A closed-form transition density expansion for elliptic and hypo-elliptic SDEs

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

[1]*School of Mathematical Sciences Lancaster University, Lancaster, UK* , [a]*iguchi@lancaster.ac.uk*
[2]*Department of Statistical Science, University College London, London, UK* , [b]*a.beskos@ucl.ac.uk*

We introduce a closed-form expansion for the transition density of elliptic and hypo-elliptic multivariate Stochastic Differential Equations (SDEs), over a period $\Delta \in (0,1)$, in terms of powers of $\Delta^{j/2}$, $j \geq 0$. Our methodology provides approximations of the transition density, easily evaluated via any software that performs symbolic calculations. A major part of the paper is devoted to an analytical control of the remainder in our expansion for fixed $\Delta \in (0,1)$. The obtained error bounds validate theoretically the methodology, by characterising the size of the distance from the true value. It is the first time that such a closed-form expansion becomes available for the important class of hypo-elliptic SDEs, to the best of our knowledge. For elliptic SDEs, closed-form expansions are available, with some works identifying the size of the error for fixed Δ, as per our contribution. Our methodology allows for a uniform treatment of elliptic and hypo-elliptic SDEs, when earlier works are intrinsically restricted to an elliptic setting. We show numerical applications highlighting the effectiveness of our method, by carrying out parameter inference for hypo-elliptic SDEs that do not satisfy stated conditions. The latter are sufficient for controlling the remainder terms, but the closed-form expansion itself is applicable in general settings.

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

1. Introduction

Stochastic Differential Equations (SDEs) constitute an effective tool for modelling non-linear dynamics that arise in numerous application fields, including, e.g., finance, physics and neuroscience (Kloeden and Platen, 1992). Over the past few decades, a large amount of research has contributed to methodological and theoretical advances on the theme of parameter inference for SDEs. An overarching challenge is that the transition density of a non-linear SDE is in general intractable, thus appropriate proxies must be formulated to conduct likelihood-based inference. We propose a new closed-form (CF) transition density expansion for SDEs, which can approximate the true density with high precision. In contrast to previous approaches, one of the novelties of our methodology is that it covers a broad class of diffusion processes, including *hypo-elliptic* SDEs, i.e. processes with a degenerate diffusion matrix and a transition law that still admits a density with respect to (w.r.t.) the Lebesgue measure. Hypo-elliptic SDEs appear in broad areas of applications (including physics, neuroscience) and parameter inference for these models has been a very active area of research in the last years.

Let $B_t = (B_{1,t}, \ldots, B_{d,t})$, $t \geq 0$, be the standard d-dimensional Brownian motion, $d \geq 1$, defined upon the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$. We consider N-dimensional SDEs, $N \geq 1$, of the following general form:

$$dX_t = V_0(X_t,\theta)\,dt + \sum_{1\leq j\leq d} V_j(X_t,\theta)\,dB_{j,t}, \quad X_0 = x_0 \in \mathbb{R}^N, \tag{1}$$

for parameter vector $\theta \in \Theta \subseteq \mathbb{R}^{N_\theta}$, $N_\theta \geq 1$, and functions $V_j : \mathbb{R}^N \times \Theta \to \mathbb{R}^N$, $0 \leq j \leq d$. We set $\sigma = [V_1, \ldots, V_d]$ and $a = \sigma\sigma^\top$. Our work focuses on two model classes, covering a large set of SDEs used in applications. The first class is the *elliptic* one, where we consider SDEs of the following form:

$$dX_t = V_{R,0}(X_t, \theta)dt + \sum_{1 \le j \le d} V_{R,j}(X_t, \theta)dB_{j,t}, \quad X_0 = x_0 \in \mathbb{R}^N, \tag{E}$$

so that $V_j = V_{R,j}$, $0 \le j \le d$. We set $\sigma_R = [V_{R,1}, \dots, V_{R,d}]$, $a_R = \sigma_R \sigma_R^\top$, and assume that $a_R = a_R(x, \theta)$ is positive definite for all $(x, \theta) \in \mathbb{R}^N \times \Theta$. Thus, w.l.o.g. here $d = N$. Class (E) includes a multitude of models used in applications, see e.g. Kloeden and Platen (1992). The second model class we work with is the *hypo-elliptic* one, where the SDE in (1) now splits into smooth and rough components as $X_t = (X_{S,t}, X_{R,t}) \in \mathbb{R}^{N_S + N_R}$, so that $N = N_S + N_R$, $N_S \ge 1$, $N_R \ge 1$, and we re-express (1) as:

$$\begin{aligned} dX_{S,t} &= V_{S,0}(X_t, \theta)\, dt; \qquad dX_{R,t} = V_{R,0}(X_t, \theta)\, dt + \sum_{1 \le j \le d} V_{R,j}(X_t, \theta)\, dB_{j,t}, \\ X_0 &= x_0 = (x_{S,0}, x_{R,0}) \in \mathbb{R}^{N_S + N_R}. \end{aligned} \tag{H}$$

In (H), the involved functions are defined as $V_{S,0} : \mathbb{R}^N \times \Theta \to \mathbb{R}^{N_S}$, $V_{R,j} : \mathbb{R}^N \times \Theta \to \mathbb{R}^{N_R}$, $0 \le j \le d$. Model class (H) stems from the generic form (1), where we now have that, for $(x, \theta) \in \mathbb{R}^N \times \Theta$:

$$V_0(x, \theta) = \left[V_{S,0}(x, \theta)^\top, V_{R,0}(x, \theta)^\top\right]^\top, \qquad V_j(x, \theta) = \left[\mathbf{0}_{N_S}^\top, V_{R,j}(x, \theta)^\top\right]^\top, \qquad 1 \le j \le d.$$

Notice that component $X_{S,t}$ is not driven by the Brownian motion, and consequently class (H) requires a separate treatment from (E). Later on, we introduce sufficient requirements associated with the *weak Hörmander's condition*, so that $V_{S,0}(X_t, \theta)$ depends on $X_{R,t}$, thus Brownian noise propagates into the smooth component, and the law of X_t, $t > 0$, admits a density w.r.t. the Lebesgue measure. Hypo-elliptic models are used in several application fields, including, e.g.: the FitzHugh-Nagumo SDE (DeVille, Vanden-Eijnden and Muratov, 2005) and the Jansen-Rit neural mass SDE (Ableidinger, Buckwar and Hinterleitner, 2017) in neuroscience; the underdamped or generalised Langevin equation (Pavliotis, 2014) in physics.

We consider parameter inference for SDEs given a collection of discrere-time data $\{X_t\}_{t \in \mathbb{T}_n}$, for the set of time instances $\mathbb{T}_n = \{t_0, t_1, \dots, t_n\}$, $n \ge 1$. For simplicity, we consider equidistant step-sizes, with $\Delta := t_i - t_{i-1}$. The likelihood function is given as:

$$L_n(\{X_t\}_{t \in \mathbb{T}_n}; \theta) = p(X_{t_0}; \theta) \prod_{1 \le i \le n} p_\Delta^X(X_{t_{i-1}}, X_{t_i}; \theta),$$

for some initial law $p(\cdot; \theta)$, where $x' \mapsto p_\Delta^X(x, x'; \theta)$ is the transition density of SDE (1), with the latter being in general unavailable in closed form. A practical standard approach to circumvent this intractability is by introducing a time-discretisation scheme and using the induced CF approximate transition density as a proxy for the true one. For instance, a common scheme is the *Euler-Maruyama* one, which yields a conditionally Gaussian approximate density upon application to elliptic SDEs. However, it is well-understood that such an approximation cannot correctly capture the true non-linear dynamics unless the step-size Δ is close to 0. Thus, parameter estimation relying on such a simple Gaussian approximation requires a *high-frequency* observation regime, with $\Delta \ll 1$. In practice, the step-size of available data is usually fixed, and its value may not be too small.

In the context of fixed Δ, the prominent work of Aït-Sahalia (2002) proposes an elaborate approximation of the transition density for time-homogeneous univariate elliptic SDEs via a CF Hermite-series expansion. Roughly, the expansion for the transition density is of the structure:

$$p_\Delta^X(x, y; \theta) \approx q_\Delta(x, y; \theta) \times \left\{1 + (\text{correction term})\right\}. \tag{2}$$

Here, $y \mapsto q_{\Delta}(x, y; \theta)$ is a 'baseline' tractable density. The 'correction term' is given in closed-form, and includes Hermite polynomials up to a degree $J \geq 1$, obtained via working with $q_{\Delta}(x, y; \theta)$. The correction term plays a key role in capturing non-linear/non-Gaussian effects in the true transitions. In detail, Aït-Sahalia (2002) constructs the CF-expansion by first applying an 1–1 'Lamperti transform' (Roberts and Stramer, 2001), thus replacing the original scalar X_t with a process Y_t of unit diffusion coefficient, and then obtaining the Hermite series expansion for the transition density of Y_t. We refer to this line of research as the *Hermite approach*. Aït-Sahalia (2002) proves convergence of the CF-expansion to the true density for fixed $\Delta \in (0, 1)$ as the degree of Hermite polynomials, J, grows to infinity. The result is a qualitative one, as no order of convergence is provided. The Hermite approach works only for the sub-class of 'reducible' elliptic SDEs for which the Lamperti transform is applicable. Also, as stated in Aït-Sahalia (2008), convergence of the Hermite series expansion is not guaranteed when back-transforming onto the original density of X_t. To treat a wider class of non-reducible multivariate elliptic SDEs, Aït-Sahalia (2008) utilises the Kolmogorov backward/forward equations (PDEs) to construct a series expansion in Δ and $y - x$. No analytical results are provided for fixed Δ. We refer to this contribution as the *PDE approach*. Li (2013) develops a *probabilistic approach*, by making use of Malliavin calculus and carrying out an asymptotic analysis of Wiener functionals (Watanabe, 1987, Yoshida, 1992) to obtain a CF-expansion, accompanied by an analytic bound for the approximation error, for fixed Δ. The expansion is given in terms of powers $\Delta^{j/2}$, $j \geq 0$, for $\Delta \in (0, 1)$. More precisely:

$$p_{\Delta}^{X}(x, y; \theta) = q_{\Delta}(x, y; \theta) \times \Big\{ 1 + \sum_{1 \leq j \leq J} \Delta^{j/2} \cdot e_{\Delta}^{(j)}(x, y; \theta) \Big\} + R(\Delta, x, y; \theta), \qquad J \geq 1, \tag{3}$$

for tractable coefficients $e_{\Delta}^{(j)}(\cdot)$, $j \geq 1$, and a remainder term $R(\cdot)$. Li (2013) proves under conditions that the remainder is of size $O(\Delta^{(J+1-N)/2})$. The probabilistic approach is extended to elliptic SDEs with jumps in Li and Chen (2016). For time-inhomogeneous elliptic SDEs, Choi (2015) develops a CF-expansion via the PDE approach, similarly to Aït-Sahalia (2008). Yang, Chen and Wan (2019) use Itô-Taylor expansions and obtain a series of the form (3) that involves Hermite polynomials, with explicit bounds provided on residuals as in Li (2013). Even if alternative approaches have been followed in the literature, the produced expansions are closely related to each other. E.g., one can obtain a series expansion as in (3) involving Hermite polynomials via the two different approaches in Li (2013), Yang, Chen and Wan (2019). Furthermore, Lee, Song and Lee (2014) show that the Hermite expansion of Aït-Sahalia (2002) can be expressed in the form (3) by rearranging terms in the expansion w.r.t. powers $\Delta^{j/2}$, $j \geq 1$.

Importantly, for developed CF-expansions to be *theoretically validated*, the remainder terms should be controlled and vanish. This property guarantees convergence of the expansion, with a rate in Δ that grows when more terms are used in the expansion. As mentioned, such an elaborate analysis has been carried out in Li (2013), Yang, Chen and Wan (2019) in the context of elliptic SDEs.

The aforementioned works also demonstrate the effective use of a CF-expansion within parameter inference procedures. In particular, the approaches provide an approximate Maximum Likelihood Estimator (MLE). Obtained numerical results showcase that: the proxy MLEs stays close to the true ones even when the step-size Δ is not too small; the CF-expansions outperform proxy methods based on Gaussian-type quasi-likelihoods. Chang and Chen (2011) provide analytical consistency and convergence rate results for the proxy MLE, and demonstrate good performance of their CF-expansion by clarifying the effect of the length of the expansion and of the fixed step-size $\Delta \in (0, 1)$. In the context of Bayesian inference for SDEs, Stramer, Bognar and Schneider (2010) utilise the expansion-based likelihood and show advantages over the Gaussian-type (Euler-Maruyama-based) likelihood.

Critically, the development of CF-expansions in the literature is so far restricted to elliptic SDEs and a limited class of hypo-elliptic ones, e.g., with linear drift and constant diffusion coefficient (Barilari and Paoli, 2017, Habermann, 2019), thus general hypo-elliptic SDEs specified as (H) have yet to be covered even though the latter are widely used in applications. Available methods for elliptic SDE build upon steps that cannot be readily extended to the hypo-elliptic setting. In brief, one limitation derives from the definition of the reference Gaussian density $q_\Delta(x, y; \theta)$ in the expansion relying on positive definiteness of its covariance matrix, when such a property is violated within the hypo-elliptic class (H). Our work develops a novel CF-expansion that covers both elliptic and hypo-elliptic SDEs in a unified framework. To this end, we consider a *non-degenerate* baseline Gaussian density $q_\Delta(x, y; \theta)$ that is well-defined for both SDE classes, (E) and (H). We then construct a CF-expansion in the form of (3) based on such a well-posed $q_\Delta(x, y; \theta)$. We emphasise that the error analysis is much more challenging in the hypo-elliptic setting than in the elliptic one, due to varying scales across the SDE co-ordinates. We manage to provide analytical error estimates for the proposed expansion by utilising a recent result on estimates of the transition density for degenerate SDEs (Pigato, 2022), thus theoretically validating our CF-expansion both within the elliptic and the hypo-elliptic classes of SDEs.

Beyond the above-mentioned literature on CF-expansions for elliptic SDEs with fixed $\Delta \in (0, 1)$, our work is also motivated by several recent developments in the area of parametric inference for hypo-elliptic SDEs, albeit in a *high-frequency observation regime*, i.e. $n \to \infty$, $\Delta = \Delta_n \to 0$, $n\Delta_n \to \infty$, together with an extra 'design' condition on Δ_n. Indicatively, Ditlevsen and Samson (2019), Gloter and Yoshida (2021), Melnykova (2020), Pilipovic, Samson and Ditlevsen (2024) propose contrast estimators, under the design condition $\Delta_n = o(n^{-1/2})$. The latter is weakened to $\Delta_n = o(n^{-1/3})$ and $\Delta_n = o(n^{-1/p})$, for a general integer $p \geq 2$, by Iguchi, Beskos and Graham (2025) and Iguchi and Beskos (2025a), respectively. Iguchi, Beskos and Graham (2024) also treat a class of 'highly degenerate' hypo-elliptic SDEs.

