

# Parameter inference for hypo-elliptic diffusions under a weak design condition

Yuga Iguchi<sup>1</sup> and Alexandros Beskos<sup>1</sup>

<sup>1</sup>Department of Statistical Science, University College London, United Kingdom,  
 e-mail: [yuga.iguchi.21@ucl.ac.uk](mailto:yuga.iguchi.21@ucl.ac.uk); [a.beskos@ucl.ac.uk](mailto:a.beskos@ucl.ac.uk)

**Abstract:** We address the problem of parameter estimation for degenerate diffusion processes defined via the solution of Stochastic Differential Equations (SDEs) with diffusion matrix that is not full-rank. For this class of hypo-elliptic SDEs recent works have proposed contrast estimators that are asymptotically normal, provided that the step-size in-between observations  $\Delta = \Delta_n$  and their total number  $n$  satisfy  $n \rightarrow \infty$ ,  $n\Delta_n \rightarrow \infty$ ,  $\Delta_n \rightarrow 0$ , and additionally  $\Delta_n = o(n^{-1/2})$ . This latter restriction places a requirement for a so-called ‘rapidly increasing experimental design’. In this paper, we relax this limitation and develop a general contrast estimator satisfying asymptotic normality under the weaker condition  $\Delta_n = o(n^{-1/p})$  for general  $p \geq 2$ . Such a result has been obtained for elliptic SDEs, but its derivation in a hypo-elliptic setting is highly non-trivial and requires a separate approach. We provide numerical results to illustrate the advantages of the developed theory.

**MSC2020 subject classifications:** 60J60, 62F12.

**Keywords and phrases:** Stochastic differential equation, hypo-elliptic diffusion, parameter estimation, design condition.

Received January 2024.

## 1. Introduction

We consider the problem of parameter inference for multivariate Stochastic Differential Equations (SDEs) determined via a degenerate diffusion matrix. Let  $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$  be a filtered probability space and let  $W = (W_t^1, \dots, W_t^d)$ ,  $t \geq 0$ , be the  $d$ -dimensional standard Brownian motion,  $d \geq 1$ , defined on such a space. We introduce  $N$ -dimensional SDE models,  $N \geq 1$ :

$$dX_t = \mu(X_t, \theta)dt + \sum_{j=1}^d A_j(X_t, \theta)dW_t^j, \quad X_0 = x \in \mathbb{R}^N, \quad (1)$$

with  $\mu, A_j(\cdot, \theta) : \mathbb{R}^N \rightarrow \mathbb{R}^N$ ,  $1 \leq j \leq d$ , for parameter  $\theta$ . This work focuses on the case where the diffusion matrix  $A = [A_1, \dots, A_d]$  is not of full-rank, i.e. the matrix  $a = AA^\top$  is not positive definite. The law of the solution to (1) is assumed to be absolutely continuous with respect to (w.r.t.) the Lebesgue measure. Such models are referred to as *hypo-elliptic diffusions* and constitute

an important class of Markovian processes used in a wide range of applications. E.g., this class includes the (underdamped) Langevin equation, the Synaptic-Conductance model [6], the NL CAR( $p$ ) model [27, 26], the Jansen-Rit neural mass model [5], the SIR epidemiological model [25] and the quasi-Markovian generalised Langevin equations ( $q$ -GLE) [29]. We investigate two (sub-)classes of hypo-elliptic SDEs which cover a wide range of models, including the ones mentioned in the above examples. The first model class is specified via the following SDE:

$$\begin{aligned} dX_t &= \begin{bmatrix} dX_{S,t} \\ dX_{R,t} \end{bmatrix} = \begin{bmatrix} \mu_S(X_t, \beta_S) \\ \mu_R(X_t, \beta_R) \end{bmatrix} dt + \sum_{j=1}^d \begin{bmatrix} \mathbf{0}_{N_S} \\ A_{R,j}(X_t, \sigma) \end{bmatrix} dW_t^j, \\ X_0 = x &\equiv \begin{bmatrix} x_S \\ x_R \end{bmatrix} \in \mathbb{R}^N. \end{aligned} \quad (\text{Hypo-I})$$

The involved drift functions and diffusion coefficients are specified as follows:

$$\begin{aligned} \mu_S : \mathbb{R}^N \times \Theta_{\beta_S} &\rightarrow \mathbb{R}^{N_S}, \quad \mu_R : \mathbb{R}^N \times \Theta_{\beta_R} \rightarrow \mathbb{R}^{N_R}, \\ A_{R,j} : \mathbb{R}^N \times \Theta_\sigma &\rightarrow \mathbb{R}^{N_R}, \quad 1 \leq j \leq d, \end{aligned}$$

for positive integers  $N_S, N_R$  such that  $N = N_S + N_R$ . Here, the unknown parameter vector is

$$\theta = (\beta_S, \beta_R, \sigma) \in \Theta = \Theta_{\beta_S} \times \Theta_{\beta_R} \times \Theta_\sigma \subseteq \mathbb{R}^{N_{\beta_S}} \times \mathbb{R}^{N_{\beta_R}} \times \mathbb{R}^{N_\sigma},$$

for positive integers  $N_{\beta_S}, N_{\beta_R}, N_\sigma$  satisfying  $N_{\beta_S} + N_{\beta_R} + N_\sigma = N_\theta$ . The parameter space  $\Theta$  is assumed to be a compact subset of  $\mathbb{R}^{N_\theta}$ . Later on we place a condition on (Hypo-I), related to the so-called *Hörmander's condition*, so that such an SDE gives rise to a hypo-elliptic process, i.e. its finite-time transition distributions admit a Lebesgue density. In brief, the condition will guarantee that randomness from the rough component  $X_{R,t}$  propagates into the smooth component  $X_{S,t}$  via the drift  $\mu_S$ . The second hypo-elliptic model class we treat in this work is specified via the following SDE:

$$\begin{aligned} dX_t &= \begin{bmatrix} dX_{S_1,t} \\ dX_{S_2,t} \\ dX_{R,t} \end{bmatrix} = \begin{bmatrix} \mu_{S_1}(X_{S,t}, \beta_{S_1}) \\ \mu_{S_2}(X_{S,t}, \beta_{S_2}) \\ \mu_R(X_t, \beta_R) \end{bmatrix} dt + \sum_{j=1}^d \begin{bmatrix} \mathbf{0}_{N_{S_1}} \\ \mathbf{0}_{N_{S_2}} \\ A_{R,j}(X_t, \sigma) \end{bmatrix} dW_t^j, \\ X_0 = x &\equiv \begin{bmatrix} x_{S_1} \\ x_{S_2} \\ x_R \end{bmatrix} \in \mathbb{R}^N, \end{aligned} \quad (\text{Hypo-II})$$

where we have set  $X_{S,t} = [X_{S_1,t}^\top, X_{S_2,t}^\top]^\top$ . Also, we now have the parameter vector:

$$\begin{aligned} \theta = (\beta_{S_1}, \beta_{S_2}, \beta_R, \sigma) &\in \Theta = \Theta_{\beta_{S_1}} \times \Theta_{\beta_{S_2}} \times \Theta_{\beta_R} \times \Theta_\sigma \\ &\subseteq \mathbb{R}^{N_{\beta_{S_1}}} \times \mathbb{R}^{N_{\beta_{S_2}}} \times \mathbb{R}^{N_{\beta_R}} \times \mathbb{R}^{N_\sigma}, \end{aligned}$$

with integers  $N_{\beta_{S_1}}, N_{\beta_{S_2}}$  such that  $N_{\beta_S} = N_{\beta_{S_1}} + N_{\beta_{S_2}}$ , and the drift functions are now specified as:

$$\mu_{S_1} : \mathbb{R}^{N_S} \times \Theta_{N_{\beta_{S_1}}} \rightarrow \mathbb{R}^{N_{S_1}}, \quad \mu_{S_2} : \mathbb{R}^N \times \Theta_{N_{\beta_{S_2}}} \rightarrow \mathbb{R}^{N_{S_2}},$$

with integers  $N_{S_1}, N_{S_2} \geq 1$  so that  $N_{S_1} + N_{S_2} = N_S$ . Again,  $\Theta$  is assumed to be compact. Note that the drift  $\mu_{S_1}$  depends on the smooth component  $X_{S,t}$  and not on  $X_{R,t}$ , thus randomness from  $X_{R,t}$  does not directly propagate onto  $X_{S_1,t}$ . We refer to an SDE of the form (Hypo-II) as a *highly degenerate diffusion*. Throughout the paper the class (Hypo-II) is treated separately from (Hypo-I). Later on we introduce a Hörmander-type condition guaranteeing that (Hypo-II) gives rise to hypo-elliptic SDEs.

We consider parameter estimation for the classes (Hypo-I), (Hypo-II), given discrete-time observations of the full vector  $X_t$  at instances  $t_i = i\Delta_n$ ,  $1 \leq i \leq n$ , for step-size  $\Delta_n > 0$ . We develop contrast estimators for models in (Hypo-I), (Hypo-II), and study their asymptotic properties under a *high-frequency complete observation* scenario. In particular, we assume the setting  $n \rightarrow \infty$ ,  $\Delta_n \rightarrow 0$ ,  $n\Delta_n \rightarrow \infty$ . Over the last two-three decades, numerous works have studied parametric inference for diffusion processes in a high-frequency setting. These works were initially focused on *elliptic* diffusions, i.e. SDEs in (1) with a positive definite matrix  $a = AA^\top$ . The work in [16] proposed a contrast estimator for scalar elliptic diffusions and proved asymptotic normality (via a CLT) under the design condition  $\Delta_n = o(n^{-1/p})$ ,  $p \geq 2$ . [28] extended such a result to multivariate elliptic SDEs and proposed an adaptive-type contrast estimator achieving a CLT under the same design condition as in [16]. In contrast, hypo-elliptic SDEs have been relatively under-explored until recently, though an interesting empirical study, without analytical results, was provided by [22].

To obtain asymptotic results in the hypo-elliptic setting similar to the ones available for elliptic SDEs, one must deal with a number of challenges. A standard Euler-Maruyama discretisation leads to a Dirac measure due to the involved degenerate diffusion matrix. This can be resolved by considering a higher-order Itô-Taylor expansion for  $X_{S,t}$  that propagates additional Gaussian variates onto the smooth components. For class (Hypo-I) such a direction leads to a discretisation scheme with Gaussian increments of size  $\mathcal{O}(\Delta_n^{3/2})$  for the smooth components. However, recent works [6, 8, 13] have highlighted that introduction of high-order Gaussian variates in the smooth components must be accompanied by an appropriate high-order mean approximation of the rough components. If such a balance is not achieved, estimation of  $\beta_R$  becomes biased. Within class (Hypo-I), [6] developed a non-degenerate discretisation scheme in the setting of  $N_S = 1$  and of the diffusion matrix  $A$  being diagonal. [6] proposed contrast estimators separately for  $\beta_S$  and  $(\beta_R, \sigma)$  and proved a CLT under  $\Delta_n = o(n^{-1/2})$ , i.e. for a ‘rapidly increasing experimental design’ [23]. [7, 8] addressed the issue of disjoint estimation, within (Hypo-I), by working with an approximate Gaussian density for the complete vector  $X_t$  and providing a joint contrast estimator for the full parameter vector  $(\beta_R, \beta_S, \sigma)$ . They showed that the estimator is asymptotically normal under the same design condition  $\Delta_n = o(n^{-1/2})$ . For

bivariate models in class (Hypo-I), [18] used a local linearisation [4] of the drift function to obtain a non-degenerate Gaussian discretisation scheme and construct a contrast estimator attaining a CLT under  $\Delta_n = o(n^{-1/2})$ . For class (Hypo-II), [13] worked in a general multivariate setting and established a joint contrast estimator achieving a CLT under  $\Delta_n = o(n^{-1/2})$ . The design condition  $\Delta_n = o(n^{-1/2})$  assumed in the above works can be restrictive in practice. Indicatively, if a user tries to design a dataset with a large interval  $T_n = n\Delta_n$  and a large number of samples  $n$  to obtain accurate estimation results, then  $\Delta_n$  should be set to a quite small value so that condition  $n\Delta_n^2 = T_n\Delta_n \rightarrow 0$  is satisfied, e.g. for  $T_n = 1000$ ,  $\Delta_n$  could be much less than  $10^{-3}$ . Datasets with such small step-sizes are not always available in applications.

