σ_2 as explained in Remark 4.5. Thus, we focus on the estimation of the diffusion parameter σ_2 and compare the proposed contrast estimator with, first, the local Gaussian estimator and, second, the estimator based on the quadratic variation of $X^2_{R,t}$, i.e., $\hat{\sigma}^{\mathrm{QV}}_{2,n} := \sqrt{\frac{1}{T}\sum_{1\le m\le n}\big(X^2_{R,m\Delta_n} - X^2_{R,(m-1)\Delta_n}\big)^2}$. We summarise in Table 2 the mean and standard deviation of (true value of σ_2) - (estimator of σ_2) from 50 replicates of estimates under scenario **JR-1,2,3**. Also, in Figure 1 we plot the root mean squared errors (RMSEs) from 50 replicates of estimates of σ_2. Note that the new contrast estimator for σ_2 gives faster convergence to the true value $\sigma_2^\dagger = 2000$, in agreement with a relatively larger Δ_n (as a function of n) permitted in the CLT for the new contrast – recall that for the CLT, the new contrast requires $\Delta_n = o(n^{-1/3})$ versus $\Delta_n = o(n^{-1/2})$ required by the local Gaussian contrast. We also observe that the new contrast estimator outperforms $\hat{\sigma}^{\mathrm{QV}}_{2,n}$. This is arguably not a coincidence for the particular experiment, as the contrast estimator makes use of observations from all co-ordinates (not only of X^2_R) via a likelihood-based contrast function, and MLEs are well-understood to possess optimal asymptotic properties, e.g. due to the Cramér-Rao bound.

5.2. Low-frequency observation regime

We perform Bayesian inference, in a framework requiring data augmentation, for a susceptible-infected-recovered (SIR) model with a time-varying contact rate. The model is specified as the following 3-dimensional SDE:

$$\begin{bmatrix} dS_t \\ dI_t \\ dC_t \end{bmatrix} = \begin{bmatrix} -\frac{C_t S_t I_t}{N} \\ \frac{C_t S_t I_t}{N} - \lambda I_t \\ \left(\alpha(\beta - \log C_t) + \frac{\sigma^2}{2}\right) C_t \end{bmatrix} dt + \begin{bmatrix} \sqrt{\frac{C_t S_t I_t}{N}} & 0 & 0 \\ -\sqrt{\frac{C_t S_t I_t}{N}} & \sqrt{\lambda I_t} & 0 \\ 0 & 0 & \sigma \end{bmatrix} \begin{bmatrix} dB_{1,t} \\ dB_{2,t} \\ dB_{3,t} \end{bmatrix}, \tag{49}$$

where S, I account for the number of susceptible and infected individuals respectively, C for the contact rate, N is the population size and $\theta = (\alpha, \beta, \sigma, \lambda)$ the parameter vector. Motivated by an implementation in Graham, Thiery and Beskos (2022), we consider the following setting:

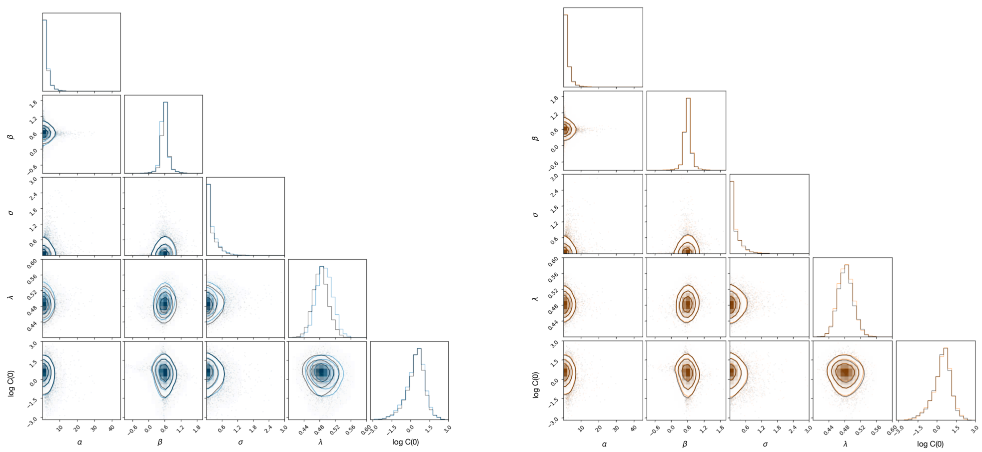
- Observations $Y = \{Y_t\}_{t=0,\dots,T}$ correspond to the number of daily infected individuals over the period of 14 days, with time interval $\Delta = 1$ (day), and they are assumed to be measured with additive Gaussian noise, as $Y_t = I_t + \sigma_y \varepsilon_t$, $0 \le t \le 13$, where $\varepsilon_t \sim \mathcal{N}(0,1)$ and $\sigma_y > 0$ is constant.
- For the parameter vector θ and the initial condition C_0, priors are determined via $\theta = g_\theta(u)$ and $C_0 = g_0(u_0)$, where u, u_0 are vectors of independent standard normal random variables and g_θ, g_0 are some tractable functions.
- Adopting a Bayesian data augmentation approach, a numerical scheme is called upon to impute instances of model (49) at times separated by discretisation step δ. The MCMC method is based on a non-centred imputation approach. That is, latent variables correspond to the Brownian increments driving the SDE, and they are collected in a vector of (apriori) independent standard normal random variables v.