Our main contributions are briefly summarised as follows:

a. We propose a CF-expansion for the transition density of both elliptic and hypo-elliptic SDEs, in (E) and (H), respectively. Within the elliptic class, a starting point for developing the CF-expansion is motivated by the work of one of the co-authors in Iguchi and Yamada (2021). This latter work lies in the area of numerical methods for SDEs and looks at the development of approximation schemes for elliptic SDEs of improved weak order of convergence. To the best of our knowledge, this is the first time in the literature that a CF-expansion is obtained for hypo-elliptic SDEs with a general form of coefficients.
b. Our proposed CF-expansion involves a linear combination of differential operators acting on an appropriately chosen baseline Gaussian density, thus is easily computable via available software with symbolic calculations. Though we initially obtain an expression of different structure from (3), we later show that our CF-expansion indeed takes up the form of (3), i.e. a series expansion in powers of $\sqrt{\Delta}$. Thus, our CF-expressions align with existing works for elliptic SDEs.
c. We theoretically validate our CF expansions by proving analytically, under appropriate conditions, that the residuals are of size $O(\Delta^{K/2})$ for a step-size $\Delta \in (0, 1)$, where $K \geq 1$ is an integer differing between the elliptic and hypo-elliptic classes, and which depends on the model dimension N. In particular, the effect of the dimensionality varies amongst the two SDE classes.
d. We present numerical results showcasing that the use of the proposed CF-expansion leads to effective parameter estimation for SDEs, with an emphasis on hypo-elliptic models. In particular, we conduct Bayesian inference for a real dataset and show that the posterior distribution is accurately estimated by the proposed density expansion.

The structure of the paper is as follows. In Section 2 we outline our strategy for constructing a CF-expansion which covers both (E), (H), and then proceed with the development of the expansion. Section 3 provides a rigorous error analysis for the proposed CF-expansion, separately for classes (E) and (H). Section 4 shows numerical applications, and the codes that reproduce the results are available at https://github.com/YugaIgu/CF-density-expansion. Section 6 provides a summary and conclusions. Most proofs are collected in a Supplementary Material.

Notation. We set $\mathbb{Z}_{\geq 0} = \mathbb{N} \cup \{0\}$. For a multi-index $\alpha \in \mathbb{Z}^k_{\geq 0}$, $k \in \mathbb{N}$, we write $\|\alpha\| = k$, $|\alpha| = \sum_{1 \leq j \leq k} \alpha_j$, $\|\alpha\|_\infty = \max_{1 \leq i \leq k} \alpha_i$, $\alpha! = \prod_{j=1}^k \alpha_j$. For $\alpha \in \mathbb{Z}^m_{\geq 0}$ and a sufficiently smooth $\varphi : \mathbb{R}^m \to \mathbb{R}$, $m \in \mathbb{N}$, we write $\partial^\alpha \varphi(y) = \partial^{\alpha_1}_{y_1} \cdots \partial^{\alpha_m}_{y_m} \varphi(y)$, where $\partial^{\alpha_i}_{y_i} \equiv \partial^{\alpha_i} / \partial y_i^{\alpha_i}$. We often write $\partial^\alpha_y \varphi(y) \equiv \partial^\alpha \varphi(y)$ to emphasise the argument upon which the derivative acts. The generator associated with SDE (1) writes as:

$$\mathscr{L}_\theta \varphi(x) = \sum_{1 \leq i \leq N} V^i_0(x, \theta)\, \partial_i \varphi(x) + \tfrac{1}{2} \sum_{1 \leq i_1, i_2 \leq N} \sum_{1 \leq j \leq d} V^{i_1}_j(x, \theta) V^{i_2}_j(x, \theta)\, \partial_{i_1 i_2} \varphi(x), \tag{4}$$

$(x, \theta) \in \mathbb{R}^N \times \Theta$, for $\varphi : \mathbb{R}^N \to \mathbb{R}$, where we use integer superscripts to indicate co-ordinates in vectors. For $x \in \mathbb{R}$, we write $\lfloor x \rfloor = \max\{m \in \mathbb{Z} \,|\, m \leq x\}$. For differential operators D_1, D_2, we define the *commutator* as $\mathrm{ad}_{D_1}(D_2) = [D_1, D_2] \equiv D_1 D_2 - D_2 D_1$. The k-times iteration of the commutator writes as $\mathrm{ad}^k_{D_1}(D_2) = [D_1, \mathrm{ad}^{k-1}_{D_1}(D_2)]$, $k \geq 1$, with $\mathrm{ad}^0_{D_1}(D_2) = D_2$.

2. Closed-Form Transition Density Expansion

We will present a new CF transition density expansion for a wide class of Itô processes in (1), including the family of hypo-elliptic SDEs specified in (H). We write the transition density of $X_{t+\Delta}$ given $X_t = x \in \mathbb{R}^N$ as $y \mapsto p^X_\Delta(x, y; \theta) := \mathbb{P}(X_{t+\Delta} \in dy \,|\, X_t = x)/dy$, with $t \geq 0$, $\Delta > 0$.

2.1. Conditions for Closed-Form Expansion

Assumption 2.1. For both model classes (E) and (H), the maps $x \mapsto V_j(x, \theta)$, $0 \leq j \leq d$, are infinitely differentiable for any $\theta \in \Theta$.

For a vector-valued $V : \mathbb{R}^N \to \mathbb{R}^N$, we make use of the standard correspondence $V \leftrightarrow \sum_{i=1}^N V^i \partial_i$.

Assumption 2.2. We distinguish between model classes (E) and (H).

I. For class (E), it holds that $a_R(x, \theta) = (\sigma_R \sigma_R)^\top(x, \theta)$ is positive definite for all $(x, \theta) \in \mathbb{R}^N \times \Theta$. This is equivalent to $\mathrm{Span}\{V_{R,j}(x, \theta),\ 1 \leq j \leq d\} = \mathbb{R}^N$, for all $(x, \theta) \in \mathbb{R}^N \times \Theta$.

II. For class (H), it holds that:

$$\mathrm{Span}\big\{V_{R,j}(x, \theta),\ 1 \leq j \leq d\big\} = \mathbb{R}^{N_R}, \qquad \mathrm{Span}\Big\{\big\{V_j(x, \theta),\ [\widetilde{V}_0, V_j](x, \theta)\big\},\ 1 \leq j \leq d\Big\} = \mathbb{R}^N,$$

for all $(x, \theta) \in \mathbb{R}^N \times \Theta$, where $\widetilde{V}_0 : \mathbb{R}^N \times \Theta \to \mathbb{R}^N$ is the drift function when the Itô SDE (H) is written in a Stratonovich form, namely $\widetilde{V}_0(x, \theta) = V_0(x, \theta) - \frac{1}{2} \sum_{j=1}^d \sum_{i=1}^N V^i_j(x, \theta) \partial_{x_i} V_j(x, \theta)$.

Assumption 2.2 is related to *Hörmander's condition* (it suffices for Hörmander's condition to hold) and implies that the law of X_t, $t > 0$, admits a Lebesgue density. For the hypo-elliptic class (H), Assumption 2.2-II guarantees that the noise in the rough component $X_{R,t}$ (of size $\sqrt{t}$ for a period of length t) propagates into the smooth component $X_{S,t}$. Inclusion of the vector fields $[\widetilde{V}_0, V_j](x, \theta)$, $1 \leq j \leq d$, relates to the appearance of terms $\int_0^t B^j_s ds$ (of a different scale $\sqrt{t^3}$) in the smooth component $X_{S,t}$ after an Itô-Taylor expansion of $\int_0^t V_{S,0}(X_u, \theta) du$. Thus, the transition law of SDE (H) is non-degenerate and admits a Lebesgue density even if not all coordinates are directly driven by the Brownian motion. A precise definition of Hörmander's condition can be found, e.g., in Nualart (2006).

2.2. Background Idea

Before presenting the CF-expansion we explain an idea that underpins its development – more precisely the starting point of the latter. Consider the elliptic class (E) and the Euler-Maruyama (EM) scheme which approximates the transition dynamics of $X_{t+\Delta}|X_t = x$, with $x \in \mathbb{R}^N$, $t > 0$, $\Delta > 0$, so that:

$$\bar{X}^{\mathrm{EM},\theta}_{t+\Delta} := x + V_{R,0}(x,\theta)\Delta + \sigma_R(x,\theta)\big(B_{t+\Delta} - B_t\big). \tag{5}$$

Under regularity conditions on $x \mapsto V_{R,j}(x,\theta)$, $0 \le j \le d$, and the requirement that the matrix $a_R(x,\theta) = (\sigma_R\sigma_R^\top)(x,\theta)$ is positive definite for all $(x,\theta) \in \mathbb{R}^N \times \Theta$, the EM scheme gives rise to a well-defined baseline Gaussian transition density, $y \mapsto p^{\bar{X}^{\mathrm{EM}}}_{\Delta}(x,y;\theta)$. In the present elliptic setting Iguchi and Yamada (2021) constructed a CF transition density approximation of the following form:

$$p^X_\Delta(x,y;\theta) \approx p^{\bar{X}^{\mathrm{EM}}}_{\Delta}(x,y;\theta) \times \big(1 + (\text{correction term})\big). \tag{6}$$

The tools utilised in Iguchi and Yamada (2021) to derive the approximation include Taylor expansion, Kolmogorov backward/forward equations, use of the infinitesimal generators for the target SDE and its EM approximation. In the above expression, the 'correction term' involves Δ, partial derivatives of the SDE coefficients and Hermite polynomials obtained via differentiating the transition density of the EM scheme. As mentioned in the Introduction, other approaches are also available, including the ones developed in Aït-Sahalia (2002, 2008), Li (2013), Yang, Chen and Wan (2019), and all such works also assume invertibility of the matrix a_R, thus are not relevant for the hypo-elliptic class (H).

To construct a CF transition density expansion for a broader family of SDEs that includes hypo-elliptic SDEs, it is critical to choose an appropriate reference Gaussian density which is *non-degenerate* for the target class of models. To achieve this, we consider the *local drift linearisation (LDL) scheme*, which, upon application on the general SDE model in (1), is defined via the following expression, for each given $t \ge 0$, $\Delta > 0$, and for $\bar{X}^\theta_t = x \in \mathbb{R}^N$:

$$\bar{X}^\theta_{t+\Delta} = x + \int_t^{t+\Delta} \big(A_{x,\theta}\bar{X}^\theta_s + b_{x,\theta}\big) ds + \sigma(x,\theta)\big(B_{t+\Delta} - B_t\big), \tag{7}$$

where $A_{x,\theta} \in \mathbb{R}^{N\times N}$ and $b_{x,\theta} \in \mathbb{R}^N$ are specified as follows:

$$A_{x,\theta} = \big[\partial_{x_j} V_0^i(x,\theta)\big]_{1\le i,j\le N}, \qquad b_{x,\theta} = V_0(x,\theta) - A_{x,\theta}\,x.$$

That is, (7) is obtained from a 1st-order Taylor expansion of the drift V_0 about the initial position x and $\sigma(\cdot)$ fixed at its initial value. Expression (7) corresponds to a linear SDE, with a solution for $\bar{X}^\theta_{t+\Delta}|\bar{X}^\theta_t = x$ that has the following explicit form:

$$\bar{X}^\theta_{t+\Delta} = e^{\Delta A_{x,\theta}}x + \int_t^{t+\Delta} e^{(t+\Delta-s)A_{x,\theta}} b_{x,\theta}\, ds + \int_t^{t+\Delta} e^{(t+\Delta-s)A_{x,\theta}} \sigma(x,\theta)\, dB_s.$$

Thus, $\bar{X}^\theta_{t+\Delta}|\bar{X}^\theta_t = x$ follows a Gaussian law, with mean $\mu(\Delta,x,\theta)$ and covariance $\Sigma(\Delta,x,\theta)$ given as:

$$\mu(\Delta,x,\theta) = e^{\Delta A_{x,\theta}}\hat{\mu}^x(\Delta,x,\theta), \qquad \hat{\mu}^z(\Delta,x,\theta) := x + \int_0^\Delta e^{-sA_{z,\theta}} b_{z,\theta} ds, \quad z \in \mathbb{R}^N; \tag{8}$$

$$\Sigma(\Delta,x,\theta) = e^{\Delta A_{x,\theta}}\hat{\Sigma}(\Delta,x,\theta)e^{\Delta A^\top_{x,\theta}}, \qquad \hat{\Sigma}(\Delta,x,\theta) := \int_0^\Delta e^{-sA_{x,\theta}} a(x,\theta) e^{-sA^\top_{x,\theta}} ds, \tag{9}$$

where we recall $a = \sigma\sigma^\top$. The introduction of the extra argument z in $\hat{\mu}^z(\Delta, x, \theta)$ will be of use in later developments.

In the sequel we show that $\hat{\Sigma}$ (thus also Σ) is positive definite for both model classes (E), (H) under Assumption 2.2. Then, under regularity conditions on the SDE coefficients together with the invertibility Σ, we appropriately expand upon the direction followed by Iguchi and Yamada (2021) to construct a CF transition density approximation that covers the model class (H) and writes as:

$$p_\Delta^X(x, y; \theta) \approx p_\Delta^{\bar{X}}(x, y; \theta) \times \big(1 + (\text{correction term})\big), \tag{10}$$

where $y \mapsto p_\Delta^{\bar{X}}(x, y; \theta)$ is the transition density of the LDL scheme (7). Similarly to the case of the expansion for elliptic diffusions, the correction term appearing in (10) involves partial derivatives of SDE coefficients w.r.t. the state argument and Hermite polynomials now defined via partial derivatives of the non-degenerate Gaussian density $p_\Delta^{\bar{X}}(x, y; \theta)$.

Remark 2.3. Iguchi and Yamada (2021) work in an elliptic setting to develop Monte-Carlo estimators of improved weak order of convergence for $\mathbb{E}[\varphi(X_T)]$, $\varphi : \mathbb{R}^N \to \mathbb{R}$, $T > 0$, and their expansion in the form of (6) is used for such a purpose. In brief, they use samples from the baseline $p_\Delta^{\bar{X}^{\text{EM}}}(x, y; \theta)$, weighted by $(1 + (\text{correction term}))$, in an iterative procedure over $\lfloor T/\Delta \rfloor$ steps. Even if the initial derivations in the CF-expansion we develop here resemble steps followed in Iguchi and Yamada (2021), our objectives and, consequently, the structure of the CF-expansion and its theoretical analysis (and, in general, the overall contribution) fully deviate from Iguchi and Yamada (2021).

Remark 2.4. LDL scheme (7) differs from the so-called Local Linearisation (LL) scheme (its definition can be found, e.g., in Jimenez, Mora and Selva (2017)) in the sense that the latter applies a first order Taylor expansion for both drift and diffusion coefficients. As shown in the next subsection, in particular in Lemma 2.5, the LDL scheme follows conditionally a non-degenerate Gaussian distribution that admits a transition density for (E) and (H) under Assumptions 2.1-2.2, leading to the development of (10). The key idea here is that the noise in the rough component X_R propagates to the smooth component X_S via the locally linearised drift. Similarly, the LL scheme can be shown to admit a well-defined tractable Lebesgue density, thus it could also form the basis of a transition density expansion like (10). In our analysis below, we employ the LDL scheme since its density admits a simpler expression and suffices to build a tractable density expansion. Also, such a non-degenerate Gaussian approximation can be constructed via a drift linearisation for partial coordinates (not full) so that the matrix $A_{x,\theta}$ is upper-triangular, thus reducing the computation cost of calculating $\exp\{A_{x,\theta}\}$. We briefly discuss a practical choice of A in the numerical experiment Section 4.1 and in Section 5 as well.