An apparent open question in the hypo-elliptic setting is the weakening of the design condition  $\Delta_n = o(n^{-1/2})$ . Indicatively, [18, 7] required  $\Delta_n = o(n^{-1/2})$  to control terms of size  $\mathcal{O}(\Delta_n^{-q})$ ,  $q \geq 1/2$ , when proving consistency for their contrast estimators. Thus, to arrive to a CLT under a weaker design condition, one needs to develop a different approach compared to previous works to prove consistency without requiring that  $\Delta_n = o(n^{-1/2})$ . Furthermore, the construction of a general contrast function for degenerate SDEs is not straightforward due to the degenerate structure of the diffusion matrix. For class (Hypo-I), [14] developed a contrast estimator achieving the CLT under the weaker condition  $\Delta_n = o(n^{-1/3})$  via a novel closed-form transition density expansion for SDEs within (Hypo-I). However, a general asymptotic theory under the condition  $\Delta_n = o(n^{-1/p})$ ,  $p \geq 2$ , has yet to be established for degenerate SDEs.

Our work closes the above gap in the research of hypo-elliptic SDEs. We propose a general contrast estimator for a wide class of hypo-elliptic diffusion models and show its asymptotic normality under the weaker design condition  $\Delta_n = o(n^{-1/p})$ ,  $p \geq 2$ . Specifically, our contributions include the following:

- (a) We develop two contrast estimators for the two model classes (Hypo-I) and (Hypo-II), for the purposes of joint estimation of the unknown parameter vector  $(\beta_S, \beta_R, \sigma)$ . The contrast functions are based on approximate log-likelihood terms which we construct by making use of Gaussian approximations with high-order mean and variance expansions, while at the same time dealing with the degenerate structure of the involved SDEs.
- (b) We show that within the high-frequency, complete observation regime, a CLT is obtained provided that the design condition  $\Delta_n = o(n^{-1/p})$ , for  $p \geq 2$ , is satisfied. To the best of our knowledge, this is the first work to define a contrast estimator achieving a CLT under the weak design condition that  $n\Delta_n^p \rightarrow 0$ , for arbitrarily large  $p \geq 2$ , in the hypo-elliptic setting. For reference, Table 1 summarises existing works with corresponding design conditions, including our contribution in this setting.
- (c) We provide numerical experiments to demonstrate that the proposed estimator is asymptotically unbiased in the high-frequency complete observation regime under the weaker design condition  $\Delta_n = o(n^{-1/p})$ , for  $p \geq 3$ . The numerical results highlight that the estimators requiring  $\Delta_n = o(n^{-1/2})$  proposed in the literature can indeed suffer from bias

TABLE 1  
*Contrast estimators for Hypo-elliptic SDEs (high-frequency, complete observation regime).*

Work	Model	Design Condition on $\Delta_n$ for CLT
[6]	(Hypo-I) with $N_S = 1$ . Diffusion matrix is diagonal.	$\Delta_n = o(n^{-1/2})$
[18]	(Hypo-I) with $N_S = N_R = 1$ .	$\Delta_n = o(n^{-1/2})$
[7, 8]	(Hypo-I)	$\Delta_n = o(n^{-1/2})$
[14]	(Hypo-I)	$\Delta_n = o(n^{-1/3})$
[13]	(Hypo-II)	$\Delta_n = o(n^{-1/2})$
This paper	(Hypo-I) & (Hypo-II)	$\Delta_n = o(n^{-1/p}), p \geq 2$

when  $n\Delta_n^2$  is not sufficiently small.

- (d) Our methodology is relevant beyond the setting of high-frequency and/or complete observations. The developed Gaussian approximation of the SDE transition density can be used as part of a broader data augmentation algorithm (e.g., MCMC or EM) given low-frequency and/or partial observations. Analytical results for such a setting are beyond the scope of this paper. We provide a numerical result showcasing the advantage of the use of the developed approximate log-likelihood under the weaker condition  $\Delta_n = o(n^{-1/p})$ ,  $p \geq 3$ , in a setting of high-frequency partial observations.

To add to point (d) above, often in applications only smooth components are observed. [6, 13] have shown empirically that filtering procedures incorporating the developed approximate likelihood of the full state vector (within a data augmentation approach) can lead to asymptotically unbiased parameter estimation in a partial observation regime.

The rest of the paper is organised as follows. Section 2 prepares some conditions related to the hypo-ellipticity of models (Hypo-I) and (Hypo-II). Section 3 develops the new contrast estimators and states their asymptotic properties, thus providing the main results of this work. Simulation studies are shown in Section 4 (the code is available at <https://github.com/YugaIgu/Parameter-estimation-hypo-SDEs>). The proofs for the main results are collected in Section 5. Section 6 concludes the paper.

**Notation.** For the model class (Hypo-II), and for  $x = (x_S, x_R) \in \mathbb{R}^{N_S} \times \mathbb{R}^{N_R} = \mathbb{R}^N$ , we write:

$$\beta_S = (\beta_{S_1}, \beta_{S_2}) \in \Theta_{\beta_S} = \Theta_{\beta_{S_1}} \times \Theta_{\beta_{S_2}}, \quad \mu_S(x, \beta_S) = \begin{bmatrix} \mu_{S_1}(x_S, \beta_{S_1}) \\ \mu_{S_2}(x, \beta_{S_2}) \end{bmatrix}.$$

We can now use the following notation, for both (Hypo-I), (Hypo-II) for  $x \in \mathbb{R}^N$ ,  $\theta = (\beta_S, \beta_R, \sigma) \in \Theta$ :

$$\mu(x, \theta) = \begin{bmatrix} \mu_S(x, \beta_S) \\ \mu_R(x, \beta_R) \end{bmatrix}, \quad A_j(x, \theta) = \begin{bmatrix} \mathbf{0}_{N_S}^\top \\ A_{R,j}(x, \sigma) \end{bmatrix}, \quad 1 \leq j \leq d.$$

For a test function  $\varphi(\cdot, \theta) : \mathbb{R}^N \rightarrow \mathbb{R}$ ,  $\theta \in \Theta$ , bounded up to second order derivatives, we introduce differential operators  $\mathcal{L}$  and  $\mathcal{L}_j$ ,  $1 \leq j \leq d$ , so that for  $(x, \theta) \in \mathbb{R}^N \times \Theta$ :

$$\begin{aligned} \mathcal{L}\varphi(x, \theta) := & \sum_{i=1}^N \mu^i(x, \theta) \frac{\partial \varphi}{\partial x_i}(x, \theta) \\ & + \frac{1}{2} \sum_{i_1, i_2=1}^N \sum_{k=1}^d A_k^{i_1}(x, \theta) A_k^{i_2}(x, \theta) \frac{\partial^2 \varphi}{\partial x_{i_1} \partial x_{i_2}}(x, \theta); \end{aligned} \quad (2)$$

$$\mathcal{L}_j \varphi(x, \theta) := \sum_{i=1}^N A_j^i(x, \theta) \frac{\partial \varphi}{\partial x_i}(x, \theta), \quad 1 \leq j \leq d. \quad (3)$$

We denote by  $\mathbb{P}_\theta$  the probability law of  $\{X_t\}_{t \geq 0}$  under  $\theta$ . We write  $\xrightarrow{\mathbb{P}_{\theta^\dagger}}$  and  $\xrightarrow{\mathcal{L}_{\theta^\dagger}}$  to denote convergence in probability and distribution, respectively, under the true parameter  $\theta^\dagger$ . The latter is assumed to be unique and to lie in the interior of  $\Theta$ . We denote by  $\mathcal{S}$  the space of functions  $f : [0, \infty) \times \mathbb{R}^N \times \Theta \rightarrow \mathbb{R}$  so that there are constants  $C, q > 0$  such that  $|f(\Delta, x, \theta)| \leq C(1 + |x|^q)\Delta$ ,  $(\Delta, x, \theta) \in [0, \infty) \times \mathbb{R}^N \times \Theta$ . We denote by  $C_p^K(\mathbb{R}^n \times \Theta; \mathbb{R}^m)$ ,  $n, m, K \in \mathbb{N}$ , the set of functions  $f : \mathbb{R}^n \times \Theta \rightarrow \mathbb{R}^m$  such that  $f$  is continuously differentiable up to order  $K$  w.r.t.  $x \in \mathbb{R}^n$  for all  $\theta \in \Theta$ , and  $f$  and its derivatives up to order  $K$  are of polynomial growth in  $x$  uniformly in  $\theta \in \Theta$ .  $C_b^\infty(\mathbb{R}^n, \mathbb{R}^m)$  is the set of smooth functions such that the function and its derivatives of any order are bounded. For a sufficiently smooth  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , for  $\alpha \in \{1, \dots, n\}^l$ ,  $l \geq 1$  and  $1 \leq i \leq n$ , we write  $\partial_\alpha^u f(u) := \frac{\partial^l}{\partial u_{\alpha_1} \dots \partial u_{\alpha_l}} f(u)$ ,  $\partial_{u,i} f(u) := \frac{\partial}{\partial u_i} f(u)$ . We write  $\partial_u = [\partial_{u,i}, \dots, \partial_{u,n}]^\top$ ,  $\partial_u^2 = \partial_u \partial_u^\top$ , for the standard differential operators acting upon maps  $\mathbb{R}^n \rightarrow \mathbb{R}$ ,  $n \geq 1$ . For a matrix  $A$ , we write its  $(i, j)$ -th element as  $[A]_{ij}$ .

## 2. Model assumptions and contrast estimators

We start by stating some basic assumptions for the classes (Hypo-I), (Hypo-II). In particular, Section 2.1 provides Hörmander-type conditions implying that the SDEs of interest are hypo-elliptic, thus their transition distribution admits a density w.r.t. the Lebesgue measure. The stated conditions also highlight that classes (Hypo-I), (Hypo-II) are distinct. We proceed to develop our estimators in Section 2.2. We often denote by  $X_t^{(I)}$ ,  $X_t^{(II)}$  the solutions to SDEs (Hypo-I), (Hypo-II), respectively.