In the above setting, the posterior law of the d_q-dimensional vector $q = [u^\top, u_0^\top, v^\top]^\top$ given the data Y has a tractable density w.r.t. the Lebesgue measure (for more details, see Section 11 in the Supporting Material of Graham, Thiery and Beskos (2022)). Thus, one can apply standard Hamiltonian Monte Carlo (HMC) to sample from the posterior with energy function $H : \mathbb{R}^{d_q} \times \mathbb{R}^{d_q} \to \mathbb{R}$ given as $H(q,p) = \ell(q) + \frac{1}{2} p^\top M^{-1} p$, where $\ell(q)$ is the negative log-posterior density and M a diagonal mass matrix. We will run HMC on the posterior induced both by the Euler-Maruyama (EM) scheme and the proposed weak second order scheme, and we will show (numerically) that the involved bias is smaller in the case of the new scheme.

The details of the design of the experiment are as follows. We use the observations taken from Anonymous (1978) as data Y. We fix $N = 763$, $S_0 = 762$, $I_0 = 0$, $\sigma_y = 5$. We assign priors $\log \alpha \sim \mathcal{N}(0,1)$, $\beta \sim \mathcal{N}(0,1)$, $\log \sigma \sim \mathcal{N}(-3,1)$, $\log \lambda \sim \mathcal{N}(0,1)$ and $\log C_0 \sim \mathcal{N}(0,1)$. We do not treat σ_y as unknown here due to the widely varying posterior scales induced by non-constant σ_y reported in Graham, Thiery and Beskos (2022). The (data-imputing) numerical schemes are applied to the log-transformation $X_t = [\log S_t, \log I_t, \log C_t]^\top$ to ensure positiveness of S, I, and C, and avoid numerical issues with the square-root terms in the diffusion coefficient. We use a dynamic integration-time HMC implementation (Betancourt (2017)) with a dual-averaging algorithm (Hoffman et al. (2014)) to adapt the integrator step-size. For the time-discretisation of the Hamiltonian dynamics we use the leapfrog integrator with Störmer-Verlet splitting. We set the mass matrix M to identity. For each of the following three choices of numerical schemes, we run four HMC chains of 1,500 iterations with the first 500 iterations used as an adaptive warm-up phase:

SIR-baseline: $\left(\text{Numerical scheme}, \delta_M\right) = \left(\text{EM}, 0.001\right)$.

SIR-EM: $\left(\text{Numerical scheme}, \delta_M\right) = \left(\text{EM}, 0.05\right)$.



(a) Euler-Maruyama scheme (b) Weak second order scheme

Figure 2: Posterior Estimates for SIR Model. The blue (left panel) and orange (right panel) histograms and contour plots are obtained by HMC that uses the EM scheme and the weak second order scheme, respectively, both with the same discretisation step $\delta_M = 0.05$. The black histograms and contour plots superimposed in both plots show the correct quantities, as obtained from HMC that uses the EM scheme with very small $\delta_M = 10^{-3}$.

SIR-Weak2nd: $\big(\text{Numerical scheme}, \delta_M\big) = \big(\text{Weak second order scheme}, 0.05\big)$.