2.3. Non-Degeneracy of the LDL Scheme

As mentioned in Section 2.2, existence of a Lebesgue density for the transition dynamics of the LDL scheme (7) is essential for the construction of our CF-expansion for both model classes (E) and (H). In this section we show that such an existence is implied by Assumption 2.2. That is, we show under Assumption 2.2, i.e. Hörmander's condition for the target model classes (E) and (H), that the vector fields defined via the LDL scheme (7) also satisfy Hörmander's condition. Specifically, the vector fields defined from the coefficients of (7) coincide with those of the original SDE, upon fixing the argument x to the initial condition of (7). We thus have the following result whose proof is given in Appendix A.

Lemma 2.5. *Let $(\Delta, x, \theta) \in (0,\infty) \times \mathbb{R}^N \times \Theta$. Under Assumptions 2.1-2.2, the law of $\bar{X}^{\theta}_{t+\Delta} | \bar{X}^{\theta}_t = x$ defined in (7) admits a smooth Lebesgue density for model classes (E) and (H).*

We describe a sub-class from (H) where the LDL scheme delivers well-posed Lebesgue densities for SDE transition dynamics while the Euler-Maruyama scheme provides degenerate distributions.

Example (Underdamped Langevin Equation). We consider the following bivariate SDE:

$$dX^1_t = X^2_t \, dt; \qquad dX^2_t \quad = \big(-V'(X^1_t) - \alpha X^2_t\big) dt + \sigma \, dB_{1,t}, \tag{11}$$

for parameters $\alpha, \sigma > 0$ and potential $V : \mathbb{R} \to \mathbb{R}$. Such dynamics are used to describe the motion of a particle on the real line $\mathbb{R}$, with X^1_t and X^2_t representing position and momentum, respectively. The coefficients in SDE (11) correspond to the following vector fields, for $x = (x_1, x_2) \in \mathbb{R}^2$:

$$V_0 \equiv \widetilde{V}_0 = x_2 \partial_{x_1} + \big(-V'(x_1) - \alpha x_2\big) \partial_{x_2}, \qquad V_1 = \sigma \partial_{x_2}.$$

The diffusion matrix is degenerate, so (11) belongs to class (H). Also, Assumption 2.2-II is satisfied as:

$$V_1(x) = [0, \sigma]^\top, \qquad [\widetilde{V}_0, V_1](x) = \widetilde{V}_0 V_1(x) - V_1 \widetilde{V}_0(x) = [-\sigma, \sigma\alpha]^\top, \tag{12}$$

and, given $\sigma, \alpha > 0$, we get that $\mathrm{Span}\{V_1(x), [V_0, V_1](x)\} = \mathbb{R}^2$, for all $x \in \mathbb{R}^2$. The Euler-Maruyama scheme for $\bar{X}^{\mathrm{EM}}_{t+\Delta} | \bar{X}^{\mathrm{EM}}_t = x$ writes as:

$$\bar{X}^{\mathrm{EM},1}_{t+\Delta} = x_1 + x_2 \Delta; \qquad \bar{X}^{\mathrm{EM},2}_{t+\Delta} = x_2 + \big(-V'(x_1) - \alpha x_2\big)\Delta + \sigma (B_{1,t+\Delta} - B_{1,t}). \tag{13}$$

So, the law of $\bar{X}^{\mathrm{EM}}_{t+\Delta} | \bar{X}^{\mathrm{EM}}_t$ involves a degenerate covariance matrix. In contrast, in this setting the LDL scheme (7) contains the 2×2 matrix A_x and the vector b_x specified as follows:

$$[A_x]_{11} = 0, \;\; [A_x]_{12} = 1, \;\; [A_x]_{21} = -V''(x_1), \;\; [A_x]_{22} = -\alpha, \quad b^1_x = 0, \;\; b^2_x = -V'(x_1) + V''(x_1) x_1.$$

The vector fields associated with the coefficients of the LDL scheme are given as follows, for $x, z \in \mathbb{R}^2$:

$$\bar{V}^z_0 = \sum_{i=1,2} [A_z x + b_z]_i \partial_{x_i}, \qquad \bar{V}_1 = \sigma \partial_{x_2},$$

where (with some abuse of notation) we introduce $z \in \mathbb{R}^2$ to represent the initial condition for (7), thus distinguish the latter from argument $x \in \mathbb{R}^2$ upon which the linear drift in (7) applies. For the above vector fields, Hörmander's condition holds via:

$$\bar{V}_1(x) = \big[0, \sigma\big]^\top, \qquad [\bar{V}^z_0, \bar{V}_1](x) = -\bar{V}_1 \bar{V}^z_0(x) = \big[-\sigma, \sigma\alpha\big]^\top. \tag{14}$$

Thus, the law of $\bar{X}_{t+\Delta} | \bar{X}_t$ admits a well-defined (Gaussian) transition density. Note that the vector fields in (14) coincide with the ones in (12) defined for the original SDE (11), so the hypo-ellipticity of the target SDE is inherited by its induced LDL scheme.

2.4. Transition Density CF Expansion

2.4.1. Preliminaries

We prepare some ingredients for the construction of our CF-expansion. We define the LDL scheme starting from a point $x \in \mathbb{R}^N$ with its coefficients being frozen at a point $z \in \mathbb{R}^N$ as:

$$d\bar{X}^{\theta,z}_t = \big(A_{z,\theta} \bar{X}^{\theta,z}_t + b_{z,\theta}\big) dt + \sigma(z,\theta) dB_t, \qquad \bar{X}^{\theta,z}_0 = x_0 \in \mathbb{R}^N. \tag{15}$$

Notice that $\left[\bar{X}_t^{\theta,z}|_{z=x} \,\middle|\, \bar{X}_0^{\theta,z} = x\right] \equiv [\bar{X}_t^{\theta} | \bar{X}_0^{\theta} = x]$. The generator corresponding to (15) is given as:

$$\mathscr{L}_\theta^{0,z}\varphi(x) = \sum_{1\le i\le N} \left[A_{z,\theta}x + b_{z,\theta}\right]_i \partial_i \varphi(x) + \tfrac{1}{2}\sum_{1\le i_1,i_2\le N}\sum_{1\le j\le d} V_j^{i_1}(z,\theta)V_j^{i_2}(z,\theta)\partial_{i_1 i_2}\varphi(x),$$

for $\varphi : \mathbb{R}^N \to \mathbb{R}$, $x, z \in \mathbb{R}^N$, $\theta \in \Theta$. Notice that $\bar{X}_{t+\Delta}^{\theta,z} | \bar{X}_t^{\theta,z} = x \sim \mathscr{N}\left(e^{\Delta A_{z,\theta}}\hat{\mu}^z(\Delta, x, \theta), \Sigma(\Delta, z, \theta)\right)$, where $\hat{\mu}^z$ and Σ are defined in (8) and (9), respectively. We write the density of $\bar{X}_{t+\Delta}^{\theta,z} | \bar{X}_t^{\theta,z} = x$ as $y \mapsto p_\Delta^{\bar{X}^z}(x, y; \theta)$ and note that $p_\Delta^{\bar{X}^z}(x, y; \theta)|_{z=x} \equiv p_\Delta^{\bar{X}}(x, y; \theta)$, where the right-hand-side is the transition density of the LDL scheme (7).

We introduce semi-groups $\{P_t^\theta\}_{t\ge 0}$ and $\{\bar{P}_t^{\theta,z}\}_{t\ge 0}$ associated with the Markov processes $\{X_t\}_{t\ge 0}$ and $\{\bar{X}_t^{\theta,z}\}_{t\ge 0}$, respectively as follows:

$$P_t^\theta \varphi(x) = \int_{\mathbb{R}^N} \varphi(y) p_t^X(x, y; \theta) dy, \qquad \bar{P}_t^{\theta,z}\varphi(x) = \int_{\mathbb{R}^N}\varphi(y) p_t^{\bar{X}^z}(x, y; \theta) dy, \quad z \in \mathbb{R}^N, \tag{16}$$

for $\varphi : \mathbb{R}^N \to \mathbb{R}$ and $(t, x, \theta) \in (0, \infty) \times \mathbb{R}^N \times \Theta$. For notational simplicity, we introduce:

$$\widetilde{\mathscr{L}}_\theta^z := \mathscr{L}_\theta - \mathscr{L}_\theta^{0,z}, \tag{17}$$

where we recall that $\mathscr{L}_\theta$ is the generator associated with the target SDE, given in (4). The first steps in the derivation of our CF-expansion are provided in the following two results whose proofs are provided in Appendices B and C:

Lemma 2.6. *Let $t > 0$, $x, y \in \mathbb{R}^N$ and $\theta \in \Theta$. Also, let $\varphi \in C^\infty(\mathbb{R}^N; \mathbb{R})$. It holds that:*

$$p_t^X(x, y; \theta) = p_t^{\bar{X}^z}(x, y; \theta)|_{z=x} + \int_0^t P_s^\theta \widetilde{\mathscr{L}}_\theta^z p_{t-s}^{\bar{X}^z}(\cdot, y; \theta)(x)|_{z=x} ds; \tag{18}$$

$$P_t^\theta\varphi(x) = \bar{P}_t^{\theta,z}\varphi(x)|_{z=x} + \int_0^t P_s^\theta \widetilde{\mathscr{L}}_\theta^z \bar{P}_{t-s}^{\theta,z}\varphi(x)|_{z=x} ds. \tag{19}$$

Lemma 2.7. *Let $0 < s < t$, $\theta \in \Theta$ and $z, x, y \in \mathbb{R}^N$. Then it holds that, for any $K \in \mathbb{N}$,*

$$\bar{P}_s^{\theta,z}\widetilde{\mathscr{L}}_\theta^z \bar{P}_{t-s}^{\theta,z}\varphi(x) = \sum_{0\le k\le K}\frac{s^k}{k!}\left\{\mathrm{ad}^k_{\mathscr{L}_\theta^{0,z}}(\widetilde{\mathscr{L}}_\theta^z)\right\}\bar{P}_t^{\theta,z}\varphi(x) + \mathscr{R}^{K+1,z}(s, t, x; \theta); \tag{20}$$

$$\bar{P}_s^{\theta,z}\widetilde{\mathscr{L}}_\theta^z p_{t-s}^{\bar{X}^z}(\cdot, y; \theta)(x) = \sum_{0\le k\le K}\frac{s^k}{k!}\left\{\mathrm{ad}^k_{\mathscr{L}_\theta^{0,z}}(\widetilde{\mathscr{L}}_\theta^z)\right\}\{p_t^{\bar{X}^z}(\cdot, y; \theta)\}(x) + \widetilde{\mathscr{R}}^{K+1,z}(s, t, x, y; \theta) \tag{21}$$

where the remainder terms $\mathscr{R}^{K+1,z}$ and $\widetilde{\mathscr{R}}^{K+1,z}$ are specified in the proof of Lemma 2.7 in Appendix C.

Results similar to the above, for the elliptic case and for the Euler-Maruyama scheme used as a baseline transition density, are obtained in Iguchi and Yamada (2021).

2.4.2. *Construction of CF-Expansion*

Based on the auxiliary results (Lemma 2.5, 2.6 and 2.7) in the previous subsections, we construct a CF transition density expansion in the following three steps:

Step 1. We recursively apply formula (19) within (18), from Lemma 2.6, to obtain for any $M \in \mathbb{N}$:

$$\begin{aligned} p_\Delta^X(x,y;\theta) &= p_\Delta^{\bar{X}^z}(x,y;\theta)|_{z=x} + \int_0^\Delta P_{s_1}^\theta \widetilde{\mathscr{L}}_\theta^z p_{\Delta-s_1}^{\bar{X}^z}(\cdot,y;\theta)(x)|_{z=x} ds_1 \\ &= p_\Delta^{\bar{X}^x}(x,y;\theta) + \sum_{1\le j\le M-1} \mathscr{T}^j(\Delta,x,y;\theta) \;+\; \mathscr{R}_1^M(\Delta,x,y;\theta), \end{aligned} \tag{22}$$

where we have set:

$$\begin{aligned} \mathscr{T}^j(\Delta,x,y;\theta) &= \int_{I(s_{1:j})} \bar{P}_{s_j}^{\theta,z} \widetilde{\mathscr{L}}_\theta^z \bar{P}_{s_{j-1}-s_j}^{\theta,z} \cdots \widetilde{\mathscr{L}}_\theta^z \bar{P}_{s_1-s_2}^{\theta,z} \widetilde{\mathscr{L}}_\theta^z p_{\Delta-s_1}^{\bar{X}^z}(\cdot,y;\theta)(x)\big|_{z=x} ds_j \cdots ds_1; \\ \mathscr{R}_1^M(\Delta,x,y;\theta) &= \int_{I(s_{1:M})} P_{s_M}^\theta \widetilde{\mathscr{L}}_\theta^z \bar{P}_{s_{M-1}-s_M}^{\theta,z} \cdots \widetilde{\mathscr{L}}_\theta^z \bar{P}_{s_1-s_2}^{\theta,z} \widetilde{\mathscr{L}}_\theta^z p_{\Delta-s_1}^{\bar{X}^z}(\cdot,y;\theta)(x)\big|_{z=x} ds_M \cdots ds_1, \end{aligned} \tag{23}$$

$$I(s_{1:k}) := \{s_{1:k} = (s_1,\ldots,s_k) : 0 \le s_k \le \cdots \le s_1 \le \Delta\}, \quad k \ge 0,$$

with the convention $s_0 \equiv \Delta$.
Step 2. Since $\mathscr{T}^j(\Delta,x,y;\theta)$, $1 \le j \le M$, is not tractable, we obtain a computable quantity for it via use of Lemma 2.7. Let $\mathscr{T}_{s_{1:j}}(\Delta,x,y;\theta)$ be the integrand of $\mathscr{T}^j(\Delta,x,y;\theta)$, so that $\mathscr{T}^j(\Delta,x,y;\theta) = \int_{I(s_{1:j})} \mathscr{T}_{s_{1:j}}(\Delta,x,y;\theta) ds_j \cdots ds_1$. Recursive application of Lemma 2.7 to $\mathscr{T}_{s_{1:j}}$, $1 \le j \le M-1$, gives:

$$\mathscr{T}_{s_{1:j}}(\Delta,x,y;\theta) = \sum_{\alpha \le \beta^{[j]}} \frac{\prod_{k=1}^j (s_k)^{\alpha_k}}{\alpha!} \cdot \mathscr{D}_\alpha^{z,\theta}\{p_\Delta^{\bar{X}^z}(\cdot,y;\theta)\}(x)\big|_{z=x} + \mathscr{E}_{s_{1:j}}^{\beta^{[j]}}(\Delta,x,y;\theta), \tag{24}$$

for a multi-index $\beta^{[j]} = (\beta_1^{[j]},\ldots,\beta_j^{[j]}) \in \mathbb{Z}_{\ge 0}^j$, where we have defined:

$$\mathscr{D}_\alpha^{z,\theta} \equiv \Big(\mathrm{ad}^{\alpha_j}_{\mathscr{L}_\theta^{0,z}}(\widetilde{\mathscr{L}}_\theta^z)\Big)\Big(\mathrm{ad}^{\alpha_{j-1}}_{\mathscr{L}_\theta^{0,z}}(\widetilde{\mathscr{L}}_\theta^z)\Big)\cdots\Big(\mathrm{ad}^{\alpha_1}_{\mathscr{L}_\theta^{0,z}}(\widetilde{\mathscr{L}}_\theta^z)\Big), \qquad z \in \mathbb{R}^N,\ \theta \in \Theta, \tag{25}$$

and the residual $\mathscr{E}_{s_{1:j}}^{\beta^{[j]}}(\Delta,x,y;\theta)$ is given in (A.54) of Supplementary Material. We now obtain, for $1 \le j \le M-1$:

$$\mathscr{T}^j(\Delta,x,y;\theta) = \sum_{\alpha \le \beta^{[j]}} \Delta^{|\alpha|+j} \cdot \frac{K(\alpha)}{\alpha!} \cdot \mathscr{D}_\alpha^{z,\theta}\{p_\Delta^{\bar{X}^z}(\cdot,y;\theta)\}(x)|_{z=x} + \mathscr{R}_2^{j,\beta^{[j]}}(\Delta,x,y,\theta), \tag{26}$$

$$\begin{aligned} K(\alpha) &:= \int_{0\le s_j\le\cdots\le s_1\le 1} \prod_{1\le k\le j} (s_k)^{\alpha_k} ds_j \cdots ds_1, \\ \mathscr{R}_2^{j,\beta^{[j]}}(\Delta,x,y,\theta) &:= \int_{I(s_{1:j})} \mathscr{E}_{s_{1:j}}^{\beta^{[j]}}(\Delta,x,y;\theta) ds_j \cdots ds_1. \end{aligned} \tag{27}$$