### 2.1. Conditions for hypo-ellipticity

We first introduce some notation. We write the drift function of the Stratonovich-type SDE corresponding to the Itô-type one, given in (Hypo-I) or (Hypo-II), as

$A_0(x, \theta) := \mu(x, \theta) - \frac{1}{2} \sum_{k=1}^d \mathcal{L}_k A_k(x, \theta)$ ,  $(x, \theta) \in \mathbb{R}^N \times \Theta$ . We often treat vector fields  $V : \mathbb{R}^N \rightarrow \mathbb{R}^N$  as differential operators via the relation  $V \leftrightarrow \sum_{i=1}^N V^i \partial_{x,i}$ . For two vector fields  $V, W : \mathbb{R}^N \rightarrow \mathbb{R}^N$ , their Lie bracket is defined as  $[V, W] = VW - WV$ . That is, for the SDE vector fields  $A_k, A_l$ ,  $0 \leq k, l \leq d$ , we have that:

$$\begin{aligned} & [A_k, A_l](x, \theta) \\ &= \sum_{i=1}^N A_k^i(x, \theta) \partial_{x,i} A_l(x, \theta) - \sum_{i=1}^N A_l^i(x, \theta) \partial_{x,i} A_k(x, \theta), \quad (x, \theta) \in \mathbb{R}^N \times \Theta. \end{aligned}$$

For  $1 \leq i \leq j \leq N$ , we define the projection operator  $\text{proj}_{i,j} : \mathbb{R}^N \rightarrow \mathbb{R}^{j-i+1}$  as  $x = [x_1, \dots, x_N]^\top \mapsto \text{proj}_{i,j}(x) = [x_i, \dots, x_j]^\top$ . We impose the following conditions on classes (Hypo-I) and (Hypo-II).

H1 (i). For class (Hypo-I), it holds that, for any  $(x, \theta) \in \mathbb{R}^N \times \Theta$ :

$$\begin{aligned} & \text{span}\{A_{R,k}(x, \sigma), 1 \leq k \leq d\} = \mathbb{R}^{N_R}; \\ & \text{span}\{\{A_k(x, \sigma), [A_0, A_k](x, \theta)\}, 1 \leq k \leq d\} = \mathbb{R}^N. \end{aligned} \quad (4)$$

(ii). For class (Hypo-II), it holds that, for any  $(x, \theta) \in \mathbb{R}^N \times \Theta$ :

$$\begin{aligned} & \text{span}\{A_{R,k}(x, \sigma), 1 \leq k \leq d\} = \mathbb{R}^{N_R}; \\ & \text{span}\{\text{proj}_{N_{S_1}+1, N}\{A_k(x, \sigma)\}, \text{proj}_{N_{S_1}+1, N}\{[A_0, A_k](x, \theta)\}, 1 \leq k \leq d\} \\ & \quad = \mathbb{R}^{N_{S_2}+N_R}; \\ & \text{span}\{\{A_k(x, \sigma), [A_0, A_k](x, \theta), [A_0, [A_0, A_k]](x, \theta)\}, 1 \leq k \leq d\} \\ & \quad = \mathbb{R}^N. \end{aligned} \quad (5)$$

H2(K) (i). For class (Hypo-I):

$$\mu_S, A_j \in C_p^{K+2}(\mathbb{R}^N \times \Theta; \mathbb{R}^{N_S}), \quad \mu_R \in C_p^K(\mathbb{R}^N \times \Theta; \mathbb{R}^{N_R}), \quad 1 \leq j \leq d.$$

(ii). For class (Hypo-II):

$$\begin{aligned} & \mu_{S_1} \in C_p^{K+4}(\mathbb{R}^N \times \Theta; \mathbb{R}^{N_{S_1}}), \quad \mu_{S_2}, A_j \in C_p^{K+2}(\mathbb{R}^N \times \Theta; \mathbb{R}^{N_{S_2}}); \\ & \mu_R \in C_p^K(\mathbb{R}^N \times \Theta; \mathbb{R}^{N_R}), \quad 1 \leq j \leq d. \end{aligned}$$

**Remark 2.1.** We use H1 to introduce some structure upon the degenerate SDE models (Hypo-I) and (Hypo-II) and so that contrast functions developed later on are well-defined. H1 is stronger than Hörmander's condition. The latter states that there is  $M \in \mathbb{N}$  so that for any  $(x, \theta) \in \mathbb{R}^N \times \Theta$ :

$$\text{span}\{V(x, \theta) : V(x, \theta) \in \bigcup_{1 \leq m \leq M} \mathbf{V}_m\} = \mathbb{R}^N,$$

where we have set  $\mathbf{V}_0 = \{A_1, \dots, A_d\}$ ,  $\mathbf{V}_k = \{[A_l, A] : A \in \mathbf{V}_{k-1}, 0 \leq l \leq d\}$ ,  $k \geq 1$ .

If  $\mu(\cdot, \theta), A_j(\cdot, \theta) \in C_b^\infty(\mathbb{R}^N; \mathbb{R}^N)$  then Hörmander's condition implies that there exists a smooth Lebesgue density for the law of  $X_t$ ,  $t > 0$ , for any initial condition  $x \in \mathbb{R}^N$  (see e.g. [20]). Thus, (Hypo-I), (Hypo-II) belong to the class of hypo-elliptic SDEs. We note that H1 is satisfied for most hypo-elliptic SDEs used in applications, as e.g. is the case for the models cited in Section 1. Condition H2 allows the SDE coefficients to lie in the larger class  $C_p^K(\mathbb{R}^N; \mathbb{R}^N)$  for some  $K \in \mathbb{N}$  rather than  $C_b^\infty(\mathbb{R}^N; \mathbb{R}^N)$ .  $K$  is determined so that the contrast functions proposed later are well-defined and error terms obtained from Itô-Taylor expansion in constructing the contrast functions are appropriately controlled. We also note that the drift functions in the smooth components, i.e.  $\mu_S$  in (Hypo-I) and  $\mu_{S_1}, \mu_{S_2}$  in (Hypo-II), require more regularity compared to  $\mu_R$ . This is because our construction of the contrast functions uses higher order stochastic Taylor expansions for those functions. More details are given in Sections 2.2.2, 2.2.3.

**Example.** We provide some examples of SDEs satisfying the above conditions in applications.

- i. Underdamped (standard) Langevin equation for one-dimensional particle with a unit mass:

$$\begin{aligned} dq_t &= p_t dt; \\ dp_t &= (-U'(q_t) + \gamma p_t) dt + \sigma dW_t, \end{aligned} \quad (6)$$

where  $\theta = (\gamma, \sigma)$  is the parameter vector and  $U : \mathbb{R} \rightarrow \mathbb{R}$  is some smooth potential with polynomial growth. The values of  $q_t$  and  $p_t$  represent the position and momentum of the particle, respectively. This model belongs in class (Hypo-I) and indeed satisfies condition H1(i).

- ii. Quasi-Markovian Generalised Langevin Equation ( $q$ -GLE) for the case of an one-dimensional particle with an one-dimensional auxiliary variable:

$$\begin{aligned} dq_t &= p_t dt; \\ dp_t &= (-U'(q_t) + \lambda s_t) dt; \\ ds_t &= (-\lambda p_t - \alpha s_t) dt + \sigma dW_t, \end{aligned} \quad (7)$$

where  $\theta = (\lambda, \alpha, \sigma)$  is the parameter vector and  $U$  is as in (6). Note that component  $s_t$  is now introduced as an auxiliary variable to capture the non-Markovianity of the memory kernel. The  $q$ -GLE class has been recently actively studied as an effective model in physics (see e.g. [17]) and parameter estimation of the model also has been investigated in [15]. The drift function of the smooth component  $q_t$  is now independent of the rough component  $s_t$ . Model (7) belongs in the SDE class (Hypo-II) and satisfies condition H1(ii). Details can be found, e.g., in [13].

## 2.2. Contrast estimators for degenerate SDEs

Under H1-H2, we define contrast functions for the classes of hypo-elliptic SDEs (Hypo-I) and (Hypo-II), so that the corresponding parameter estimators attain



a CLT under the design condition  $\Delta_n = o(n^{-1/p})$ ,  $p \geq 2$ . The development of our contrast functions is related to the approaches of [16, 28] where, in an elliptic setting, contrast functions delivering CLTs with  $\Delta_n = o(n^{-1/p})$  are obtained. However, carrying forward such earlier approaches to the hypo-elliptic diffusion is far from straightforward, mainly due to the degeneracy of the diffusion matrix. After a brief review of the construction of contrast functions that deliver a CLT with  $\Delta_n = o(n^{-1/p})$  in the elliptic case, we proceed to the treatment of the hypo-elliptic class of models. Hereafter, we make use of superscripts (I) and (II) as necessary to specify the class of hypo-elliptic SDEs we work with.

### 2.2.1. Review of contrast estimators for elliptic SDEs

We review the construction of contrast estimators for elliptic SDEs in [16, 28], where the diffusion matrix  $a(x, \sigma) = A(x, \sigma)A(x, \sigma)^\top$  for the SDE in (1) is now assumed to be positive definite for any  $(x, \sigma) \in \mathbb{R}^N \times \Theta_\sigma$ .

**Remark 2.2.** We introduce the following notation. For a vector  $y \in \mathbb{R}^N$  and a symmetric matrix  $\Sigma \in \mathbb{R}^{N \times N}$ ,  $N \in \mathbb{N}$ , we define  $\mathbf{q}(y; \Sigma) = y^\top \Sigma y$ .

**Step 1.** Via an Itô-Taylor expansion, one obtains high-order approximations for the mean and variance of  $X_{t_i}$  given  $X_{t_{i-1}}$ , that is:

$$\begin{aligned}\mathbb{E}_\theta[X_{t_i} | \mathcal{F}_{t_{i-1}}] &= r_{K_p}(\Delta, X_{t_{i-1}}, \theta) + \mathcal{O}(\Delta^{K_p+1}); \\ \text{Var}_\theta[X_{t_i} | \mathcal{F}_{t_{i-1}}] &= \Delta \cdot \Xi_{K_p}(\Delta, X_{t_{i-1}}, \theta) + \mathcal{O}(\Delta^{K_p+1}),\end{aligned}$$

where  $K_p = [p/2]$ , with  $r_{K_p}(\Delta, X_{t_{i-1}}, \theta) \in \mathbb{R}^N$  and  $\Xi_{K_p}(\Delta, X_{t_{i-1}}, \theta) \in \mathbb{R}^{N \times N}$  determined as follows:

$$\begin{aligned}r_K(\Delta, X_{t_{i-1}}, \theta) &= X_{t_{i-1}} + \sum_{1 \leq k \leq K_p} \Delta^k \cdot \mathcal{L}^{k-1} \mu(X_{t_{i-1}}, \theta); \\ \Xi_{K_p}(\Delta, X_{t_{i-1}}, \theta) &= \sum_{0 \leq k \leq K_p} \Delta^k \cdot \Sigma_k(X_{t_{i-1}}, \theta).\end{aligned}$$

In the above expression for  $\Xi_{K_p}$ , we have  $\Sigma_0 = a = AA^\top$ , and  $\Sigma_k$ ,  $k \geq 1$ , are matrices available in closed-form and which include high-order derivatives of the SDE coefficients.