Results from the HMC implementations are summarised in Figure 2 and Table 3. Figure 2 shows estimated pair-wise and marginal posteriors for parameters α, β, σ, λ and $\log C_0$. Results for **SIR-EM** and **SIR-Weak2nd** are shown in separate plots, and each plot superimposes corresponding results from **SIR-baseline**, this latter scheme treated as providing the 'correct' posterior quantities due to the use of very small $\delta_M = 0.001$. Table 3 shows summary statistics that monitor the performance of the HMC algorithm, in particular, bulk effective sample size (ESS), tail ESS and improved $\hat{R}$ with rank-normalisation and folding (see Vehtari et al. (2021) for analytical definitions). These are computed from the non-warm-up steps of the four HMC chains. Note that for all parameters and choices of numerical schemes, $\hat{R}$ is smaller than 1.01 and ESS is larger than 400, as Vehtari et al. (2021) recommend. One can thus be reasonably confident that the estimated posteriors shown in Figure 2 are reliable representations of the true ones. In agreement with the analytical theory in this work, Figure 2 illustrates that for all parameters the estimated posteriors obtained via the weak second order scheme capture more accurately the 'correct' baseline posteriors than the corresponding ones obtained via the EM scheme.

6. Conclusions

This work begins by putting forward weak second order sampling schemes for elliptic and hypo-elliptic SDEs. Then, we develop a small time density expansion of the scheme as a proxy for the intractable SDE transition density. Via appropriate choice of the higher order expansion terms, we have provided analytical results both: in a high-frequency classical setting, showcasing the advantageous rate of $\Delta_n = o(n^{-1/3})$, achieved for hypo-elliptic models; and in a low-frequency Bayesian data augmentation setting, where we have deduced two local weak third order density schemes (41) and (44) and shown that the induced bias by the schemes is of size $O(M^{-2})$ when covering the fixed time interval $\Delta > 0$ with inner time-steps $\delta = \Delta/M$, $M > 0$ specified by user.

Table 3. Summary statistics of HMC for the SIR model.

Parameter	Scenario	bulk ESS	tail ESS	improved $\hat{R}$
α	SIR-baseline	4052.0	2811.0	1.0
	SIR-EM	3874.0	3131.0	1.0
	SIR-Weak2nd	3876.0	2813.0	1.0
β	SIR-baseline	4272.0	2934.0	1.0
	SIR-EM	4058.0	2878.0	1.0
	SIR-Weak2nd	3880.0	2969.0	1.0
σ	SIR-baseline	1720.0	2836.0	1.0
	SIR-EM	1525.0	2634.0	1.0
	SIR-Weak2nd	1781.0	2513.0	1.0
λ	SIR-baseline	6596.0	2952.0	1.0
	SIR-EM	7063.0	3052.0	1.0
	SIR-Weak2nd	7024.0	2941.0	1.0
$\log C_0$	SIR-baseline	4555.0	2889.0	1.0
	SIR-EM	5036.0	3092.0	1.0
	SIR-Weak2nd	4695.0	3053.0	1.0

We iterate here that there is the flexibility to apply MCMC methods based only on the sampling schemes without reference to transition densities, via, e.g., non-centred model parameterisation approaches (see e.g. Beskos, Dureau and Kalogeropoulos (2015), Papaspiliopoulos, Roberts and Sköld (2007)), particle-filtering based MCMC methods (Andrieu, Doucet and Holenstein, 2010), and recent manifold-based algorithms for the case of observations without/with noise (Graham, Thiery and Beskos, 2022). Via the derivation of approximate transition densities, one is still given the option to use the wealth of data augmentation methods for diffusion models that require such a density expression.

Acknowledgements

We thank two referees and the Associate Editor for their comments that led to major improvements in the content of the paper. YI is supported by the Additional Funding Programme for Mathematical Sciences, delivered by EPSRC (EP/V521917/1) and the Heilbronn Institute for Mathematical Research.

Supplementary Material

"Supplementary Material" (Iguchi, Beskos and Graham, 2024) provides the proofs of the main results and related technical proofs, and also contains an additional numerical experiment of implementation of the proposed contrast function under the high-frequency observations regime.

References

Ableidinger, M., Buckwar, E. and Hinterleitner, H. (2017). A stochastic version of the Jansen and Rit neural mass model: Analysis and numerics. *J. Math. Neurosci.* **7** 1–35.

Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. *Econometrica* **70** 223–262.

Aït-Sahalia, Y. (2008). Closed-form likelihood expansions for multivariate diffusions. *Ann. Statist.* **36** 906–937.

Andrieu, C., Doucet, A. and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 269–342.

ANONYMOUS (1978). News and notes: influenza in a boarding school. *Br. Med. J.* **1** 586–590.