Step 3. From (22) in *Step 1.* and (26) in *Step 2.*, we obtain the following CF-expansion for the true (intractable) transition density. For any $M \ge 1$ and multi-indices $\beta^{[j]} \in \mathbb{Z}_{\ge 0}^j$, $1 \le j \le M-1$:

$$p_\Delta^X(x,y;\theta) = p_\Delta^{\bar{X}}(x,y;\theta) + \sum_{1\le j\le M-1} \sum_{\alpha\le\beta^{[j]}} \Delta^{|\alpha|+j} \cdot \frac{K(\alpha)}{\alpha!} \cdot \mathscr{D}_\alpha^{z,\theta}\{p_\Delta^{\bar{X}^z}(\cdot,y;\theta)\}(x)\Big|_{z=x}$$

$$+\mathscr{R}_1^M(\Delta,x,y;\theta)+\sum_{1\le j\le M-1}\mathscr{R}_2^{j,\beta^{[j]}}(\Delta,x,y,\theta), \qquad (28)$$

where $\mathscr{R}_1^M$, $\mathscr{R}_2^{j,\beta^{[j]}}$ are defined in (23), (27), respectively. Under assumptions, we show in Section 3.1 that the remainder terms are of size $O(\Delta^p)$ for an arbitrary $p>0$ by choosing a large enough M and appropriate $\beta^{[j]}$, $1\le j\le M-1$. The double sum in (28) involves tractable terms and can be utilised as a proxy for the true transition density. In particular, the expansion is well-defined for both model classes (E) and (H) since the Gaussian density $p^{\bar{X}^z}(x,y;\theta)$ and its partial derivatives (involved in $\mathscr{D}_\alpha^{z,\theta}\{p_\Delta^{\bar{X}^z}(\cdot,y;\theta)(x)\}|_{z=x}$) are well-defined from Lemma 2.5. We note that the tractable double sum in (28) is regarded as a CF-expansion, but the current form of the expansion does not yet correspond to the 'Δ'-expansion (3). For instance, the exponents of the step-size Δ are integers in (28), while they are given as $k/2$, $k\in\mathbb{N}$, in (3). However, we emphasise that (28) will be ultimately expressed as a Δ-expansion of the form in (3) after carefully working with the terms $\mathscr{D}_\alpha^{z,\theta}\{p_\Delta^{\bar{X}^z}(\cdot,y;\theta)\}(x)$. Indeed, taking partial derivatives of $x\mapsto p^{\bar{X}^z}(x,y;\theta)$, will give Hermite polynomials and powers $\Delta^{-k/2}$, where the integer k depends on the number of derivatives. We explain this in detail later on in Section 3.2.

3. Error Analysis for the CF-Expansion

In Section 2.4 we have constructed a CF-expansion (28) for the true transition density. Our objective now is to provide rigorous error estimates for this expansion, thus theoretically justifying its derivation, similarly to results obtained by a few earlier works in the case of the elliptic class (E). We also describe that the obtained expansion (33) can be given in the form (3), namely a series in powers of Δ. As the error estimates vary for classes (E), (H), we make use of the notation $w\in\{\text{E},\text{H}\}$ and write $p^{X,(w)}\equiv p^X$, $p^{\bar{X},(w)}\equiv p^{\bar{X}}$, $\mathscr{R}_1^{M,(w)}\equiv\mathscr{R}_1^M$ and $\mathscr{R}_2^{j,\beta^{[j]},(w)}\equiv\mathscr{R}_2^{j,\beta^{[j]}}$ to indicate the class under consideration.

3.1. Main Result

We derive upper bounds for the residuals of the CF expansion $\mathscr{R}_1^M$ and $\mathscr{R}_2^{j,\beta^{[j]}}$ specified in (23) and (27), respectively. We will need the following additional assumptions.

Assumption 3.1. The parameter space Θ is compact. Also, for each $x\in\mathbb{R}^N$, the function $\theta\mapsto V_j(x,\theta)$ is continuous, $0\le j\le d$.

Assumption 3.2. Let $x\in\mathbb{R}^N$ be the initial state of the transition dynamics. The SDE coefficients satisfy the following properties:

1 (Boundedness of drift at initial state): There exists a constant $\kappa>0$ such that $|x|+|V_0(x,\theta)|\le\kappa$ for all $\theta\in\Theta$;
2 (Uniform boundedness of diffusion coefficients): There exists a constant $C>0$ such that $|V_j(y,\theta)|\le C$, $1\le j\le d$, for all $(y,\theta)\in\mathbb{R}^N\times\Theta$;
3 (Uniform boundedness of derivatives): There is a constant $C>0$ such that $\sum_{j=0}^d|\partial_y^\alpha V_j(y,\theta)|\le C$ for all $\alpha\in\mathbb{Z}_{\ge0}^N$ with $|\alpha|\ge1$ and all $(y,\theta)\in\mathbb{R}^N\times\Theta$.

Assumption 3.3. For the hypo-elliptic model class (H), $N_S=N_R=d$.

Assumptions 3.1–3.3 suffice for obtaining appropriate bounds for the residuals of the CF-expansion. The uniform boundedness for the derivatives of the SDE coefficients is a standard assumption for the existence of a smooth transition density, when combined with Hörmander's condition. Such uniform boundedness is also assumed in Li (2013) to control the residual of the expansion developed therein for elliptic SDEs. Assumptions 3.2–3.3 are used mainly in the proof of Theorem 3.4, where we need an upper bound for the true density $y \mapsto p_{\Delta}^{X}(x, y; \theta)$. Pigato (2022) shows that under Assumptions 3.2–3.3 the true transition density has a Gaussian-type bound as given later at (31). Based on this result, we show that the errors are appropriately bounded, analogously to Yang, Chen and Wan (2019) who also used a Gaussian-type bound for the true density to control the residuals of an expansion for inhomogeneous elliptic SDEs. We stress that Assumptions 3.1–3.3 are not necessary for the construction of the CF-expansion, in the sense that our formulae can still be evaluated for SDEs with coefficients whose partial derivatives exhibit, e.g., polynomial growth as assumed in earlier works (Aït-Sahalia, 2008, Yang, Chen and Wan, 2019) for elliptic SDEs. Assumption 3.3 can be weakened as there is a possibility to obtain an upper bound for the true transition density by carefully developing the arguments in Pigato (2022). However, this is not straightforward and is beyond the scope of the present work. The relaxation of our conditions is left for future research.

To provide a statement of our main result, we introduce some notation. We set:

$$m(\mathrm{E}) := N, \qquad m(\mathrm{H}) := 4d. \tag{29}$$

Also, $\mathscr{G}^{(w)} : (0,\infty) \times \mathbb{R}^N \times \mathbb{R}^N \times \Theta \to \mathbb{R}$, $w \in \{\mathrm{E}, \mathrm{H}\}$, is a mapping characterised as follows. There exist constants $C, c > 0$ such that:

$$\left|\mathscr{G}^{(\mathrm{E})}(t, x, y, \theta)\right| \leq C t^{-\frac{m(\mathrm{E})}{2}} \times \exp\Big(-c \tfrac{|y-x|^2}{t}\Big); \tag{30}$$

$$\left|\mathscr{G}^{(\mathrm{H})}(t, x, y, \theta)\right| \leq C t^{-\frac{m(\mathrm{H})}{2}} \times \exp\left(-c\left(\frac{|y_S - x_S - V_{S,0}(x,\theta)t|^2}{t^3} + \frac{|y_R - x_R|^2}{t}\right)\right), \tag{31}$$

for all $(t, x, y, \theta) \in (0,\infty) \times \mathbb{R}^N \times \mathbb{R}^N \times \Theta$. Notice that for some constant $C > 0$, for $\Delta \in (0,1)$:

$$\begin{aligned} &\sup_{(x,y,\theta)\in\mathbb{R}^N\times\mathbb{R}^N\times\Theta} \left|\mathscr{G}^{(w)}(\Delta, x, y, \theta)\right| \leq C \Delta^{-\frac{m(w)}{2}}, \\ &\sup_{(x,\theta)\in\mathbb{R}^N\times\Theta} \int_{\mathbb{R}^N} \left|\mathscr{G}^{(w)}(\Delta, x, y, \theta)\right| dy \leq C, \end{aligned} \qquad w \in \{\mathrm{E}, \mathrm{H}\}, \tag{32}$$

which implies that the size of $\mathscr{G}^{(w)}$ in L_1-norm is $O(1)$ irrespective of the model classes (E) or (H).

Theorem 3.4 (Bound for $\mathscr{R}_1^{M,(w)}$). *Let $x \in \mathbb{R}^N$ be the initial state of the transition dynamics and $M \geq 1$. Under Assumptions 2.1–3.3, there exists a constant $C > 0$ such that for all $(\Delta, y, \theta) \in (0,1) \times \mathbb{R}^N \times \Theta$:*

$$\left|\mathscr{R}_1^{M,(w)}(\Delta, x, y; \theta)\right| \leq C \Delta^{\frac{M}{2}} \times \left|\mathscr{G}^{(w)}(\Delta, x, y, \theta)\right|, \qquad w \in \{E, H\}.$$

Theorem 3.5 (Bound for $\mathscr{R}_2^{j,\beta^{[j]},(w)}$). *Let $x \in \mathbb{R}^N$ be the initial state of the transition dynamics, and let $1 \leq j \leq M-1$, $M \in \mathbb{N}$, $\beta^{[j]} = (\beta_1^{[j]}, \ldots, \beta_j^{[j]}) \in \mathbb{Z}_{\geq 0}^{j}$. Under Assumptions 2.1–3.3, there exists constant $C > 0$ such that for all $(\Delta, y, \theta) \in (0,1) \times \mathbb{R}^N \times \Theta$:*

$$\left|\mathscr{R}_2^{j,\beta^{[j]},(w)}(\Delta, x, y; \theta)\right| \leq C \Delta^{K^{[j],w}(\beta^{[j]})} \times \left|\mathscr{G}^{(w)}(\Delta, x, y, \theta)\right|, \quad w \in \{E, H\},$$

$$K^{[j],E}(\beta^{[j]}) := \min_{1\le i\le j} \frac{\beta_i^{[j]}}{2} + \frac{j}{2}, \qquad K^{[j],H}(\beta^{[j]}) := \min_{1\le i\le j} \frac{1}{2}\big(\lfloor \frac{\beta_i^{[j]}}{2} \rfloor - \mathbf{1}_{\beta_i^{[j]}\ge 2}\big) + \frac{j}{2}.$$

The proofs of Theorems 3.4–3.5 are given in Section A of Supplementary Material. From Theorem 3.5, by selecting the multi-index $\beta^{[j],(w)} \in \mathbb{Z}^j_{\ge 0}$ so that $K^{[j],w}(\beta^{[j],(w)}) \ge \frac{M}{2}$, we have that:

$$\begin{aligned} p_\Delta^{X,(w)}(x,y;\theta) &= p_\Delta^{\bar{X},(w)}(x,y;\theta) \\ &+ \sum_{1\le j\le M-1} \sum_{\alpha\le\beta^{[j],(w)}} \Delta^{|\alpha|+j} \cdot \frac{K(\alpha)}{\alpha!} \cdot \mathscr{D}_\alpha^{z,\theta}\{p_\Delta^{\bar{X}^z,(w)}(\cdot,y;\theta)\}(x)\Big|_{z=x} + \mathscr{E}^{(w)}(\Delta,x,y;\theta), \end{aligned} \tag{33}$$

for a residual $\mathscr{E}^{(w)}$ so that for any initial state $x \in \mathbb{R}^N$ there exists a constant $C > 0$ such that for all $(\Delta, y, \theta) \in (0,1) \times \mathbb{R}^N \times \Theta$:

$$\big|\mathscr{E}^{(w)}(\Delta,x,y;\theta)\big| \le C\Delta^{\frac{M}{2}} \times \big|\mathscr{G}^{(w)}(\Delta,x,y,\theta)\big|, \qquad w \in \{\mathrm{E},\mathrm{H}\}.$$

3.2. Series Expansion in Δ

We study the CF-expansion given in (33) in detail. Expression (33) involves $\Delta^{|\alpha|+j}$ in front of each summand, but an additional $\Delta^{-K_\alpha/2}$, for some $K_\alpha \in \mathbb{N}$, is produced from $\mathscr{D}_\alpha^{z,\theta}\{p_\Delta^{\bar{X}^z}(\cdot,y;\theta)\}(x)$, where we recall that $\mathscr{D}_\alpha^{z,\theta}$ is the differential operator defined in (25). We show that upon rearrangement of terms in powers of Δ, the right-hand-side of (33) attains the form of the Δ-expansion in (3), i.e. a series expansion in (positive) powers of $\sqrt{\Delta}$. In particular, we clarify below that differentiating the Gaussian density $p^{\bar{X}^z}(x,y;\theta)$ w.r.t. the initial state $x \in \mathbb{R}^N$ produces additional powers $\Delta^{-K/2}$, for $K \ge 1$ depending on the number of derivatives. For the model class (H), the value of K varies depending on whether the differentiation acts on smooth or rough components. We define, for $\alpha \in \mathbb{Z}^N_{\ge 0}$:

$$\|\alpha\|_{\mathrm{E}} := \tfrac{1}{2}|\alpha|, \qquad \|\alpha\|_{\mathrm{H}} = \tfrac{3}{2}|\alpha_S| + \tfrac{1}{2}|\alpha_R|, \tag{34}$$

where, for class (H), we interpret $\alpha = (\alpha_{S,1},\dots,\alpha_{S,N_S},\alpha_{R,1},\dots,\alpha_{R,N_R})$ for given $\alpha_S \in \mathbb{Z}^{N_S}_{\ge 0}$, $\alpha_R \in \mathbb{Z}^{N_R}_{\ge 0}$. We then have the following key result whose proof is provided in Section A.2 of Supplementary Material:

Lemma 3.6. *Let $x, y \in \mathbb{R}^N$, $\theta \in \Theta$, $\Delta > 0$. Also, let $\alpha \in \mathbb{Z}^N_{\ge 0}$. Under Assumptions 2.1–3.3, we have that:*

$$\partial_x^\alpha\, p_\Delta^{\bar{X}^z,(w)}(x,y;\theta)|_{z=x} = \Delta^{-\|\alpha\|_w} \times \mathscr{H}_\alpha^{(w)}(\Delta,x,y;\theta)\, p_\Delta^{\bar{X},(w)}(x,y;\theta), \qquad w \in \{E,H\},$$

where $y \mapsto \mathscr{H}_\alpha^{(w)}(\Delta,x,y;\theta)$ is obtained explicitly defined via differentiation of $x \mapsto p_\Delta^{\bar{X}^z,(w)}(x,y;\theta)$ and is characterised as follows. There exists a constant $C > 0$ such that for all $\Delta > 0$, $x, y \in \mathbb{R}^N$, $\theta \in \Theta$:

$$\big|\mathscr{H}_\alpha^{(w)}(\Delta,x,y;\theta)\, p_\Delta^{\bar{X},(w)}(x,y;\theta)\big| \le C\big|\mathscr{G}^{(w)}(\Delta,x,y,\theta)\big|.$$

In brief, Lemma 3.6 states the following. For model class (E), taking $k \in \mathbb{N}$ partial derivatives of $x \mapsto p_\Delta^{\bar{X}^z,(\mathrm{E})}(x,y;\theta)$ yields a term $\Delta^{-\frac{k}{2}}$. For class (H), taking $k \in \mathbb{N}$ partial derivatives of $p_\Delta^{\bar{X}^z,(\mathrm{H})}(x,y;\theta)$

w.r.t. the smooth components (resp. the rough components) produces the term $\Delta^{-\frac{3}{2}k}$ (resp. the term $\Delta^{-\frac{k}{2}}$). Based upon Lemma 3.6, the form of the CF-expansion is determined from the expression of the differential operator $\mathscr{D}_{\alpha}^{z,\theta}$ and the number of derivatives involved therein. A detailed characterisation for the differential operator $\mathscr{D}_{\alpha}^{z,\theta}$ is provided in Supplementary Material. In particular, Lemma B.3 in Supplementary Material states that the operator admits the following expression. For $\varphi \in C^{\infty}(\mathbb{R}^N;\mathbb{R})$, $\alpha \in \mathbb{Z}_{\geq 0}^{j}$, $j \in \mathbb{N}$ and $(x,\theta) \in \mathbb{R}^N \times \Theta$:

$$\mathscr{D}_{\alpha}^{z,\theta,(w)} \varphi(x)|_{z=x} = \sum_{\gamma \in \mathscr{J}_w(\alpha)} \mathscr{W}_{\gamma}^{[\alpha]}(x,\theta) \partial^{\gamma} \varphi(x), \qquad w \in \{\mathrm{E}, \mathrm{H}\}, \tag{35}$$

where $\mathscr{J}_w(\alpha)$ is a set of multi-indices $\mathbb{Z}_{\geq 0}^{N}$ defined in (B.21) in Supplementary Material and $\mathscr{W}_{\gamma}^{[\alpha]} : \mathbb{R}^N \times \Theta \to \mathbb{R}$ is explicitly determined from products of partial derivatives of the SDE coefficients and can be evaluated in applications using software performing symbolic calculations. Due to (35) and Lemma 3.6, we have that for $w \in \{\mathrm{E}, \mathrm{H}\}$:

$$\begin{aligned}
&\sum_{1 \leq j \leq M-1} \sum_{\alpha \leq \beta^{[j]}} \Delta^{|\alpha|+j} \cdot \frac{K(\alpha)}{\alpha!} \cdot \mathscr{D}_{\alpha}^{z,\theta,(w)} \{p_{\Delta}^{\bar{X}^z,(w)}(\cdot, y;\theta)\}(x)\Big|_{z=x} \\
&= \sum_{1 \leq j \leq M-1} \sum_{\alpha \leq \beta^{[j]}} \sum_{\gamma \in \mathscr{J}_w(\alpha)} \Delta^{|\alpha|+j-\|\gamma\|_w} \frac{K(\alpha)}{\alpha!} \mathscr{W}_{\gamma}^{[\alpha]}(x,\theta)\, \mathscr{H}_{\gamma}^{(w)}(\Delta,x,y;\theta)\, p_{\Delta}^{\bar{X},(w)}(x,y;\theta) \\
&\equiv \sum_{1 \leq k \leq M-1} \Delta^{\frac{k}{2}} \cdot e_k^{(w)}(\Delta,x,y;\theta) \cdot p_{\Delta}^{\bar{X},(w)}(x,y;\theta) + \mathscr{R}_3^{(w)}(\Delta,x,y;\theta)\, p_{\Delta}^{\bar{X},(w)}(x,y;\theta),
\end{aligned} \tag{36}$$

where in the last line, we rearrange the sum in ascending order in powers of $\sqrt{\Delta}$ and have defined:

$$\begin{aligned}
&e_k^{(w)}(\Delta,x,y;\theta) \\
&\quad := \sum_{1 \leq j \leq M-1} \sum_{\alpha \leq \beta^{[j]}} \sum_{\gamma \in \mathscr{J}_w(\alpha)} \frac{K(\alpha)}{\alpha!} \mathscr{W}_{\gamma}^{[\alpha]}(x,\theta) \mathscr{H}_{\gamma}^{(w)}(\Delta,x,y;\theta) \cdot \mathbf{1}_{|\alpha|+j-\|\gamma\|_w = \frac{k}{2}}; \qquad (37) \\
&\mathscr{R}_3^{(w)}(\Delta,x,y;\theta) \\
&\quad := \sum_{1 \leq j \leq M-1} \sum_{\alpha \leq \beta^{[j]}} \sum_{\gamma \in \mathscr{J}_w(\alpha)} \Delta^{|\alpha|+j-\|\gamma\|_w} \frac{K(\alpha)}{\alpha!} \mathscr{W}_{\gamma}^{[\alpha]}(x,\theta)\, \mathscr{H}_{\gamma}^{(w)}(\Delta,x,y;\theta) \cdot \mathbf{1}_{|\alpha|+j-\|\gamma\|_w \geq \frac{M}{2}}.
\end{aligned}$$

Under Assumptions 2.1–3.3, from Lemma 3.6, there exists a constant $C > 0$ such that:

$$\left|\mathscr{R}_3^{(w)}(\Delta,x,y;\theta) p_{\Delta}^{\bar{X},(w)}(x,y;\theta)\right| \leq C \Delta^{\frac{M}{2}} \left|\mathscr{G}^{(w)}(\Delta,x,y,\theta)\right|, \tag{38}$$

for all $(\Delta, y, \theta) \in (0,1) \times \mathbb{R}^N \times \Theta$. Working with Theorems 3.4–3.5, (33), (36) and (38), we finally obtain the following 'minimal' representation of density expansion for diffusion models (E) and (H).

Theorem 3.7 (Δ-Expansion). *Let $(\Delta, x, y, \theta) \in (0,1) \times \mathbb{R}^N \times \mathbb{R}^N \times \Theta$. Under Assumptions 2.1–3.3, the transition density admits the following expansion. For every $J \in \mathbb{N}$ and $w \in \{E, H\}$:*

$$p_{\Delta}^{X,(w)}(x,y;\theta) = p_{\Delta}^{\bar{X},(w)}(x,y;\theta) \cdot \left\{1 + \sum_{1 \leq j \leq J} \Delta^{\frac{j}{2}} \cdot e_j^{(w)}(\Delta,x,y;\theta)\right\} + \mathscr{R}^{J,(w)}(\Delta,x,y;\theta). \tag{39}$$

The coefficients $e_j^{(w)}$ *are explicitly determined in (37). Also, for* $1 \le j \le J$*, there exists a constant* $C > 0$ *such that for all* $(\Delta, y, \theta) \in (0,1) \times \mathbb{R}^N \times \Theta$*,*

$$\left|e_j^{(w)}(\Delta, x, y; \theta) \times p_\Delta^{\bar{X},(w)}(x, y; \theta)\right| \le C \left|\mathscr{G}^{(w)}(\Delta, x, y, \theta)\right|.$$

For the residual $\mathscr{R}^{J,(w)}$*, there exist constants* $C_1, C_2, C_3 > 0$ *such that for all* $(\Delta, y, \theta) \in (0,1) \times \mathbb{R}^N \times \Theta$*:*

$$\left|\mathscr{R}^{J,(w)}(\Delta, x, y; \theta)\right| \le C_1 \Delta^{\frac{J+1}{2}} \left|\mathscr{G}^{(w)}(\Delta, x, y, \theta)\right| \le C_2 \Delta^{\frac{J+1}{2} - \frac{m(w)}{2}} \tag{40}$$

and

$$\int_{\mathbb{R}^N} \left|\mathscr{R}^{J,(w)}(\Delta, x, y; \theta)\right| dy \le C_3 \Delta^{\frac{J+1}{2}}. \tag{41}$$

We note that the pointwise error bound (40) differs across model classes (E) and (H) due to $m(w)$ taking a larger value in the latter case. In brief, this is due to X_S in (H) being a smooth component, driven by a Gaussian noise $\int_0^\Delta B_s ds$ of size $O(\Delta^{3/2})$ rather than by B_Δ of size $O(\Delta^{1/2})$ in the case of X_R, thus X_S has a smaller variance for a fixed $\Delta \in (0,1)$. I.e., the existence of the smooth component X_S in (H) leads to a sharper density and/or concentration around the mode in the X_S coordinate. However, in terms of L_1-error, its order only depends on the choice of J and not on the model class because $y \mapsto \mathscr{G}^{(w)}(x, y, \theta)$ can be treated as a Gaussian density for any $(x, \theta, w) \in \mathbb{R}^d \times \Theta \times \{\mathrm{E}, \mathrm{H}\}$; recall also (32).

Remark 3.8. Let $\pi^{[J],(w)}(\Delta, x, y; \theta) := \sum_{1 \le j \le J} \Delta^{\frac{j}{2}} \cdot e_j^{(w)}(\Delta, x, y; \theta)$. To avoid negative values for $\pi^{[J],(w)}$, we use a standard technique (see e.g. Stramer, Bognar and Schneider (2010) for a related approach) where $1 + \xi = \exp\left\{\log(1+\xi)\right\} = \exp\{T_{J'}(\xi)\} \exp\{R_{J'}(\xi)\}$, for $J' \ge 1$, with $T_{J'}(\xi) := \sum_{j=1}^{J'} (-1)^{j+1} \frac{\xi^j}{j}$ the J'-order Taylor expansion of $\xi \mapsto \log(1+\xi)$ and $R_{J'}(\xi)$ its residual. Via simple arguments, for $|\xi| < \delta < 1$ one has $\left|(1+\xi) - \exp\{T_{J'}(\xi)\}\right| \le C\delta^{J'+1}$, for $C > 0$. The above suggests the use of the following proxy:

$$\widetilde{p}_\Delta^{(w)}(x, y; \theta) := p_\Delta^{\bar{X},(w)}(x, y; \theta) \cdot \exp\left\{T_{J'}\left(\pi^{[J],(w)}(\Delta, x, y; \theta)\right)\right\}. \tag{42}$$

Thus, $\pi^{[J],(w)}$ includes powers $\Delta^{1/2}, \dots, \Delta^{J/2}$ (assuming non-zero e_j's), so is of size $\delta = O(\Delta^{1/2})$ and the residual in (39) is $O(\Delta^{(J+1)/2})$ – in the sense of the first bound in (40). For the replacement by the Taylor approximation to only affect terms of size $O(\Delta^{(J+1)/2})$, one should select J' as the smallest even integer so that $J' \ge J$. An even J' guarantees integrability of the density proxy.

4. Numerical Experiments

We focus on the bivariate *FitzHugh-Nagumo (FHN)* SDE used in neuroscience. This model writes as:

$$dV_t = \tfrac{1}{\varepsilon}\left(V_t - V_t^3 - U_t - s\right)dt; \qquad dU_t \quad = \left(\gamma V_t - U_t + \beta\right)dt + \sigma\, dB_{1,t}, \tag{43}$$

with V describing the membrane potential of a single neuron and the recovery variable U expressing the ion channel kinetics. Also, s is the magnitude of the stimulus current and is often controlled and $\theta =$

$(\epsilon, \gamma, \beta, \sigma)$ is the parameter. This SDE does not satisfy the boundedness conditions in Assumption 3.2 as there is a non-Lipschitz term in the drift. Statistical inference for the FHN SDE is an important topic from a theoretical and a practical viewpoint, see Ditlevsen and Samson (2019), Melnykova (2020), Samson, Tamborrino and Tubikanec (2025). SDE (43) belongs in class (H) as the weak Hörmander's condition (specifically, Assumption 2.2-II) holds in this case. The transition density is intractable and we approximate it with the CF-expansion given in Section 3.2. We investigate the accuracy of the CF-expansion in Section 4.1 and use the expansion to carry out Bayesian inference with real data in Section 4.2.

4.1. Accuracy of the CF-Expansion

We produce two expansions via use of different baseline Gaussian densities. In particular, for a given initial value $x = (x_1, x_2) \in \mathbb{R}^2$, $\Delta > 0$ and $J \in \mathbb{N}$, we work with the CF-expansions:

$$y \mapsto p_{\Delta}^{(\iota),[J]}(x, y; \theta) := \bar{p}_{\Delta}^{(\iota)}(x, y; \theta) \times \Big\{ 1 + \sum_{1 \le k \le J} \Delta^{\frac{k}{2}} \cdot e_k^{(\iota)}(\Delta, x, y) \Big\}, \qquad \iota \in \{\mathrm{I}, \mathrm{II}\}, \tag{44}$$

with the reference density $\bar{p}_{\Delta}^{(\iota)}(\cdot)$, $\iota \in \{\mathrm{I}, \mathrm{II}\}$, corresponding to the following 'full' (for $\iota = \mathrm{I}$) or 'partial' (for $\iota = \mathrm{II}$) LDL scheme:

$$d\bar{X}_t^{(\iota)} = \big(A_{x,\theta}^{(\iota)} \bar{X}_t^{(\iota)} + b_{x,\theta}^{(\iota)} \big) dt + \begin{bmatrix} 0 \\ \sigma \end{bmatrix} dB_{1,t}, \tag{45}$$

where we consider the following two choices:

$$A_{x,\theta}^{(\mathrm{I})} := \begin{bmatrix} \big(1 - 3(x_1)^2\big)/\varepsilon & -1/\varepsilon \\ \gamma & -1 \end{bmatrix}, \qquad A_{x,\theta}^{(\mathrm{II})} := \begin{bmatrix} \big(1 - 3(x_1)^2\big)/\varepsilon & -1/\varepsilon \\ 0 & -1 \end{bmatrix};$$

$$b_{x,\theta}^{(\iota)} = \begin{bmatrix} \big(x_1 - (x_1)^3 - x_2 + s\big)/\varepsilon \\ \gamma x_1 - x_2 + \beta \end{bmatrix} - A_{x,\theta}^{(\iota)} x.$$

The $e_k^{(\iota)}$'s are found starting from expansion (28), with $\mathscr{L}_{\theta}^{0,z}$ (used by the differential operator $\mathscr{D}_{\alpha}^{z,\theta}$) corresponding to the generator associated with (45) for $\iota \in \{\mathrm{I}, \mathrm{II}\}$, and then re-arranging terms in powers of $\sqrt{\Delta}$ as described in Section 3.2. Matrix $A^{(\mathrm{II})}$ is upper-triangular, so the baseline $\bar{p}_{\Delta}^{(\mathrm{II})}$ takes a simpler form compared to when using $A^{(\mathrm{I})}$. In both cases, the reference Gaussian laws are non-degenerate. To calculate the $e_k^{(\iota)}$'s we use `Mathematica` with full expressions given in Section C in Supplementary Material. Due to the SDE noise being additive, we have $w_k^{(\iota)} = 0$, for $k = 1, 2$, $\iota \in \{\mathrm{I}, \mathrm{II}\}$. Thus, the CF-expansions with $J \in \{1, 2\}$ coincide with the baseline. The reference density for $\iota = \mathrm{I}$ involves full linearisation, so the $e_k^{(\mathrm{I})}$'s have simpler expressions than the $e_k^{(\mathrm{II})}$'s, see Section C.1 in Supplementary Material for details.