**Step 2.** [16, 28] make use of a Gaussian density with mean  $r_{K_p}(\Delta_n, X_{t_{i-1}}, \theta)$  and variance  $\Delta_n \cdot \Xi_{K_p}(\Delta_n, X_{t_{i-1}}, \theta)$  as a proxy for the intractable transition density of  $X_{t_i}$  given  $X_{t_{i-1}}$ . Furthermore, to ensure that the approximation is well-defined (notice that  $\Xi_{K_p}$  is not guaranteed to be positive-definite) and avoid cumbersome technicalities, they apply a formal Taylor expansion on  $\Xi_{K_p}^{-1}(\Delta_n, X_{t_{i-1}}, \theta)$  and  $\log \det \Xi_{K_p}(\Delta_n, X_{t_{i-1}}, \theta)$  around  $\Delta_n = 0$  so that the positive definiteness of the matrix  $a(x, \sigma)$  is exploited. Thus, they define the estimator as

$\hat{\theta}_{p,n}^{\text{Elliptic}} = \arg \min_{\theta \in \Theta} \ell_{p,n}^{\text{Elliptic}}(\theta)$ ,  $p \geq 2$ , for the contrast function:

$$\ell_{p,n}^{\text{Elliptic}}(\theta) = \sum_{i=1}^n \sum_{k=0}^{K_p} \Delta_n^k \cdot \left\{ \frac{1}{\Delta_n} \mathbf{q}(X_{t_i} - r_{K_p}(\Delta_n, X_{t_{i-1}}, \theta); \mathbf{G}_k(X_{t_{i-1}}, \theta)) \right. \\ \left. + \mathbf{H}_k(X_{t_{i-1}}, \theta) \right\}, \quad (8)$$

where  $\mathbf{G}_k : \mathbb{R}^N \times \Theta \rightarrow \mathbb{R}^{N \times N}$  and  $\mathbf{H}_k : \mathbb{R}^N \times \Theta \rightarrow \mathbb{R}$  are analytically available and correspond to the coefficients of the  $\Delta^k$ -term in the formal Taylor expansion of  $\Xi_{K_p}^{-1}(\Delta, X_{t_{i-1}}, \theta)$  and of  $\log \det \Xi_{K_p}(\Delta, X_{t_{i-1}}, \theta)$  at  $\Delta = 0$ , respectively.

**Remark 2.3.** Due to the fact that  $K_p = [p/2]$ ,  $p \geq 2$ , the contrast function/estimator has the same form for  $p = 2k, 2k+1$ ,  $k \in \mathbb{N}$ . However, as [16, 28] noted in their works, when  $p = 2$ , a simpler estimator based upon the Euler-Maruyama discretisation is asymptotically normal under the condition  $\Delta_n = o(n^{-1/2})$ , with a contrast function given as:

$$\ell_n^{\text{EM}}(\theta) = \sum_{i=1}^n \left\{ \frac{1}{\Delta_n} \cdot \mathbf{q}(X_{t_i} - X_{t_{i-1}} - \Delta_n \mu(X_{t_{i-1}}, \beta); a^{-1}(X_{t_{i-1}}, \sigma)) \right. \\ \left. + \log \det a(X_{t_{i-1}}, \sigma) \right\},$$

### 2.2.2. Contrast estimators for SDE class (Hypo-I)

We will adapt the strategy of [16, 28] to construct contrast functions for degenerate SDEs, starting from the hypo-elliptic class (Hypo-I). However, the development is not straightforward because now the diffusion matrix  $a(x, \sigma)$  is not positive definite. Another important difference between the hypo-elliptic and the elliptic setting is that an Itô-Taylor expansion for moments of the SDE can involve  $\Delta$  with varying orders across smooth and rough components.

**Example.** We consider the under-damped Langevin equation defined in (6). For the model, the Itô-Taylor expansion gives:

$$\text{Var}_{\theta}[X_{t_{i+1}} | \mathcal{F}_{t_i}] = \sigma^2 \begin{bmatrix} \frac{\Delta^3}{3} & \frac{\Delta^2}{2} \\ \frac{\Delta^2}{2} & \Delta \end{bmatrix} + \begin{bmatrix} \mathcal{O}(\Delta^4) & \mathcal{O}(\Delta^3) \\ \mathcal{O}(\Delta^3) & \mathcal{O}(\Delta^2) \end{bmatrix}, \quad \Delta = t_{i+1} - t_i.$$

Thus, the order of  $\Delta$  is larger for the smooth component. Note that the leading term  $\sigma^2 \Delta^3/3$  in the variance of  $X_{\Delta}^1$  derives from the Gaussian variate  $\sigma \int_0^{\Delta} W_s ds$  arising in the Itô-Taylor expansion of  $X_{\Delta}^1$ .

One must now find a positive definite matrix instead of  $a(x, \theta)$  to obtain a general contrast function in the form of (8) in a hypo-elliptic setting while dealing with such structure of varying scales amongst the components of degenerate SDEs. We thus begin by considering a standardisation of  $X_{\Delta}^{(1)}$  (conditionally on

$X_0^{(I)} = x$ ) via subtracting high-order mean approximations from smooth/rough components and dividing with appropriate  $\Delta$ -terms. In particular, we introduce the  $\mathbb{R}^N$ -valued random variables as:

$$Y_{p,\Delta}^{(I)} := m_p^{(I)}(\Delta, x, X_\Delta^{(I)}, \theta) = \begin{bmatrix} \frac{1}{\sqrt{\Delta^3}} \left( X_{S,\Delta}^{(I)} - r_{S,K_p+1}^{(I)}(\Delta, x; \theta) \right) \\ \frac{1}{\sqrt{\Delta}} \left( X_{R,\Delta}^{(I)} - r_{R,K_p}^{(I)}(\Delta, x; \theta) \right) \end{bmatrix},$$

where we have set  $K_p = [p/2]$  and for  $q \in \mathbb{N}$ ,

$$\begin{bmatrix} r_{S,q}^{(I)}(\Delta, x, \theta) \\ r_{R,q}^{(I)}(\Delta, x, \theta) \end{bmatrix} = \begin{bmatrix} x_S \\ x_R \end{bmatrix} + \sum_{k=1}^q \frac{\Delta^k}{k!} \begin{bmatrix} \mathcal{L}^{k-1} \mu_S(x, \theta) \\ \mathcal{L}^{k-1} \mu_R(x, \theta) \end{bmatrix}, \quad x = \begin{bmatrix} x_S \\ x_R \end{bmatrix} \in \mathbb{R}^N \quad (9)$$

with this latter quantity obtained from an Itô-Taylor expansion of  $\mathbb{E}_\theta[X_\Delta^{(I)}]$ . We note that  $Y_{p,\Delta}$  is well-defined under [H2](#)( $2K_p$ ). Also, note that different orders (by one) of mean approximation are used for smooth and rough components in the above standardisation. We will provide some more details on this in [Remark 2.6](#) later in the paper. We denote the Lebesgue density of the distribution of  $X_\Delta^{(I)}$  given  $X_0^{(I)} = x$  and  $\theta \in \Theta$  as  $y \mapsto \mathbb{P}_\theta(X_\Delta^{(I)} \in dy)/dy = p^{X_\Delta^{(I)}}(x, y; \theta)$ . Transformation of random variables gives that, for  $(x, y, \theta) \in \mathbb{R}^N \times \mathbb{R}^N \times \Theta$ :

$$p^{X_\Delta^{(I)}}(x, y; \theta) = \frac{1}{\sqrt{\Delta^{3N_S+N_R}}} p^{Y_{p,\Delta}^{(I)}}(\xi; \theta) \Big|_{\xi=m_p^{(I)}(\Delta, x, y, \theta)},$$

where  $p^{Y_{p,\Delta}^{(I)}}(\xi; \theta) \equiv \mathbb{P}_\theta(Y_{p,\Delta}^{(I)} \in d\xi)/d\xi$  denotes the Lebesgue density of the law of  $Y_{p,\Delta}^{(I)}$  given the parameter  $\theta \in \Theta$ . Following the standardisation, an Itô-Taylor expansion now gives, under [H2](#)( $2J$ ),  $J \in \mathbb{N}$ :

$$\mathbb{E}_\theta \left[ Y_{p,\Delta}^{(I)} (Y_{p,\Delta}^{(I)})^\top \right] = \Sigma^{(I)}(x, \theta) + \sum_{j=1}^J \Delta^j \cdot \Sigma_j^{(I)}(x, \theta) + R^{(I)}(\Delta^{J+1}, x, \theta), \quad (10)$$

where  $R^{(I)} \in \mathcal{S}$ , for some analytically available  $N \times N$  matrices  $\Sigma_j^{(I)}(x, \theta)$ . In particular, matrix  $\Sigma^{(I)}(x, \theta)$  has the following block expression:

$$\Sigma^{(I)}(x, \theta) \equiv \begin{bmatrix} \Sigma_{SS}^{(I)}(x, \theta) & \Sigma_{SR}^{(I)}(x, \theta) \\ \Sigma_{RS}^{(I)}(x, \theta) & \Sigma_{RR}^{(I)}(x, \theta) \end{bmatrix},$$

where we have set:

$$\begin{aligned} \Sigma_{RR}^{(I)}(x, \theta) &= \sum_{k=1}^d A_{R,k}(x, \sigma) A_{R,k}(x, \sigma)^\top \equiv a_R(x, \sigma); \\ \Sigma_{SR}^{(I)}(x, \theta) &= \frac{1}{2} \sum_{k=1}^d \mathcal{L}_k \mu_S(x, \theta) A_{R,k}(x, \sigma)^\top, \quad \Sigma_{RS}^{(I)}(x, \theta) = \Sigma_{SR}^{(I)}(x, \theta)^\top; \end{aligned}$$

$$\Sigma_{SS}^{(I)}(x, \theta) = \frac{1}{3} \sum_{k=1}^d \mathcal{L}_k \mu_S(x, \theta) \mathcal{L}_k \mu_S(x, \theta)^\top \equiv \frac{1}{3} a_S(x, \theta).$$

Matrix  $\Sigma^{(I)}(x, \theta)$  plays a role similar to  $a(x, \theta)$  in the elliptic setting due to the following result whose proof is in Section A.2 of the Supplementary Material [12].