BESKOS, A., DUREAU, J. and KALOGEROPOULOS, K. (2015). Bayesian inference for partially observed stochastic differential equations driven by fractional Brownian motion. *Biometrika* **102** 809–827.

BETANCOURT, M. (2017). A conceptual introduction to Hamiltonian Monte Carlo. *arXiv preprint arXiv:1701.02434*.

BUCKWAR, E., TAMBORRINO, M. and TUBIKANEC, I. (2020). Spectral density-based and measure-preserving ABC for partially observed diffusion processes. An illustration on Hamiltonian SDEs. *Stat. Comput.* **30** 627–648.

DITLEVSEN, S. and SAMSON, A. (2019). Hypoelliptic diffusions: filtering and inference from complete and partial observations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 361–384.

GLOTER, A. and YOSHIDA, N. (2020). Adaptive and non-adaptive estimation for degenerate diffusion processes. *arXiv preprint arXiv:2002.10164*.

GLOTER, A. and YOSHIDA, N. (2021). Adaptive estimation for degenerate diffusion processes. *Electron. J. Stat.* **15** 1424–1472.

GOBET, E. and LABART, C. (2008). Sharp estimates for the convergence of the density of the Euler scheme in small time. *Electron. Commun. Probab.* **13** 352–363.

GRAHAM, M. M., THIERY, A. H. and BESKOS, A. (2022). Manifold Markov chain Monte Carlo methods for Bayesian inference in diffusion models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 1229–1256.

HOFFMAN, M. D., GELMAN, A. et al. (2014). The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **15** 1593–1623.

IGUCHI, Y., BESKOS, A. and GRAHAM, M. (2024). Supplement to "Parameter estimation with increased precision for elliptic and hypo-elliptic Diffusions".

IGUCHI, Y. and YAMADA, T. (2021). Operator splitting around Euler-Maruyama scheme and high order discretization of heat kernels. *ESAIM: Math. Model. Numer. Anal.* **55** S323–S367.

IGUCHI, Y. and YAMADA, T. (2022). Weak approximation of SDEs for tempered distributions and applications. *Adv. Comput. Math* **48** 269–342.

IKEDA, N. and WATANABE, S. (2014). *Stochastic Differential Equations and Diffusion Processes*. Elsevier.

KESSLER, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scand. J. Stat.* **24** 211–229.

KINGMA, D. P. and BA, J. (2015). Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR* (Y. BENGIO and Y. LECUN, eds.).

KLOEDEN, P. E. and PLATEN, E. (1992). *Numerical Solution of Stochastic Differential Equations*. Springer.

LI, C. (2013). Maximum-likelihood estimation for diffusion processes via closed-form density expansions. *Ann. Statist.* **41** 1350–1380.

MILSTEIN, G. N. and TRETYAKOV, M. V. (2021). *Stochastic Numerics for Mathematical Physics*. Springer Cham.

NUALART, D. (2006). *The Malliavin Calculus and Related Topics*. Springer Berlin, Heidelberg.

PAPASPILIOPOULOS, O., ROBERTS, G. O. and SKÖLD, M. (2007). A general framework for the parametrization of hierarchical models. *Stat. Sci.* **22** 59–73.

PAVLIOTIS, G. A. (2014). *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations* **60**. Springer.

POKERN, Y., STUART, A. M. and WIBERG, P. (2009). Parameter estimation for partially observed hypoelliptic diffusions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 49–73.

PRAKASA RAO, B. L. S. (1988). Statistical inference from sampled data for stochastic processes. In *Statistical inference from stochastic processes (Ithaca, NY, 1987). Contemp. Math.* **80** 249–284. Amer. Math. Soc., Providence, RI.

STRAMER, O., BOGNAR, M. and SCHNEIDER, P. (2010). Bayesian inference for discretely sampled Markov processes with closed-form likelihood expansions. *J. Financ. Econom.* **8** 450–480.

UCHIDA, M. and YOSHIDA, N. (2012). Adaptive estimation of an ergodic diffusion process based on sampled data. *Stoch. Process. Their Appl.* **122** 2885–2924.

VEHTARI, A., GELMAN, A., SIMPSON, D., CARPENTER, B. and BÜRKNER, P.-C. (2021). Rank-normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of MCMC (with discussion). *Bayesian Anal.* **16** 667–718.

WATANABE, S. (1987). Analysis of Wiener functionals (Malliavin calculus) and its applications to heat kernels. *Ann. Probab.* **15** 1–39.