We choose $s = 0.01$, initial value $x = (V_0, U_0) = (-0.1, 0.2)$ and $\theta = (\varepsilon, \gamma, \beta, \sigma) = (0.1, 1.2, 0.3, 0.8)$. We consider $\Delta \in \{0.1, 0.05, 0.02\}$ and compute CF-expansions using the transform $\widetilde{p}_{\Delta}$ described in Remark 3.8, which we denote here $\widetilde{p}_{\Delta}^{(\iota),[J]}$, $\iota \in \{\mathrm{I}, \mathrm{II}\}$. We try $J = 2, 3, 4, 5$, and for $\widetilde{p}_{\Delta}$ we set $J' = 2$, as the correction term includes powers $\Delta^{3/2}, \ldots, \Delta^{J/2}$, $J \le 5$, and the transform can only affect terms of size $O((\Delta^{3/2})^{(J'+1)}) = O(\Delta^{9/2})$. We find the benchmark 'true' density via a simulation that: (i) uses 2×10^7 samples from the FHN SDE at Δ via an EM scheme with discretisation step $\Delta/800$; (ii) applies a

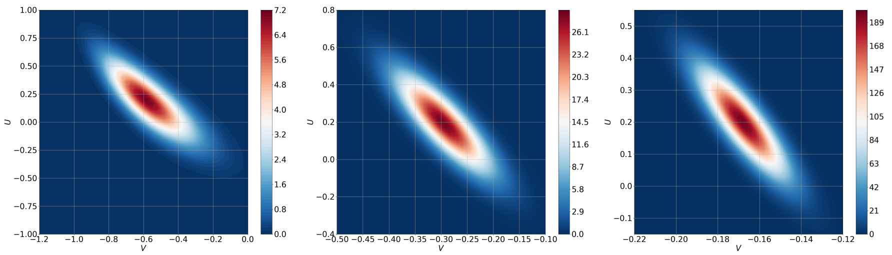


Figure 1. Contours of the benchmark densities $p_\Delta^{(\mathrm{B})}$. Left: $\Delta = 0.1$. Middle: $\Delta = 0.05$. Right: $\Delta = 0.02$.

standard Kernel Density Estimator (KDE) approach to reconstruct the density. We write the benchmark density as $p_\Delta^{(\mathrm{B})}$. Densities are evaluated on a regular 51×51 grid $D = \{(s_i, r_j) \,|\, 0 \le i, j \le 50\} \subset \mathbb{R}^2$ for reals $r_0 < \cdots < r_{50}$ and $s_0 < \cdots < s_{50}$ defined in an apparent way.

Fig. 1 shows the contours of the benchmark $p_\Delta^{(\mathrm{B})}$, which indicate the unimodality of the target transition densities. Fig. 2 plots the absolute errors,

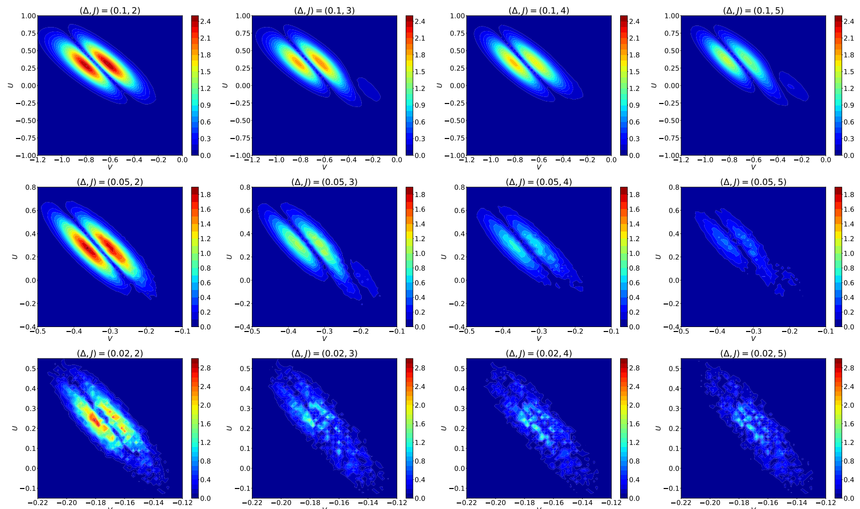
$$\mathscr{E}_\Delta^{(\iota),[J]}(x, y; \theta) := \left| p_\Delta^{(B)}(x, y; \theta) - \widetilde{p}_\Delta^{(\iota),[J]}(x, y; \theta) \right|, \quad y \in D,$$

between $p_\Delta^{(\mathrm{B})}$ and the CF-expansions of order $J = 2, 3, 4, 5$. Fig. 3 summarises the overall performance of the CF-expansions. Fig. 3(a) gives the L_1-error of the CF-expansions, defined as $\mathrm{L}_1^{(\iota),[J]}(\Delta, x; \theta) := \sum_{y \in D} \mathscr{E}_\Delta^{(\iota),[J]}(x, y; \theta) \times \delta_V \times \delta_U$, where $\delta_V = (s_{50} - s_0)/50$, $\delta_V = (r_{50} - r_0)/50$. Fig. 3(b) shows the average running time of DE-I and DE-II (denoting the two density expansions for $\iota \in \{\mathrm{I}, \mathrm{II}\}$), with the average taken from the 3 choices of Δ. Fig. 2 demonstrates that absolute errors diminish as J increases. In particular, the error of single mode and variance (or higher order moments) between the benchmark and approximate transition densities gradually diminishes as J increases. We observe a similar decrease in L_1-error in Fig. 3(a). Note that the errors by the two CF-expansions with $J = 5$ are less than half of those with $J = 2$. Also, errors decrease for smaller Δ. In terms of computing cost, for the DE with $J = 2$, i.e. the baseline Gaussian density without correction, DE-II is computationally cheaper due to the simpler expression in the matrix exponential. Costs are similar between DE-I with $J = 2$ and DE-II with $J = 4$. Costs grow as J increases, but the growth rate seems faster in DE-II since the latter makes use of the simpler but slightly less accurate baseline density, thus involves more correction terms as J grows.

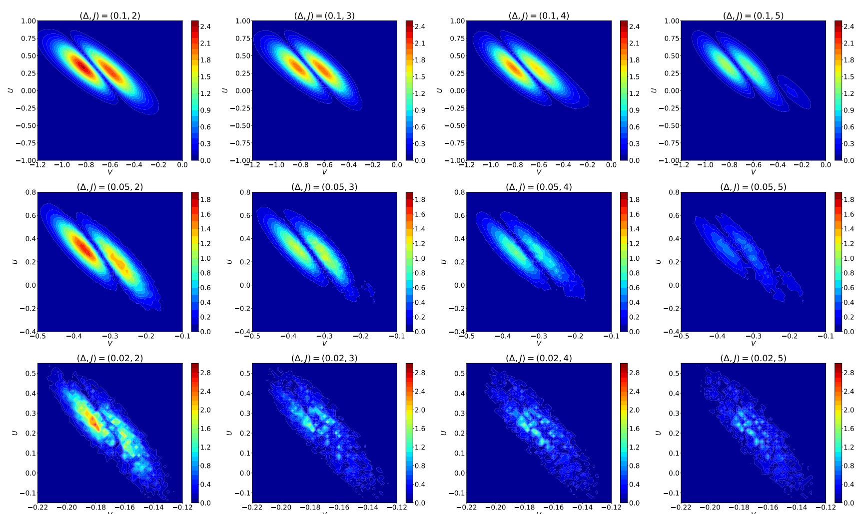
4.2. Application to Bayesian Inference

4.2.1. MCMC via CF-expansion – Design of Posterior

We use our CF-expansion to carry out Bayesian inference for SDEs. In this subsection we consider general SDEs rather than just the FHN SDE as the approach is relevant in a wide setting. Via the CF-expansion we obtain a posterior law that can be integrated within well-established MCMC methodologies, including centred/non-centred parameterisations (Papaspiliopoulos, Roberts and Sköld, 2007), Particle MCMC and Particle Gibbs algorithms (Andrieu, Doucet and Holenstein, 2010). Note that early literature (Stramer, Bognar and Schneider, 2010) investigated the use of CF-expansions (for elliptic models) within a standard Metropolis-Hastings method under centred parametrisation, thus the options



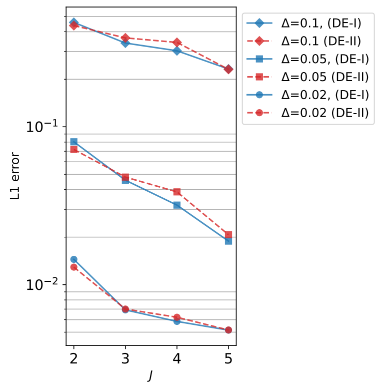
(a) DE-I.



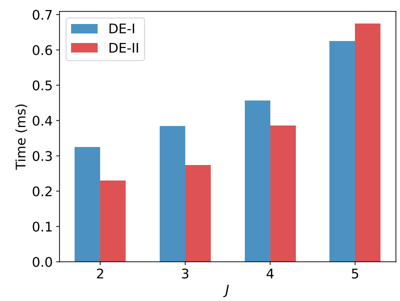
(b) DE-II.

Figure 2. Heatmap of the absolute error for the CF-expansion in the case of the hypo-elliptic FHN model (see Section 4). Rows correspond to 3 choices $\Delta = (0.1, 0.05, 0.02)$ and columns to 4 choices $J = 2, 3, 4, 5$.

provided were limited. Particle-based MCMC methods require sampling from the SDE transition density, i.e. in our case from the CF-expansion used as its proxy. It is typically difficult to simulate from the CF-expansion. However, notice that the expansion writes as ‘Gaussian density’ $\times$ ‘correction term’. Thus, particle-based MCMC and general Sequential Monte Carlo (SMC) methodology can be implemented using the baseline density (which we can sample from) with the correction term being attached in the ‘weights’ within the algorithm. Furthermore, the CF-expansion structure of ‘Gaussian density’ $\times$ ‘correction term’ permits a non-centred approach – such an algorithm turns out to be the most effective



(a) L_1-error



(b) Computational time per grid

Figure 3. Summary for performance of density expansions.

one in our numerics in the next section. We provide more details on the mentioned algorithms directly below.

Consider the data $\mathscr{Y}_n = \{Y_{t_k}\}_{0\le k\le n}$ at instances t_k, $0 \le k \le n$, for which we assume an equidistant step-size Δ. We consider the setting of noisy observations, so that there is a density $p(Y_{t_k}|X_{t_k})$, assumed known. Under a *data augmentation* approach, we set $\mathbf{q} := (\theta, \mathscr{X}_n) \in \mathbb{R}^{d_\theta} \times \mathbb{R}^{N\times(n+1)}$, where $\mathscr{X}_n := \{X_{t_k}\}_{1\le k\le n}$. The posterior density on the augmented state $\mathbf{q} = (\theta, \mathscr{X}_n)$ writes as:

$$P(\mathbf{q}\,|\,\mathscr{Y}_n) \propto \Big\{\prod_{0\le k\le n} p(Y_{t_k}|X_{t_k})\Big\} \times \Big\{\prod_{1\le k\le n} p^X_\Delta(X_{t_{k-1}}, X_{t_k};\theta)\Big\} \times p_0(X_0) \times p_\theta(\theta), \tag{46}$$

where p_0, p_θ denote priors on the initial value X_0 and the parameter θ, respectively. We replace the true transition density with the CF-expansion as given in (42), that is:

$$\prod_{1\le k\le n} p^X_\Delta(X_{t_{k-1}}, X_{t_k};\theta) \approx \prod_{1\le k\le n} p^{\bar{X}}_\Delta(X_{t_{k-1}}, X_{t_k};\theta) \times \prod_{1\le k\le n} \exp\Big(T_{J'}\big(\pi^{[J]}(\Delta, X_{t_{k-1}}, X_{t_k};\theta)\big)\Big). \tag{47}$$

The approximate posterior obtained via (46)-(47) can now be used within standard or particle-based MCMC methods: (i) For standard MCMC, the 'correction terms' can be treated as a part of the likelihood function, so that a-priori the dynamics of the X-process are determined by the baseline density. This allows for application of centred/non-centred algorithms, as in the latter case one can use as latent components the standard Gaussian noise that generates samples from the baseline density; (ii) For particle-based methods, the 'correction terms' can become part of the weights and one can apply, e.g., particle filters by sampling from the tractable baseline density.

Remark 4.1. In the above, we have discussed the use of baseline Gaussian density as a 'proposal' within the standard/particle-based MCMC computational framework with 'correction terms' becoming part of the weights, rather than directly sampling from the approximate density of the form (baseline Gaussian density) × (1 + (correction)), as the latter approach is in general unavailable. However, one may employ a methodology proposed by Davie (2022) who constructed a tractable sampling scheme via a corresponding density expansion in an elliptic setting, in a way so that the used expansion preserves a high order proximity in Wasserstein distance. Extension to hypo-elliptic SDEs is not straightforward and could be an interesting future research direction.