**Lemma 2.4.** *Under H1(i), the matrices  $a_R(x, \sigma)$ ,  $a_S(x, \theta)$  and  $\Sigma^{(I)}(x, \theta)$  are positive definite for all  $(x, \theta) \in \mathbb{R}^N \times \Theta$ .*

We form a Gaussian approximation for  $p_{p,\Delta}^{Y^{(I)}}(\xi; \theta)$  to obtain a contrast function that will be well-defined due to the positive definiteness of matrix  $\Sigma^{(I)}(x, \theta)$ . We introduce some notation. For  $h > 0$ ,  $\theta \in \Theta$ ,  $0 \leq i \leq n$ ,  $1 \leq j \leq K$  and  $K \in \mathbb{N}$ ,

$$\begin{aligned} \mathbf{m}_{p,i}^{(I)}(h, \theta) &:= m_p^{(I)}(h, X_{t_{i-1}}^{(I)}, X_{t_i}^{(I)}, \theta); \\ \Sigma_i^{(I)}(\theta) &:= \Sigma^{(I)}(X_{t_i}, \theta), \quad \Sigma_{i,j}^{(I)}(\theta) := \Sigma_j^{(I)}(X_{t_i}, \theta); \\ \Xi_{K,i}^{(I)}(h, \theta) &:= \Sigma_i^{(I)}(\theta) + \sum_{1 \leq j \leq K} h^j \cdot \Sigma_{i,j}^{(I)}(\theta), \quad \Lambda_i^{(I)}(\theta) := (\Sigma_i^{(I)}(\theta))^{-1}. \end{aligned}$$

Note that the above functions are well-defined under H2( $2K_p$ ); see (9)–(10). We write the formal Taylor expansion of  $(\Xi_{K,i}^{(I)}(h, \theta))^{-1}$  and  $\log \det \Xi_{K,i}^{(I)}(h, \theta)$  up to the level  $K \in \mathbb{N}$  as  $\sum_{0 \leq k \leq K} h^k \cdot \mathbf{G}_{i,k}^{(I)}(\theta)$ ,  $\sum_{0 \leq k \leq K} h^k \cdot \mathbf{H}_{i,k}^{(I)}(\theta)$ , respectively, where

$$\begin{aligned} \mathbf{G}_{i,k}^{(I)}(\theta) &= \frac{1}{k!} \partial_h^k (\Xi_{K,i}^{(I)}(h, \theta))^{-1} \big|_{h=0}; \\ \mathbf{H}_{i,k}^{(I)}(\theta) &= \frac{1}{k!} \partial_h^k (\log \det \Xi_{K,i}^{(I)}(h, \theta)) \big|_{h=0}, \quad 0 \leq k \leq K. \end{aligned}$$

For instance, one has:

$$\begin{aligned} \mathbf{G}_{i,0}^{(I)}(\theta) &= \Lambda_i^{(I)}(\theta), \quad \mathbf{G}_{i,1}^{(I)}(\theta) = -\Lambda_i^{(I)}(\theta) \Sigma_{i,1}^{(I)}(\theta) \Lambda_i^{(I)}(\theta); \\ \mathbf{G}_{i,2}^{(I)}(\theta) &= -(\mathbf{G}_{i,1}^{(I)}(\theta) \Sigma_{i,1}^{(I)}(\theta) + \Lambda_i^{(I)}(\theta) \Sigma_{i,2}^{(I)}(\theta)) \Lambda_i^{(I)}(\theta) \end{aligned}$$

and

$$\begin{aligned} \mathbf{H}_{i,0}^{(I)}(\theta) &= \log \det \Sigma_i^{(I)}(\theta), \quad \mathbf{H}_{i,1}^{(I)}(\theta) = \text{Tr} \left[ \Lambda_i^{(I)}(\theta) \Sigma_{i,1}^{(I)}(\theta) \right]; \\ \mathbf{H}_{i,2}^{(I)}(\theta) &= \text{Tr} \left[ \frac{1}{2} \mathbf{G}_{i,1}^{(I)}(\theta) \Sigma_{i,1}^{(I)}(\theta) + \Lambda_i^{(I)}(\theta) \Sigma_{i,2}^{(I)}(\theta) \right]. \end{aligned}$$

Note that the terms  $\mathbf{G}_{i,k}^{(I)}(\theta)$  and  $\mathbf{H}_{i,k}^{(I)}(\theta)$  involve the inverse of the matrix  $\Sigma_i^{(I)}(\theta)$ , and are well-defined following Condition H1(i) and Lemma 2.4. We now construct the contrast function  $\ell_{p,n,\Delta}^{(I)}(\theta)$ ,  $\theta = (\beta_S, \beta_R, \sigma) \in \Theta$ , for the hypo-elliptic class (Hypo-I) as follows. Under Condition H2( $2K_p$ ):

(i) For  $p = 2$ ,

$$\ell_{2,n}^{(I)}(\theta) = \sum_{i=1}^n \left\{ \mathbf{m}_{2,i}^{(I)}(\Delta_n, \theta)^\top \mathbf{\Lambda}_{i-1}^{(I)}(\theta) \mathbf{m}_{2,i}^{(I)}(\Delta_n, \theta) + \log \det \mathbf{\Sigma}_{i-1}^{(I)}(\theta) \right\}. \quad (11)$$

(ii) For  $p \geq 3$ ,

$$\ell_{p,n}^{(I)}(\theta) = \sum_{i=1}^n \sum_{j=0}^{K_p} \Delta_n^j \cdot \left\{ \mathbf{m}_{p,i}^{(I)}(\Delta_n, \theta)^\top \mathbf{G}_{i-1,j}^{(I)}(\theta) \mathbf{m}_{p,i}^{(I)}(\Delta_n, \theta) + \mathbf{H}_{i-1,j}^{(I)}(\theta) \right\}. \quad (12)$$

Thus, the contrast estimator  $\hat{\theta}_{p,n}^{(I)} = (\hat{\beta}_{S,p,n}, \hat{\beta}_{R,p,n}, \hat{\sigma}_{p,n})$ ,  $p \geq 2$ , is defined as

$$\hat{\theta}_{p,n}^{(I)} = \arg \min_{\theta \in \Theta} \ell_{p,n}^{(I)}(\theta).$$

### 2.2.3. Contrast estimators for SDE class (*Hypo-II*)

To construct the contrast function and corresponding estimator for the highly degenerate diffusion class defined via (*Hypo-II*), we proceed as above, that is we consider an appropriate standardisation that takes under consideration the scales of the three components  $(X_{S_1,\Delta}, X_{S_2,\Delta}, X_{R,\Delta})$ ,  $\Delta > 0$ . Note that the first smooth component  $X_{S_1,\Delta}$  has a variance of size  $\mathcal{O}(\Delta^5)$  due to the largest (as  $\Delta \rightarrow 0$ ) variate in the Itô-Taylor expansion being  $\int_0^\Delta \int_0^s W_u du ds$ , where we have made use of the fact that the drift  $\mu_{S_1}$  is a function only of the smooth components. Thus, under Condition [H2](#) ( $2K_p$ ), we introduce:

$$\mathbf{m}_{p,i}^{(II)}(\Delta, \theta) := \begin{bmatrix} \frac{1}{\sqrt{\Delta_n^5}} (X_{S_1,t_i}^{(II)} - r_{S_1,K_p+2}^{(II)}(\Delta, X_{t_{i-1}}^{(II)}, \theta)) \\ \frac{1}{\sqrt{\Delta_n^3}} (X_{S_2,t_i}^{(II)} - r_{S_2,K_p+1}^{(II)}(\Delta, X_{t_{i-1}}^{(II)}, \theta)) \\ \frac{1}{\sqrt{\Delta_n}} (X_{R,t_i}^{(II)} - r_{R,K_p}^{(II)}(\Delta, X_{t_{i-1}}^{(II)}, \theta)) \end{bmatrix},$$

where we have set, for  $q \in \mathbb{N}$ ,  $(\Delta, x, \theta) \in (0, \infty) \times \mathbb{R}^N \times \Theta$ :

$$\begin{bmatrix} r_{S_1,q}^{(II)}(\Delta, x, \theta) \\ r_{S_2,q}^{(II)}(\Delta, x, \theta) \\ r_{R,q}^{(II)}(\Delta, x, \theta) \end{bmatrix} = \begin{bmatrix} x_{S_1} \\ x_{S_2} \\ x_R \end{bmatrix} + \sum_{k=1}^q \frac{\Delta^k}{k!} \begin{bmatrix} \mathcal{L}^{k-1} \mu_{S_1}(x, \theta) \\ \mathcal{L}^{k-1} \mu_{S_2}(x, \theta) \\ \mathcal{L}^{k-1} \mu_R(x, \theta) \end{bmatrix}.$$

We will explain the choice of the truncation levels used above in the mean approximation  $r^{(II)}$  in Remark [2.6](#). Via an Itô-Taylor expansion, we obtain under Condition [H2](#) ( $2K_p$ ) that:

$$\begin{aligned} \mathbb{E}_{\theta^\dagger} [\mathbf{m}_{p,i+1}^{(II)}(\Delta, \theta)^\top \mathbf{m}_{p,i+1}^{(II)}(\Delta, \theta) | \mathcal{F}_{t_i}] \\ = \mathbf{\Sigma}_i^{(II)}(\theta) + \sum_{j=1}^{K_p} \Delta^j \cdot \mathbf{\Sigma}_{i,j}^{(II)}(\theta) + R(\Delta^{K_p+1}, X_{t_i}^{(II)}, \theta), \end{aligned} \quad (13)$$

for some analytically available matrices  $\Sigma_i^{(\text{II})}(\theta)$ ,  $\Sigma_{j,i}^{(\text{II})}(\theta) \in \mathbb{R}^{N \times N}$ ,  $1 \leq i \leq n$ ,  $j \in \mathbb{N}$ , and a residual  $R(\cdot)$ , such that  $[R(\cdot)]_{k_1 k_2} \in \mathcal{S}$ ,  $1 \leq k_1, k_2 \leq N$ . In particular, the matrix  $\Sigma_i^{(\text{II})}(\theta) \equiv \Sigma^{(\text{II})}(X_{t_i}^{(\text{II})}, \theta)$  admits the following block expression:

$$\Sigma^{(\text{II})}(x, \theta) = \begin{bmatrix} \Sigma_{S_1 S_1}^{(\text{II})}(x, \theta) & \Sigma_{S_1 S_2}^{(\text{II})}(x, \theta) & \Sigma_{S_1 R}^{(\text{II})}(x, \theta) \\ \Sigma_{S_2 S_1}^{(\text{II})}(x, \theta) & \Sigma_{S_2 S_2}^{(\text{II})}(x, \theta) & \Sigma_{S_2 R}^{(\text{II})}(x, \theta) \\ \Sigma_{R S_1}^{(\text{II})}(x, \theta) & \Sigma_{R S_2}^{(\text{II})}(x, \theta) & \Sigma_{R R}^{(\text{II})}(x, \theta) \end{bmatrix} \quad (14)$$