4.2.2. *Experimental Design and MCMC Results*

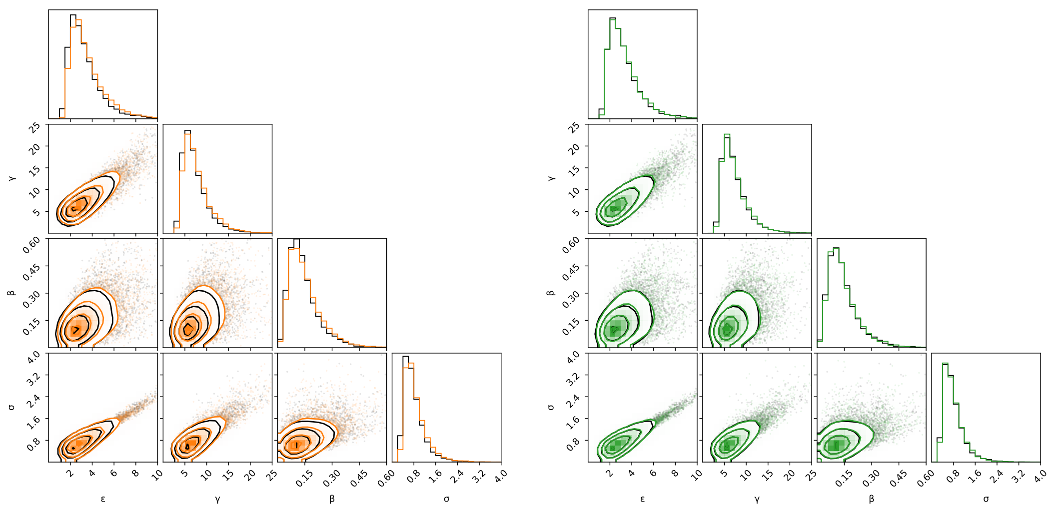
We apply our CF-expansion to carry out Bayesian inference for the FHN SDE (43) with the real dataset used in Samson, Tamborrino and Tubikanec (2025). The data are available at https://data.mendeley.com/datasets/ybhwtngzmm/1 which provides 20 neural recordings of the 5th lumbar dorsal rootlet from a single adult female rat with time length 250ms and equidistant step-size 0.02ms. In our study we choose a particular dataset, specifically the file 1554.mat from the above URL, which was obtained while the 5th lumber dermatome was stimulated. We subsample the first 40ms of data with a step-size 0.08ms, i.e. we have $(T, \Delta) = (40, 0.08)$ and a number of datapoints $n = 501$, so Δ is relatively large. As in Samson, Tamborrino and Tubikanec (2025), we set $s = 0$ and focus on the parameter $\theta = (\epsilon, \gamma, \beta, \sigma)$. We assume that the data $\mathscr{Y}_n = \{Y_{t_k}\}_{0\le k\le n}$ are observed with a small measurement noise as $Y_{t_k} = V_{t_k} + \epsilon_{t_k}$, with V the smooth coordinate in the FHN SDE and $\epsilon_{t_0}, \ldots, \epsilon_{t_n} \overset{i.i.d}{\sim} \mathscr{N}(0, 0.01^2)$. We adopt a non-centred parametrisation, assign log-normal priors on θ, i.e., $\log\varepsilon, \log\beta, \log\gamma, \log\sigma \overset{\text{i.i.d.}}{\sim} \mathscr{N}(0,1)$ and set $V_0 \sim \mathscr{N}(0, 0.1^2)$, $U_0 \sim \mathscr{N}(0, 0.2^2)$ for the initial state. We employ Hybrid Monte Carlo (HMC) to sample from the posterior, using the Python package Mici (https://pypi.org/project/mici/) which offers a variety of MCMC methods based on Hamiltonian dynamics. We use a dynamic integration-time HMC implementation (Betancourt, 2017) with a dual-averaging algorithm (Hoffman et al., 2014) to adapt the step-size of the leapfrog integrator. The mass matrix is set to identity.

We consider 3 designs of tractable posteriors: [**P0**] Benchmark 'true' posterior. This is constructed via a local Gaussian (LG) transition density scheme (Gloter and Yoshida, 2021), which provides an approximation of the transition density of the hypo-elliptic SDE (H) for a sufficiently small step-size. A data augmentation step is applied, whereby $d_M = 100$ signal points are added in-between observation pairs to eliminate the bias. The obtained posterior values are treated as the benchmark true values; [**P1**] Posterior based on the partial LDL scheme given in (45) with $\iota = \text{II}$; [**P2**] Posterior produced via implementation of the non-centred parameterisation of the initial target given by (46)-(47), based on the CF-expansion around the partial LDL scheme with $J = 3$. For each posterior, we run two HMC chains of 8,000 iterations with the first 4,000 iterations used as an adaptive warm-up phase.

Fig. 4 shows results for targets P0-P2. Results for the true posterior P0 are given in black and are overlaid in the sub-figures to observe the accuracy of posteriors P1-P2. Table 1 shows average running times per iteration from two chains. Additional convergence diagnostics provided in Table 1 of Section C.2 in Supplementary Material show similarly good convergence performance for all 3 cases, we can thus conclude that the posteriors shown in Fig. 4 are reliable. In Fig. 4 it is clear that P2 (i.e. scheme based upon the CF-expansion) captures P0 more accurately than P1 (i.e. scheme without correction terms) does. Thus, the inclusion of the correction term eliminates the bias even for $J = 3$, with the algorithm targeting P2 having a computing cost approximately 10 times smaller than that of the benchmark P0 (see Table 1). Our experiment implies that, in this case, the CF-expansion is effective both from the perspectives of computational cost and estimation accuracy. We remark that a centred-parametrisation led to MCMC chains with very poor convergence performance.

Table 1. Computational cost of MCMC chains for the FHN model. Schemes: P0$\rightarrow$ benchmark; P1$\rightarrow$ Modified LDL; P2 $\rightarrow$ CF-Expansion, $J = 3$.

scheme	P0	P1	P2
time(sec)/iter	4.745	0.237	0.460



(a) P1 (Partial LDL scheme) (b) P2 (CF-expansion with $J = 3$)

Figure 4. Posterior estimates. P0 (benchmark posterior) is overlaid in each figure in black.

5. Discussion on Practical Perspectives

1. *Data augmentation between data points*. We have developed a density expansion and discussed its use in statistical inference for the setting where the step-size Δ between observations is less than 1. In a general setting where $\Delta \geq 1$, one can still employ the density expansion in a *Data Augmentation* framework (Papaspiliopoulos, Roberts and Stramer (2013)), i.e., by imputing the latent variables between data points via time-discretisation. This can indeed be realised by generating a Markov chain of the baseline Gaussian scheme with the products of correction terms attached to the test function as a weight. As illustrated at the experiments in Section 4.2, the use of the correction terms can lead to efficient inference with a smaller number of discretisations in-between data points, compared to the case without the corrections. Similar efficiency gains were studied and illustrated in Iguchi, Beskos and Graham (2025) where a weak second-order sampling scheme is compared with the Euler-Maruyama (weak first order) one in a Bayesian data augmentation framework.

2. *Numerical properties of the LDL scheme.* Motivated by the use of the LDL scheme in a data augmentation framework, one may be interested in its numerical properties such as stability or (geometric) ergodicity. Though a full investigation is beyond the scope of this paper, we will make some comments below on the preservation of ergodicity by the LDL scheme. Let us consider the following Langevin-type equation:

$$dX_t = b(X_t)dt + \Sigma dB_t, \tag{48}$$

where $X_t \in \mathbb{R}^d$, $\Sigma \in \mathbb{R}^{d \times d}$ and $\{B_t\}_{t \geq 0}$ is a d-dimensional Wiener process. We assume standard sufficient conditions for (48) to be ergodic, specifically, (i) minorisation and (ii) Lyapunov condition, see e.g. Lemma 2.3. and Assumption 2.2., respectively, in Mattingly, Stuart and Higham

(2002). We will check if such conditions are inherited by the LDL scheme applied to (48). We first notice that this is not generally true for the Euler-Maruyama scheme unless the drift function is globally Lipschitz. In particular, the second condition can break down when the drift is only locally Lipschitz, while the minorisation condition can still hold, see, e.g., the proof of (Mattingly, Stuart and Higham, 2002, Corollary 7.4.), regardless of the growth of the drift. In a similar manner, the minorisation condition should hold for the LDL scheme as well, thus we focus on the Lyapunov condition. Here, we will show that such a preservation can occur for the LDL scheme applied to the following 1-dimensional SDE with non-globally Lipschitz drift:

$$dX_t = -X_t^3 dt + dB_{1,t}, \quad X_0 = x \in \mathbb{R}. \tag{49}$$

We note that (Mattingly, Stuart and Higham, 2002, Lemma 6.3.) proved that for any step-size and initial state, the Euler-Maruyama scheme applied to (49) can be unstable with a positive probability, thus does not preserve the ergodic property. Then, the LDL scheme $\{Y_k\}_{k\geq 0}$ is defined as follows, for $0 < t_{n-1} < t_n$ and $\Delta = t_n - t_{n-1}$:

$$Y_n = \exp(-3Y_{n-1}^2\Delta)Y_{n-1} + 2Y_{n-1}^3 \int_0^{\Delta} \exp(-3Y_{n-1}^2 s)ds + \int_{t_{n-1}}^{t_n} \exp(-3Y_{n-1}^2(t_n - s))dB_{1,s}. \tag{50}$$

We use the Lyapunov function $V(x) = x^2$ and write $\mathcal{F}_{n-1}$ as the σ-algebra generated by the Markov chain $\{Y_{t_k}\}_{k\leq n-1}$. We have that:

$$\mathbb{E}[V(Y_n)|\mathcal{F}_{t_{n-1}}] = m(\Delta, Y_{n-1})^2 + \int_0^{\Delta} \exp(-6Y_{n-1}^2 s)ds \leq m(\Delta, Y_{n-1})^2 + \Delta,$$

where $m(\Delta, Y_{n-1}) = \exp(-3Y_{n-1}^2\Delta)Y_{n-1} + 2Y_{n-1}^3 \int_0^{\Delta} \exp(-3Y_{n-1}^2 s)ds$. We derive an upper bound for $m(\Delta, Y_{n-1})^2$ and for a fixed $\varepsilon > 0$ consider the following three cases separately: (a) $Y_{n-1} = 0$; (b) $|Y_{n-1}| > \varepsilon$; (c) $|Y_{n-1}| \leq \varepsilon, Y_{n-1} \neq 0$. For (a), we immediately see that $m(\Delta, Y_{n-1}) = 0$. We note that, when $Y_{n-1} \neq 0$:

$$m(\Delta, Y_{n-1}) = \exp(-3Y_{n-1}^2\Delta)Y_{n-1} + 2Y_{n-1}^3 \times \frac{1-\exp(-3Y_{n-1}^2\Delta)}{3Y_{n-1}^2} = \Big(\tfrac{2}{3} + \tfrac{1}{3}\exp(-3Y_{n-1}^2\Delta)\Big)Y_{n-1}.$$

Therefore, for case (b), we have that:

$$m(\Delta, Y_{n-1})^2 \leq \Big(\tfrac{2}{3} + \tfrac{1}{3}\exp(-3\varepsilon^2\Delta)\Big)^2 \times Y_{n-1}^2 \equiv \rho \times V(Y_{n-1}),$$

with $\rho \in (0, 1)$. In case (c), it also follows that:

$$\begin{aligned} m(\Delta, Y_{n-1})^2 &= \rho \times Y_{n-1}^2 + 2\sqrt{\rho}\Big(\exp(-3Y_{n-1}^2\Delta) - \exp(-3\varepsilon^2\Delta)\Big)Y_{n-1}^2 \\ &\quad + \Big(\exp(-3Y_{n-1}^2\Delta) - \exp(-3\varepsilon^2\Delta)\Big)^2 Y_{n-1}^2 \\ &\leq \rho \times V(Y_{n-1}) + (2\sqrt{\rho} + 1)\varepsilon^2. \end{aligned}$$

We thus conclude that the discrete-time Lyapunov condition holds for (50), i.e., there exists $\alpha \in (0, 1)$ and $\beta \geq 0$ s.t.:

$$\mathbb{E}[V(Y_n)|\mathcal{F}_{n-1}] \leq \alpha V(Y_{n-1}) + \beta, \qquad \forall n \in \mathbb{N}.$$

In summary, due to minorisation and discrete-time Lyapunov conditions, the LDL scheme (50) preserves (geometric) ergodicity for the SDE (49) with locally Lipschitz drift. This example is used as an indication that the LDL scheme can preserve ergodicity for general SDEs with non-globally Lipschitz drift. Detailed analysis is left as future work.

3. *Design of local drift linearisation – choice of matrix A*. In the development of the density expansion, we considered full-drift linearisation, i.e., first-order Taylor expansion of the drift for all coordinates to define the matrix A. However, as mentioned in Section 4.1, one can also consider a partially linearised drift approximation, e.g., with the matrix A being upper-triangular, in order to reduce the computational cost of calculating $\exp(A)$, as long as the vector field defined in the baseline scheme satisfies Hörmander's condition as in Lemma 2.5. For preservation of hypoellipticity, at least linearisation of $V_{S,0}$ (drift of the smooth component X_S) w.r.t. X_R is required so that the noise in the rough component X_R is lifted to X_S. The optimal way of linearisation (choice of A) would depend on the model at hand, but if a user considers a lower level of density expansion 'J', say, $J = 2, 3, 4$, which is indeed sufficient to see improvements in estimation accuracy, then the partial drift linearisation will be a better option in terms of computational cost; recall e.g., DE-II in Figure 3-(b).

4. *Computational cost w.r.t. the state dimension.* We stress that our CF-expansion converges exponentially fast with $J \geq 1$, so small values of J will typically provide accurate proxies. Such a consideration counterbalances the computing cost for increasingx state dimension N. Following the analytical expressions of the e_k'e for the FHN model in Section C of Supplementary Material, in the case of additive noise, one has $e_1 = e_2 = 0$, while the calculation of e_5 requires all 3rd order derivatives of the baseline Gaussian density, at a cost of $O(N^3)$. An extra derivative is added in the calculation when increasing k in e_k by one. Note that calculations involving just the baseline Gaussian transition density will typically already involve costs of $O(N^3)$ due to matrix inversions, so in the additive noise setting using $J = 5$ will not increase computing costs vs $J = 0$ as an order of N.

6. Conclusion

We propose a new CF-expansion for the transition density of multivariate SDEs over a time interval with fixed length $\Delta \in (0, 1)$, of the form 'baseline Gaussian density' × 'correction term', where the 'correction term' involves quantities of size $\Delta^{j/2}$, $j = 0, \ldots, J$, for $J \geq 1$. Analytical expressions can be obtained via any software that carries out symbolic calculations. We have shown analytically that the error has a size of $O(\Delta^{(J+1)/2})$. The proposed CF-expansion covers hypo-elliptic classes of SDEs, whereas most of the developments in earlier works are dedicated to elliptic SDEs. In the numerical studies in Section 4 the errors from our CF-expansion are fastly eliminated as J increases for a fixed $\Delta \in (0, 1)$.

We also mention the following. First, we take the direction described in the paper to produce our CF-expansion because potential alternative approaches used in the literature for the elliptic class (involving, e.g., Malliavin calculus) are arguably much more challenging in terms of producing a practical and theoretically validated methodolology. Second, several recent works on the theme of parameter inference for hypo-elliptic SDEs produce methodology and analytical results in the high-frequency scenario $\Delta \to 0$, see e.g. Ditlevsen and Samson (2019), Gloter and Yoshida (2021), Iguchi, Beskos and Graham (2024), Iguchi and Beskos (2025a), Iguchi, Beskos and Graham (2025), Melnykova (2020), Pilipovic, Samson and Ditlevsen (2024). Then, numerical experiments are used to check the precision for a fixed

$\Delta > 0$ given in practice. In contrast, our contribution assumes a fixed Δ, thus is expected to be more robust in deviations of Δ from 0 than high-frequency approaches.

Acknowledgments

We are grateful to the Editor-in-Chief, Associate Editor, and two anonymous referees for their constructive suggestions that enhanced the quality of this paper. We also thank Dr Karen Habermann for suggesting related references on closed-form heat kernel expansions of hypo-elliptic diffusions.