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

$$\begin{aligned} \Sigma_{RR}^{(\text{II})}(x, \theta) &= \sum_{k=1}^d A_{R,k}(x, \sigma) A_{R,k}(x, \sigma)^\top \equiv a_R(x, \sigma); \\ \Sigma_{S_2 R}^{(\text{II})}(x, \theta) &= \frac{1}{2} \sum_{k=1}^d \mathcal{L}_k \mu_{S_2}(x, \theta) A_{R,k}(x, \sigma)^\top, \quad \Sigma_{RS_2}^{(\text{II})}(x, \theta) = \Sigma_{S_2 R}^{(\text{II})}(x, \theta)^\top; \\ \Sigma_{S_1 R}^{(\text{II})}(x, \theta) &= \frac{1}{6} \sum_{k=1}^d \mathcal{L}_k \mathcal{L} \mu_{S_1}(x, \theta) A_{R,k}(x, \sigma)^\top, \quad \Sigma_{RS_1}^{(\text{II})}(x, \theta) = \Sigma_{S_1 R}^{(\text{II})}(x, \theta)^\top; \\ \Sigma_{S_2 S_2}^{(\text{II})}(x, \theta) &= \frac{1}{3} \sum_{k=1}^d \mathcal{L}_k \mu_{S_2}(x, \theta) \mathcal{L}_k \mu_{S_2}(x, \theta)^\top \equiv \frac{1}{3} a_{S_2}(x, \theta), \\ \Sigma_{S_1 S_2}^{(\text{II})}(x, \theta) &= \frac{1}{8} \sum_{k=1}^d \mathcal{L}_k \mathcal{L} \mu_{S_1}(x, \theta) \mathcal{L}_k \mu_{S_2}(x, \theta)^\top, \quad \Sigma_{S_2 S_1}^{(\text{II})}(x, \theta) = \Sigma_{S_1 S_2}^{(\text{II})}(x, \theta)^\top, \\ \Sigma_{S_1 S_1}^{(\text{II})}(x, \theta) &= \frac{1}{20} \sum_{k=1}^d \mathcal{L}_k \mathcal{L} \mu_{S_1}(x, \theta) \mathcal{L}_k \mathcal{L} \mu_{S_1}(x, \theta)^\top \equiv \frac{1}{20} a_{S_1}(x, \theta). \end{aligned}$$

As is the case with the matrix  $\Sigma^{(\text{II})}(x, \theta)$ ,  $(x, \theta) \in \mathbb{R}^N \times \Theta$ , for the matrix  $\Sigma^{(\text{II})}(x, \theta)$ , we have the following result whose proof is provided in Appendix B of [13].

**Lemma 2.5.** *Under condition H1(ii), the matrices  $a_R(x, \theta)$ ,  $a_{S_2}(x, \theta)$ ,  $a_{S_1}(x, \theta)$  and  $\Sigma^{(\text{II})}(x, \theta)$  are positive definite for all  $(x, \theta) \in \mathbb{R}^N \times \Theta$ .*

For  $h > 0$ ,  $0 \leq i \leq n$  and  $0 \leq k \leq K_p = [p/2]$ , we write:

$$\begin{aligned} \Lambda_i^{(\text{II})}(\theta) &= (\Sigma_i^{(\text{II})}(\theta))^{-1}, \quad \Xi_{K_p, i}^{(\text{II})}(h, \theta) = \Sigma_i^{(\text{II})}(\theta) + \sum_{j=1}^{K_p} h^j \cdot \Sigma_{i, j}^{(\text{II})}(\theta); \\ \mathbf{G}_{i, k}^{(\text{II})}(\theta) &= \frac{1}{k!} \partial_h^k (\Xi_{K_p, i}^{(\text{II})}(h, \theta))^{-1} \Big|_{h=0}, \quad \mathbf{H}_{i, k}^{(\text{II})}(t, \theta) = \frac{1}{k!} \partial_h^k (\log \det \Xi_{K_p, i}^{(\text{II})}(h, \theta)) \Big|_{h=0}. \end{aligned}$$

We now obtain our contrast function  $\ell_{p, n}^{(\text{II})}(\theta)$ ,  $p \geq 2$ , for the hypo-elliptic class (Hypo-II). Under Condition H2 ( $2K_p$ ):

(i) For  $p = 2$ :

$$\ell_{2, n}^{(\text{II})}(\theta) = \sum_{i=1}^n \left\{ \mathbf{m}_{2, i}^{(\text{II})}(\Delta_n, \theta)^\top \Lambda_{i-1}^{(\text{II})}(\theta) \mathbf{m}_{2, i}^{(\text{II})}(\Delta_n, \theta) + \log \det \Sigma_{i-1}^{(\text{II})}(\theta) \right\}.$$

(ii) For  $p \geq 3$ :

$$\ell_{p,n}^{(\text{II})}(\theta) = \sum_{i=1}^n \sum_{j=0}^{K_p} \Delta_n^j \cdot \{ \mathbf{m}_{p,i}^{(\text{II})}(\Delta_n, \theta)^\top \mathbf{G}_{i-1,j}^{(\text{II})}(\theta) \mathbf{m}_{p,i}^{(\text{II})}(\Delta_n, \theta) + \mathbf{H}_{i-1,j}^{(\text{II})}(\theta) \}.$$

Thus, the contrast estimator is defined as:

$$\hat{\theta}_{p,n}^{(\text{II})} = (\hat{\beta}_{S_1,p,n}^{(\text{II})}, \hat{\beta}_{S_2,p,n}^{(\text{II})}, \hat{\beta}_{R,p,n}^{(\text{II})}, \hat{\sigma}_{p,n}^{(\text{II})}) = \arg \min_{\theta \in \Theta} \ell_{p,n}^{(\text{II})}(\theta).$$

**Remark 2.6.** In the definition of contrast estimator, we make use of mean approximations with different number of terms for the various components, so that it holds that, for any  $\theta \in \Theta$ ,  $p \geq 2$ ,  $1 \leq i \leq n$ , and  $1 \leq k \leq N$ :

$$\mathbb{E}_\theta [\mathbf{m}_{p,i}^{(w),k}(\Delta_n, \theta) | \mathcal{F}_{t_{i-1}}] = R_k(\sqrt{\Delta_n^{2K_p+1}}, X_{t_{i-1}}^{(w)}, \theta), \quad w \in \{\text{I}, \text{II}\}, \quad (15)$$

with some  $R_k \in \mathcal{S}$ . This is one of the key developments to obtain the CLT under the weaker design condition  $\Delta_n = o(n^{-1/p})$ ,  $p \geq 2$ . Notice that we divide the smooth components by smaller quantities in terms of powers of  $\Delta_n$ , e.g.,  $\sqrt{\Delta_n^5}$  and  $\sqrt{\Delta_n^3}$  for the components  $X_{S_1}^{(\text{II})}$  and  $X_{S_2}^{(\text{II})}$ , respectively, thus we require more accurate mean approximations for those components to obtain (15).

### 3. Asymptotic properties of the contrast estimators

To state our main results, we introduce a set of additional conditions for the two classes (Hypo-I) and (Hypo-II). Recall that we write the true parameter as  $\theta^\dagger = (\beta_S^\dagger, \beta_R^\dagger, \sigma^\dagger)$ , under the interpretation that  $\beta_S^\dagger = (\beta_{S_1}^\dagger, \beta_{S_2}^\dagger)$  for class (Hypo-II), and that  $\theta^\dagger$  is assumed to be unique and to lie in the interior of  $\Theta$ .

H3 For any  $x \in \mathbb{R}^N$  and any multi-index  $\alpha \in \{1, \dots, N\}^l$ ,  $l \geq 0$ , the functions:

$$\theta \mapsto \partial_\alpha^x \mu^i(x, \theta), \quad \theta \mapsto \partial_\alpha^x A_j^i(x, \theta), \quad 1 \leq j \leq d, \quad 1 \leq i \leq N,$$

are three times differentiable. For any multi-index  $\beta \in \{1, \dots, N_\theta\}^l$ ,  $l \in \{1, 2, 3\}$ , the functions:

$$x \mapsto \partial_\beta^\theta \partial_\alpha^x \mu^i(x, \theta), \quad x \mapsto \partial_\beta^\theta \partial_\alpha^x A_j^i(x, \theta), \quad 1 \leq j \leq d, \quad 1 \leq i \leq N,$$

are of polynomial growth, uniformly in  $\theta \in \Theta$ .

H4 The diffusion process  $\{X_t\}_{t \geq 0}$  defined via (Hypo-I) or (Hypo-II) is ergodic under  $\theta = \theta^\dagger$ , with invariant distribution denoted by  $\nu_{\theta^\dagger}$ . Furthermore, all moments of  $\nu_{\theta^\dagger}$  are finite.

H5 It holds that for all  $r \geq 1$ ,  $\sup_{t \geq 0} \mathbb{E}_{\theta^\dagger}[|X_t|^r] < \infty$ .

H6 If it holds:

$$\mu_S(x, \beta_S) = \mu_S(x, \beta_S^\dagger), \quad \mu_R(x, \beta_R) = \mu_R(x, \beta_R^\dagger), \quad A_R(x, \sigma) = A_R(x, \sigma^\dagger),$$

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

We first show that the proposed contrast estimators  $\hat{\theta}_{p,n}^{(I)}$ ,  $\hat{\theta}_{p,n}^{(II)}$  are consistent in the high-frequency, complete observation regime:

**Theorem 3.1** (Consistency). *Let  $p \geq 2$  be an arbitrary integer and  $K_p = [p/2]$ . Assume that conditions [H1](#), [H2](#)( $2K_p$ ), [H3](#), [H4](#), [H5](#) and [H6](#) hold. If  $n \rightarrow \infty$ ,  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$ , then:*

$$\hat{\theta}_{p,n}^{(I)} \xrightarrow{\mathbb{P}_{\theta^\dagger}} \theta^\dagger, \quad \hat{\theta}_{p,n}^{(II)} \xrightarrow{\mathbb{P}_{\theta^\dagger}} \theta^\dagger.$$

Also, the proposed estimators are asymptotically normal under the condition  $\Delta_n = o(n^{-1/p})$ ,  $p \geq 2$ .

**Theorem 3.2** (CLT). *Let  $p \geq 2$  be an arbitrary integer and  $K_p = [p/2]$ . Assume that conditions [H1](#), [H2](#)( $2K_p$ ), [H3](#), [H4](#), [H5](#) and [H6](#) hold. If  $n \rightarrow \infty$ ,  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$ , with  $\Delta_n = o(n^{-1/p})$ , then:*

I. For the hypo-elliptic class ([Hypo-I](#)):

$$\begin{bmatrix} \sqrt{\frac{n}{\Delta_n}}(\hat{\beta}_{S,p,n}^{(I)} - \beta_S^\dagger) \\ \sqrt{n\Delta_n}(\hat{\beta}_{R,p,n}^{(I)} - \beta_R^\dagger) \\ \sqrt{n}(\hat{\sigma}_{R,p,n}^{(I)} - \sigma^\dagger) \end{bmatrix} \xrightarrow{\mathcal{L}_{\theta^\dagger}} \mathcal{N}(\mathbf{0}_{N_\theta}, \Gamma^{(I)}(\theta^\dagger)^{-1}),$$

where  $\Gamma^{(I)}(\theta^\dagger) = \text{Diag}[\Gamma_{\beta_S}^{(I)}(\theta^\dagger), \Gamma_{\beta_R}^{(I)}(\theta^\dagger), \Gamma_\sigma^{(I)}(\theta^\dagger)]$  is the asymptotic precision matrix with the involved block matrices specified as:

$$\begin{aligned} [\Gamma_{\beta_S}^{(I)}(\theta^\dagger)]_{i_1 j_1} &= 12 \int \partial_{\beta_S, i_1} \mu_S(y, \beta_S^\dagger)^\top a_S^{-1}(y, \theta^\dagger) \partial_{\beta_S, j_1} \mu_S(y, \beta_S^\dagger) \nu_{\theta^\dagger}(dy); \\ [\Gamma_{\beta_R}^{(I)}(\theta^\dagger)]_{i_2 j_2} &= \int \partial_{\beta_R, i_2} \mu_R(y, \beta_R^\dagger)^\top a_R^{-1}(y, \sigma^\dagger) \partial_{\beta_R, j_2} \mu_R(y, \beta_R^\dagger) \nu_{\theta^\dagger}(dy); \\ [\Gamma_\sigma^{(I)}(\theta^\dagger)]_{i_3 j_3} &= \frac{1}{2} \int \text{Tr}[\partial_{\sigma, i_3} \Sigma^{(I)}(y, \theta^\dagger) \Lambda^{(I)}(y, \theta^\dagger) \partial_{\sigma, j_3} \Sigma^{(I)}(y, \theta^\dagger) \Lambda^{(I)}(y, \theta^\dagger)] \nu_{\theta^\dagger}(dy), \end{aligned}$$

for  $1 \leq i_1, j_1 \leq N_{\beta_S}$ ,  $1 \leq i_2, j_2 \leq N_{\beta_R}$  and  $1 \leq i, j \leq N_\sigma$ .

II. For the hypo-elliptic class ([Hypo-II](#)):

$$\begin{bmatrix} \sqrt{\frac{n}{\Delta_n^3}}(\hat{\beta}_{S_1,p,n}^{(II)} - \beta_{S_1}^\dagger) \\ \sqrt{\frac{n}{\Delta_n}}(\hat{\beta}_{S_2,p,n}^{(II)} - \beta_{S_2}^\dagger) \\ \sqrt{n\Delta_n}(\hat{\beta}_{R,p,n}^{(II)} - \beta_S^\dagger) \\ \sqrt{n}(\hat{\sigma}_{R,p,n}^{(II)} - \sigma^\dagger) \end{bmatrix} \xrightarrow{\mathcal{L}_{\theta^\dagger}} \mathcal{N}(\mathbf{0}_{N_\theta}, \Gamma^{(II)}(\theta^\dagger)^{-1}),$$



where  $\Gamma^{(\text{II})}(\theta^\dagger) = \text{Diag}[\Gamma_{\beta_{S_1}}^{(\text{II})}(\theta^\dagger), \Gamma_{\beta_{S_2}}^{(\text{II})}(\theta^\dagger), \Gamma_{\beta_R}^{(\text{II})}(\theta^\dagger), \Gamma_\sigma^{(\text{II})}(\theta^\dagger)]$  is the asymptotic precision matrix with the involved block matrices specified as:

$$\begin{aligned} [\Gamma_{\beta_{S_1}}^{(\text{II})}(\theta^\dagger)]_{i_1 j_1} &= 720 \int \partial_{\beta_{S_1}, i_1} \mu_{S_1}(y, \beta_{S_1}^\dagger)^\top a_{S_1}^{-1}(y, \theta^\dagger) \partial_{\beta_{S_1}, j_1} \mu_{S_1}(y, \beta_{S_1}^\dagger) \nu_{\theta^\dagger}(dy); \\ [\Gamma_{\beta_{S_2}}^{(\text{II})}(\theta^\dagger)]_{i_2 j_2} &= 12 \int \partial_{\beta_{S_2}, i_2} \mu_{S_2}(y, \beta_{S_2}^\dagger)^\top a_{S_2}^{-1}(y, \theta^\dagger) \partial_{\beta_{S_2}, j_2} \mu_{S_2}(y, \beta_{S_2}^\dagger) \nu_{\theta^\dagger}(dy); \\ [\Gamma_{\beta_R}^{(\text{II})}(\theta^\dagger)]_{i_3 j_3} &= \int \partial_{\beta_R, i_3} \mu_R(y, \beta_R^\dagger)^\top a_R^{-1}(y, \theta^\dagger) \partial_{\beta_R, j_3} \mu_R(y, \beta_R^\dagger) \nu_{\theta^\dagger}(dy); \\ [\Gamma_\sigma^{(\text{II})}(\theta^\dagger)]_{i_4 j_4} &= \frac{1}{2} \int \text{Tr}[\partial_{\sigma, i_4} \Sigma^{(\text{II})}(y, \theta^\dagger) \Lambda^{(\text{II})}(y, \theta^\dagger) \partial_{\sigma, j_4} \Sigma^{(\text{II})}(y, \theta^\dagger) \Lambda^{(\text{II})}(y, \theta^\dagger)] \nu_{\theta^\dagger}(dy), \end{aligned}$$

for  $1 \leq i_1, j_1 \leq N_{\beta_{S_1}}, 1 \leq i_2, j_2 \leq N_{\beta_{S_2}}, 1 \leq i_3, j_3 \leq N_{\beta_R}$  and  $1 \leq i_4, j_4 \leq N_\sigma$ .

The proofs are given in Section 5.

**Remark 3.3.** The design condition  $\Delta_n = o(n^{-1/p})$ ,  $p \geq 2$ , i.e.  $n\Delta_n^p \rightarrow 0$ , appears in the CLT result (Theorem 3.2). As we explain later in the proof in Section 5.2, the condition is required so that the expectation of the score function, specifically, the gradient of the contrast function  $\ell_{p,n}(\theta^\dagger)$ , tends to 0. This is relevant for the mean of the asymptotic Gaussian distribution to converge to 0. If the given design condition is not satisfied, the distribution of estimators will tend to concentrate on an area that deviates from the true value, thus the estimators will suffer from bias. We observe this issue in the numerical experiments in Section 4, where the standard contrast estimator  $\hat{\theta}_{2,n}$  exhibits the described bias when  $n\Delta_n^2$  is not sufficiently small, while  $\hat{\theta}_{p,n}$  for  $p \geq 3$  avoids such a bias.

## 4. Numerical applications

In the numerical examples shown below, for given choices of  $\Delta_n$  and  $n$ , we observe gradual improvements in the behaviour of the estimates when moving from  $p = 2$  to  $p = 3$  and then to  $p = 4$ . Indeed, discrepancies are stronger in the experiment that contrasts  $p = 2$  with  $p = 4$ .

### 4.1. Quasi-Markovian generalised Langevin equation

We study numerically the properties of the proposed contrast estimator for the  $q$ -GLE model defined via (7) in Example 2.1 and which belongs in class (Hypo-II). We consider the following two choices of potential function  $U$ , leading to a linear/non-linear system of degenerate SDEs:

**Case I. Quadratic potential:**  $\mathbb{R} \ni q \mapsto U(q) = Dq^2/2$  with some parameter  $D > 0$ .

**Case II. Double-well potential:**  $\mathbb{R} \ni q \mapsto U(q) = (q^2 - D)^2/4$ , with some parameter  $D > 0$ .

TABLE 2  
*Designs of experiment.*

	Case I.	Case II.
Design A	$(\Delta_n, T_n) = (0.008, 1000)$	$(\Delta_n, T_n) = (0.01, 1000)$
Design B	$(\Delta_n, T_n) = (0.005, 500)$	$(\Delta_n, T_n) = (0.005, 1000)$

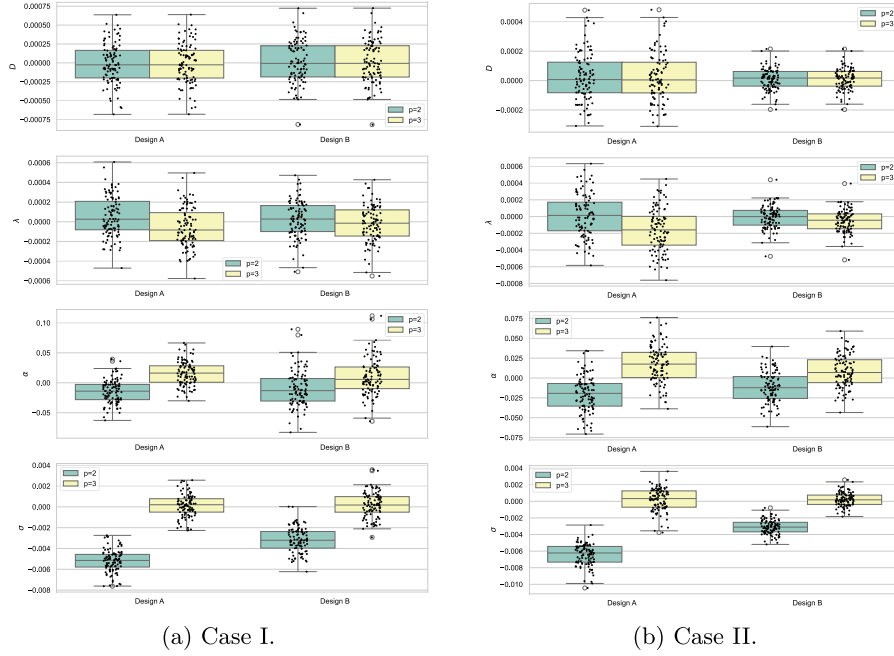


FIG 1. Boxplots of  $M = 100$  independent realisations of  $(\hat{\theta}_{p,n} - \theta^\dagger)/\theta^\dagger$  for the models Case I (left) and Case II (right). The black points show the individual realisations.

For the above two cases, we generate  $M = 100$  independent datasets by applying the locally Gaussian (LG) discretisation scheme defined in [13] with a small step-size  $\Delta = 10^{-4}$ . We treat such data as obtained from the true model (Hypo-II) in the understanding that the discretisation bias is negligible. We then obtain the synthetic complete observations by sub-sampling from the above datasets with coarser time increments. For the two choices of potential above, we consider the designs of high-frequency observations given in Table 2. We specify the true values for  $\theta = (D, \lambda, \alpha, \sigma)$  as  $(2.0, 2.0, 4.0, 4.0)$  and  $(2.0, 1.0, 4.0, 4.0)$  for Case I and Case II, respectively. We compute the estimators  $\hat{\theta}_{p,n} (= \hat{\theta}_{p,n}^{(\text{II})}) = (\hat{D}_{p,n}, \hat{\lambda}_{p,n}, \hat{\alpha}_{p,n}, \hat{\sigma}_{p,n})$  for  $p = 2, 3$ . To find the minima of the contrast functions we used the adaptive moments (Adam) optimiser with the following specifications: (step-size) = 0.1, (exponential decay rate for the first moment estimates) = 0.9, (exponential decay rate for the second moment estimates) = 0.999, (additive term for numerical stability) =  $10^{-8}$  and (number of iterations) = 8,000. Figure 1 shows boxplots of the individual realisations of relative discrepancies

TABLE 3  
Average running times (seconds) of one realisation of estimators for the  $q$ -GLE model.