Funding

YI acknowledges the supports from the Additional Funding Programme for Mathematical Sciences, delivered by EPSRC (EP/V521917/1) and the Heilbronn Institute for Mathematical Research. YI is also supported by the EPSRC grant Prob AI (EP/Y028783/1).

Supplementary Material

"Supplementary Material" (Iguchi and Beskos, 2025b) provides the proofs of the main results and related technical proofs, and also contains an additional numerical experiment and supporting information of density expansions employed in Section 4.

Appendix A: Proof of Lemma 2.5

We write the LDL scheme (15) with a frozen variable $z \in \mathbb{R}^N$ as the following differential form:

$$d\bar{X}_t^{\theta,z} = V_0^z(\bar{X}_t^{\theta,z}, \theta)dt + \sum_{j=1}^{d} V_j^z(\bar{X}_t^{\theta,z}, \theta)dB_{j,t}, \quad X_0 = x,$$

for $(x, \theta) \in \mathbb{R}^N \times \Theta$, where we have set:

$$V_0^z(x,\theta) \equiv A_{z,\theta}x + b_{z,\theta}, \qquad V_j^z(x,\theta) \equiv V_j(z,\theta), \quad 1 \le j \le d. \tag{51}$$

Since the diffusion coefficients are independent of the state $\bar{X}_t^{\theta,z}$, the above Itô-type SDE identifies with the Stratonovich one. We show under Assumption 2.2 that the vector fields determined from the coefficients of the above SDE satisfy Hörmander's condition for each model class E and H.

Elliptic model E. We immediately have from Assumption 2.2-I that

$$\mathrm{Span}\big\{V_j^z(x,\theta)|_{z=x}, \ \ 1 \le j \le d\big\} = \mathbb{R}^N, \tag{52}$$

for all $(x, \theta) \in \mathbb{R}^N \times \Theta$.

Hypo-elliptic model H. We firstly note that

$$\widetilde{V}_{S,0}(x,\theta) \equiv V_{S,0}(x,\theta), \quad V_j^z(x,\theta) \equiv \big[\mathbf{0}_{N_S}^\top, V_{R,j}(z,\theta)^\top\big]^\top, \qquad 1 \le j \le d.$$

Then Assumption 2.2-II is equivalent to the following condition:

$$\begin{aligned}&\mathrm{Span}\big\{V_{R,j}(x,\theta),\ 1\le j\le d\big\}=\mathbb{R}^{N_R};\\ &\mathrm{Span}\big\{\mathrm{Proj}_{1,N_S}\{[\widetilde{V}_0,V_j](x,\theta)\},\ 1\le j\le d\big\}=\mathrm{Span}\big\{\partial_{x_R}^{\top}V_{S,0}(x,\theta)V_{R,j}(x,\theta),\ 1\le j\le d\big\}=\mathbb{R}^{N_S}.\end{aligned} \tag{53}$$

The condition (53) leads to:

$$\mathrm{Span}\big\{V_{R,j}(z,\theta)|_{z=x},\ 1\le j\le d\big\}=\mathbb{R}^{N_R},$$

and

$$\begin{aligned}&\mathrm{Span}\big\{\mathrm{Proj}_{1,N_S}\{[V_0^z,V_j^z](x,\theta)|_{z=x}\},\ 1\le j\le d\big\}=\mathrm{Span}\big\{\mathrm{Proj}_{1,N_S}\{A_{z,\theta}V_j^z(x,\theta)|_{z=x}\},\ 1\le j\le d\big\}\\ &\quad=\mathrm{Span}\big\{\partial_{x_R}^{\top}V_{S,0}(z,\theta)V_{R,j}(z,\theta)|_{z=x},\ 1\le j\le d\big\}=\mathbb{R}^{N_S}.\end{aligned}$$

Thus, we obtain:

$$\mathrm{Span}\Big\{\big\{V_j^z(x,\theta),[V_0^z,V_j^z](x,\theta)\big\}|_{z=x},\ 1\le j\le d\Big\}=\mathbb{R}^N, \tag{54}$$

for all $(x,\theta)\in\mathbb{R}^N\times\Theta$. The proof is now complete.

Appendix B: Proof of Lemma 2.6

We define

$$G(s)\equiv\int_{\mathbb{R}^N}p_{\Delta-s}^{\bar{X}^z}(\xi,y;\theta)p_s^X(x,\xi;\theta)|_{z=x}\,d\xi,\qquad s\ge 0.$$

Noticing that $p_0^X(x,y;\theta)=\delta_y(x)$ and $p_0^{\bar{X}}(x,y;\theta)=\delta_y(x)$, we have

$$p_\Delta^X(x,y;\theta)-p_\Delta^{\bar{X}^z}(x,y;\theta)|_{z=x}=G(\Delta)-G(0)=\int_0^\Delta G'(s)ds.$$

Also, note that the transition densities $p_\Delta^X(x,y;\theta)$ and $p_\Delta^{\bar{X}^z}(x,y;\theta)$ satisfy the following backward/forward Kolmogorov equations:

$$\partial_t\,p_t^X(x,y;\theta)=\mathscr{L}_\theta\{p_t^X(\cdot,y;\theta)\}(x),\qquad \partial_t\,p_t^X(x,y;\theta)=\mathscr{L}_\theta^*\{p_t^X(x,\cdot;\theta)\}(y); \tag{55}$$

$$\partial_t\,p_t^{\bar{X}^z}(x,y;\theta)=\mathscr{L}_\theta^{0,z}\{p_t^{\bar{X}^z}(\cdot,y;\theta)\}(x). \tag{56}$$

It then follows that:

$$\begin{aligned}G'(s)=&-\int_{\mathbb{R}^N}\mathscr{L}_\theta^{0,z}\{p_{\Delta-s}^{\bar{X}^z}(\cdot,y;\theta)\}(\xi)p_s^X(x,\xi;\theta)|_{z=x}d\xi\\ &\qquad+\int_{\mathbb{R}^N}p_{\Delta-s}^{\bar{X}^z}(\xi,y;\theta)\mathscr{L}_\theta^*\{p_s^X(x,\cdot;\theta)\}(\xi)|_{z=x}d\xi\\ =&-\int_{\mathbb{R}^N}\mathscr{L}_\theta^{0,z}\{p_{\Delta-s}^{\bar{X}^z}(\cdot,y;\theta)\}(\xi)p_s^X(x,\xi;\theta)|_{z=x}d\xi\\ &\qquad+\int_{\mathbb{R}^N}\mathscr{L}_\theta\{p_{\Delta-s}^{\bar{X}^z}(\cdot,y;\theta)\}(\xi)p_s^X(x,\xi;\theta)|_{z=x}d\xi.\end{aligned}$$

The proof is now complete.

Appendix C: Proof of Lemma 2.7

We focus on showing the formula (20), and (21) is obtained from a similar argument. We make use of the approach used in Iguchi and Yamada (2021, Proposition 2.1). We define

$$g^{\widetilde{\mathscr{L}}_\theta^z}(s) \equiv \bar{P}_s^{\theta,z} \widetilde{\mathscr{L}}_\theta^z \bar{P}_{t-s}^{\theta,z} \varphi(x) = \int_{\mathbb{R}^N} \widetilde{\mathscr{L}}_\theta^z \bar{P}_{t-s}^{\theta,z} \varphi(\xi_1)\, p_s^{\bar{X}^z}(x, \xi_1; \theta)\, d\xi_1 \tag{57}$$

and consider the Taylor expansion of $g^{\widetilde{\mathscr{L}}_\theta^z}(s)$ at $s = 0$:

$$g^{\widetilde{\mathscr{L}}_\theta^z}(s) = \sum_{k=0}^{J} \partial^k g^{\widetilde{\mathscr{L}}_\theta^z}(s)|_{s=0} \times \frac{s^k}{k!} + \mathscr{R}^{J+1,z}(s,t,x;\theta), \qquad J \in \mathbb{N},$$

where $\mathscr{R}^{J+1,z}(s,t,x;\theta) \equiv s^{J+1} \int_0^1 \partial^{J+1} g^{\widetilde{\mathscr{L}}_\theta^z}(su) \frac{(1-u)^J}{J!} du$. We have from (57) that

$$\begin{aligned}
\partial g^{\widetilde{\mathscr{L}}_\theta^z}(s) &= \int_{\mathbb{R}^N} \widetilde{\mathscr{L}}_\theta^z \bar{P}_{t-s}^{\theta,z} \varphi(\xi_1)\, \partial_s p_s^{\bar{X}^z}(x, \xi_1; \theta)\, d\xi_1 + \int_{\mathbb{R}^N} \widetilde{\mathscr{L}}_\theta^z \partial_s \bar{P}_{t-s}^{\theta,z} \varphi(\xi_1)\, p_s^{\bar{X}^z}(x, \xi_1; \theta)\, d\xi_1 \\
&= \int_{\mathbb{R}^N} \widetilde{\mathscr{L}}_\theta^z \bar{P}_{t-s}^{\theta,z} \varphi(\xi_1)\, (\mathscr{L}_\theta^{0,z})^* \{p_s^{\bar{X}^z}(x, \cdot; \theta)\}(\xi_1)\, d\xi_1 - \int_{\mathbb{R}^N} \widetilde{\mathscr{L}}_\theta^z \mathscr{L}_\theta^{0,z} \bar{P}_{t-s}^{\theta,z} \varphi(\xi_1)\, p_s^{\bar{X}^z}(x, \xi_1; \theta)\, d\xi_1 \\
&= \int_{\mathbb{R}^N} [\mathscr{L}_\theta^{0,z}, \widetilde{\mathscr{L}}_\theta^z] \bar{P}_{t-s}^{\theta,z} \varphi(\xi_1)\, p_s^{\bar{X}^z}(x, \xi_1; \theta)\, d\xi_1 \\
&= g^{[\mathscr{L}_\theta^{0,z}, \widetilde{\mathscr{L}}_\theta^z]}(s),
\end{aligned}$$

where we made use of (56) and integration by parts in the second and third lines, respectively. Thus, the higher-order derivatives of $g^{\widetilde{\mathscr{L}}_\theta^z}(s)$ are given as:

$$\partial^k g^{\widetilde{\mathscr{L}}_\theta^z}(s) = g^{(\mathrm{ad}_{\mathscr{L}_\theta^{0,z}})^k (\widetilde{\mathscr{L}}_\theta^z)}(s) = \bar{P}_s^{\theta,z} \big\{ \big(\mathrm{ad}_{\mathscr{L}_\theta^{0,z}}\big)^k (\widetilde{\mathscr{L}}_\theta^z) \big\} \bar{P}_{t-s}^{\theta,z} \varphi(x), \qquad 0 \le k \le J,$$

and then,

$$\partial^k g^{\widetilde{\mathscr{L}}_\theta^z}(s)|_{s=0} = \big\{ \big(\mathrm{ad}_{\mathscr{L}_\theta^{0,z}}\big)^k (\widetilde{\mathscr{L}}_\theta^z) \big\} \bar{P}_t^{\theta,z} \varphi(x).$$

The proof is now complete.

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.

BARILARI, D. and PAOLI, E. (2017). Curvature terms in small time heat kernel expansion for a model class of hypoelliptic Hörmander operators. *Nonlinear Analysis* **164** 118–134.

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

CHANG, J. and CHEN, S. X. (2011). On the approximate maximum likelihood estimation for diffusion processes. *Ann. Statist.* **39** 2820-2851.

CHOI, S. (2015). Explicit form of approximate transition probability density functions of diffusion processes. *J. Econom.* **187** 57–73.

DAVIE, A. M. (2022). Approximation of SDE solutions using local asymptotic expansions. *Stochastics and Partial Differential Equations: Analysis and Computations* **10** 720–734.

DEVILLE, L., VANDEN-EIJNDEN, E. and MURATOV, C. (2005). Two distinct mechanisms of coherence in randomly perturbed dynamical systems. *Phys. Rev. E* **72** 031105.

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. (2021). Adaptive estimation for degenerate diffusion processes. *Electron. J. Stat.* **15** 1424–1472.

HABERMANN, K. (2019). Small-time fluctuations for the bridge in a model class of hypoelliptic diffusions of weak Hörmander type. *Electronic Journal of Probability* **24**. https://doi.org/10.1214/19-ejp274

HOFFMAN, M., 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). Parameter inference for degenerate diffusion processes. *Stoch. Process. Their Appl.* 104384.

IGUCHI, Y. and BESKOS, A. (2025a). Parameter inference for hypo-elliptic diffusions under a weak design condition. *Electron. J. Stat. (to appear)*.

IGUCHI, Y. and BESKOS, A. (2025b). Supplementary Material for “A closed-form transition density expansion for elliptic and hypo-elliptic SDEs".

IGUCHI, Y., BESKOS, A. and GRAHAM, M. (2025). Parameter estimation with increased precision for elliptic and hypo-elliptic diffusions. *Bernoulli* **31** 333–358.

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.

JIMENEZ, J. C., MORA, C. and SELVA, M. (2017). A weak Local Linearization scheme for stochastic differential equations with multiplicative noise. *Journal of Computational and Applied Mathematics* **313** 202–217. https://doi.org/10.1016/j.cam.2016.09.013

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

LEE, Y. D., SONG, S. and LEE, E. K. (2014). The delta expansion for the transition density of diffusion models. *J. Econom.* **178** 694–705.

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

LI, C. and CHEN, D. (2016). Estimating jump–diffusions using closed-form likelihood expansions. *J. Econom.* **195** 51–70.

MATTINGLY, J. C., STUART, A. M. and HIGHAM, D. J. (2002). Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stochastic Processes and their Applications* **101** 185–232. https://doi.org/10.1016/s0304-4149(02)00150-3

MELNYKOVA, A. (2020). Parametric inference for hypoelliptic ergodic diffusions with full observations. *Stat. Inference Stoch. Process.* **23** 595–635.

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

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

PAPASPILIOPOULOS, O., ROBERTS, G. O. and STRAMER, O. (2013). Data augmentation for diffusions. *Journal of Computational and Graphical Statistics* **22** 665–688. https://doi.org/10.1080/10618600.2013.783484

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

PIGATO, P. (2022). Density estimates and short-time asymptotics for a hypoelliptic diffusion process. *Stoch. Process. Their Appl.* **145** 117–142.

PILIPOVIC, P., SAMSON, A. and DITLEVSEN, S. (2024). Strang Splitting for Parametric Inference in Second-order Stochastic Differential Equations. *arXiv:2405.03606*.

ROBERTS, G. and STRAMER, O. (2001). On inference for partially observed nonlinear diffusion models using the Metropolis–Hastings algorithm. *Biometrika* **88** 603–621.

SAMSON, A., TAMBORRINO, M. and TUBIKANEC, I. (2025). Inference for the stochastic FitzHugh-Nagumo model from real action potential data via approximate Bayesian computation. *Computational Statistics and Data Analysis* **204** 108095. https://doi.org/10.1016/j.csda.2024.108095

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.

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

YANG, N., CHEN, N. and WAN, X. (2019). A new delta expansion for multivariate diffusions via the Itô-Taylor expansion. *J. Econom.* **209** 256–288.

YOSHIDA, N. (1992). Asymptotic expansions of maximum likelihood estimators for small diffusions via the theory of Malliavin-Watanabe. *Probab. Theory Relat. Fields* **92** 275–311.