Experiment	$\hat{\theta}_{2,n}$	$\hat{\theta}_{3,n}$
<b>Design A - Case I</b>	38.25	44.33
<b>Design A - Case II</b>	24.46	28.34
<b>Design B - Case I</b>	27.48	28.71
<b>Design B - Case II</b>	90.26	104.12

$(\hat{\theta}_{p,n}^j - \theta^{\dagger,j})/\theta^{\dagger,j}$ ,  $1 \leq j \leq 4$ . In all four designs, when  $p = 2$  we observe that estimator  $\hat{\sigma}_{2,n}$  suffers from a severe bias, as its realisations do not cover the true value  $\sigma^\dagger$  (see the boxplots at the bottom of Figure 1). In contrast, when  $p = 3$  the 100 realisations of  $\hat{\sigma}_{3,n}$  are concentrated around  $\sigma^\dagger$ . We observe that the estimators of parameters  $D, \lambda$  in the drift functions of the smooth components converge quickly to the true values for both  $p = 2, 3$ . This agrees with the fast convergence rate of  $\sqrt{\Delta_n/n}$  obtained in the CLT result (Theorem 3.2). In Table 3 we show the mean computational time of one realisation of estimators corresponding to  $p = 2$  and  $p = 3$  from 10 independent sets of observations. Notice from the table that in all experiments the calculation of  $\hat{\theta}_{3,n}$  does not add a significant cost compared with that of  $\hat{\theta}_{2,n}$ . We can thus conclude from Figure 1 and Table 3 that use of  $\hat{\theta}_{3,n}$  provides a reliable estimation that does not suffer from the bias observed in the case of  $\hat{\theta}_{2,n}$ , under a competitive computational cost.

#### 4.2. FitzHugh-Nagumo model

We consider the FitzHugh-Nagumo (FHN) model belonging in class (Hypo-I), and specified via the following bivariate SDE:

$$\begin{aligned} dX_t &= \frac{1}{\varepsilon}(X_t - (X_t)^3 - Y_t - s)dt; \\ dY_t &= (\gamma X_t - Y_t + \alpha)dt + \sigma dW_t^1, \end{aligned} \quad (16)$$

with  $\theta = (\varepsilon, \gamma, \alpha, \sigma)$ . We consider parameter estimation of the FHN model in both complete and partial observation regimes, where only the component  $X_t$  is observed in the latter case.

##### 4.2.1. Complete observation regime

We consider the estimator  $\hat{\theta}_{p,n}(= \hat{\theta}_{p,n}^{(1)}) = (\hat{\varepsilon}_{p,n}, \hat{\gamma}_{p,n}, \hat{\alpha}_{p,n}, \hat{\sigma}_{p,n})$  for  $p = 2, 4$ . As [18], we fix  $s = 0.01$  and set the true values to  $\theta^\dagger = (0.10, 1.50, 0.30, 0.60)$ . We generate  $M = 50$  independent datasets by using the LG discretisation for the (Hypo-I) class defined in [7, 14] with a step-size  $\Delta = 10^{-4}$ . Then, we obtain  $M = 50$  synthetic complete observations for the FHN model by sub-sampling from the above datasets so that  $n = 250,000$ ,  $T_n = 5,000$  and  $\Delta_n = 0.02$ . To obtain the minima of the contrast functions we used the Adam optimiser with the same settings as the experiments with the  $q$ -GLE earlier, except for

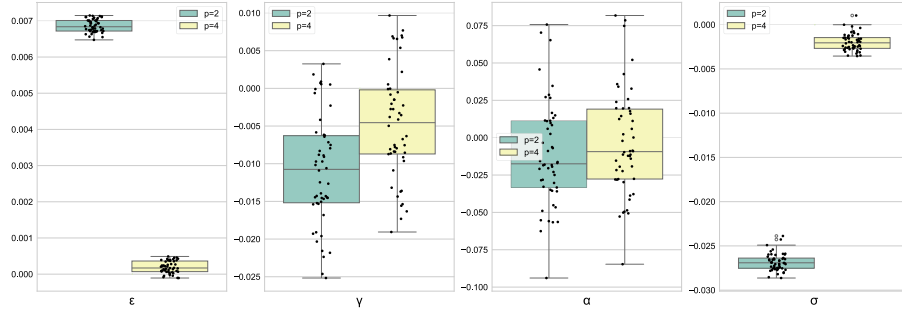


FIG 2. Boxplots of the  $M = 50$  realisations of  $(\hat{\theta}_{p,n} - \theta^\dagger)/\theta^\dagger$  for the FHN model, under the design  $n = 250,000$ ,  $T_n = 5,000$ ,  $\Delta_n = 0.02$ . The black points show the individual values.

(step-size) = 0.01. Figure 2 shows boxplots of the corresponding 50 realisations of relative discrepancies  $(\hat{\theta}_{p,n}^j - \theta^{\dagger,j})/\theta^{\dagger,j}$ ,  $1 \leq j \leq 4$ . Note that the means of  $\hat{\theta}_{p,n} - \theta^\dagger$  with  $p = 4$  are closer to 0 than those with  $p = 2$  for all  $\theta$ -coordinates while both estimators attain similar standard deviations. Importantly, when  $p = 2$ , we see in Figure 2 that  $\hat{\epsilon}_{2,n}$  and  $\hat{\sigma}_{2,n}$  are severely biased as the true values are not covered by the 50 realisations. In contrast, when  $p = 4$ , the realisations of  $\hat{\theta}_{4,n}$  are spread on areas close to the true parameters.

#### 4.2.2. Partial observation regime

We use the proposed contrast function, and the corresponding approximate log-likelihood, under the partial observation regime where only the smooth component  $\{X_{t_i}\}$  is observed. In particular, we compute maximum likelihood estimators (MLEs) based upon a closed-form marginal likelihood constructed from the Gaussian approximation corresponding to the contrast function  $\ell_{p,n}^{(1)}(\theta)$  for  $p = 2, 3$ . To be precise, we use the high-order expansions for the mean and the variance, but do not make use of the Taylor expansions for the inverse of the covariance and of its log-determinant, so that we obtain quantities that correspond to proper density functions. Here, we mention that due to the structure of the FHN model, the obtained one-step Gaussian approximation corresponding to the developed contrast function for  $p = 2, 3$  is a linear Gaussian model w.r.t. the hidden component  $Y_{t_i}$  given the observation  $X_{t_i}$ . Thus, one can make use of the Kalman Filter (KF) and calculate the marginal likelihood of the partial observations. See e.g. [13] where a marginal likelihood is built upon KF for the  $q$ -GLE model belonging in class (Hypo-II). We provide the details of the Gaussian approximation and the marginal likelihood for the FHN model in Section D of the Supplementary Material [12]. Then, the MLE for partial observation is defined as:

$$\hat{\theta}_{p,n} = (\hat{\epsilon}_{p,n}, \hat{\gamma}_{p,n}, \hat{\alpha}_{p,n}, \hat{\sigma}_{p,n}) = \arg \max_{\theta \in \Theta} \log f_{p,n}(\{X_{t_i}\}_{i=0,\dots,n}; \theta), \quad (17)$$

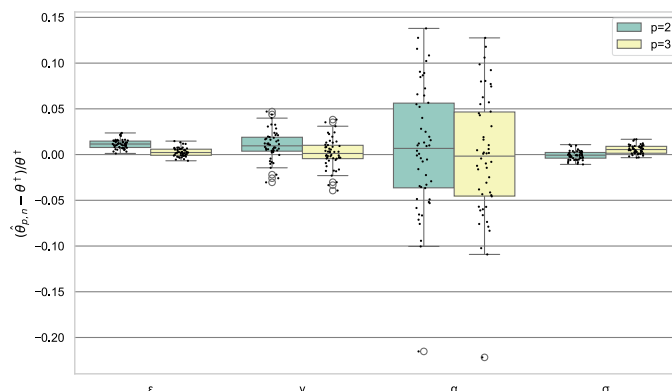


FIG 3. Boxplots of  $(\hat{\theta}_{p,n} - \theta^\dagger)/\theta^\dagger$  from  $M = 50$  sets of partial observations for the FHN model, under the design  $n = 200,000$ ,  $T_n = 1,000$  and  $\Delta_n = 0.005$ . The black points show the individual values.

TABLE 4

Average running times (seconds) of one realisation of estimators for the FHN model, from partial observations.

$\hat{\theta}_{2,n}$	$\hat{\theta}_{3,n}$
100.976	138.428

for  $p = 2, 3$ , where  $f_{p,n}(\{X_{t_i}\}_{i=0,\dots,n}; \theta)$  is the marginal likelihood. As with the experiment under complete observations, we fix  $s = 0.01$  and set the true values to  $\theta^\dagger = (0.10, 1.50, 0.30, 0.60)$ . We generate  $M = 50$  independent datasets from the LG discretisation with a step-size  $\Delta = 10^{-4}$ . Then, we obtain 50 synthetic partial observations of  $\{X_{t_i}\}_{i=0,\dots,n}$  with  $n = 200,000$ ,  $T_n = 1,000$  and  $\Delta_n = 0.005$  by sub-sampling from the datasets and removing the rough component  $\{Y_{t_i}\}_{i=0,\dots,n}$ . The Nelder-Mead method is applied to optimise the log-marginal likelihood for  $p = 2, 3$  with the initial guess  $\theta_0 = (0.5, 0.5, 0.5, 0.5)$ . Figure 3 shows boxplots of the corresponding 50 realisations of relative discrepancies  $(\hat{\theta}_{p,n}^j - \theta^{\dagger,j})/\theta^{\dagger,j}$ ,  $1 \leq j \leq 4$ . First, we observe that both estimators share almost the same standard deviations. Second, the bias in  $\hat{\varepsilon}_{2,n}$  is clear from the boxplot in the sense that the true value  $\varepsilon^\dagger$  is not included in the interval between the minimum and the maximum of the 50 realisations of  $\hat{\varepsilon}_{2,n}$ . Finally, the square of the mean of  $\hat{\theta}_{p,n} - \theta^\dagger$  (eq. bias) is reduced in the drift parameters ( $\varepsilon$ ,  $\gamma$  and  $\alpha$ ) for  $p = 3$ . Table 4 presents the average computational times of one realisation of estimators  $\hat{\theta}_{2,n}$  and  $\hat{\theta}_{3,n}$  from 10 independent sets of partial observations. The increase in the average time of  $\hat{\theta}_{3,n}$  from that of  $\hat{\theta}_{2,n}$  is about 40%, thus use of  $\hat{\theta}_{3,n}$  is quite feasible.

## 5. Proof of main results

We provide the proof for our main results, Theorems 3.1 and 3.2, which demonstrate the asymptotic normality of the proposed contrast estimators under the design condition  $\Delta_n = o(n^{-1/p})$ ,  $p \geq 2$ . We show all details of the proof for class (Hypo-II) as the proof for (Hypo-I) follows from similar arguments. Throughout this section, for simplicity of notation, we frequently omit the subscript (II) appearing in the estimator and the functions introduced in Section 2.2.3. A number of technical results are collected in the Supplementary Material [12]. We make use of the notation  $o_{\mathbb{P}_{\theta^\dagger}}(\cdot)$ : for a sequence of random variables  $\{F_n\}$  and a numerical sequence  $\{\alpha_{n,\Delta_n}\}$  depending on  $n$  or  $\Delta_n$ , we write  $F_n = o_{\mathbb{P}_{\theta^\dagger}}(\alpha_{n,\Delta_n})$  if  $F_n/\alpha_{n,\Delta_n} \xrightarrow{\mathbb{P}_{\theta^\dagger}} 0$  as  $n \rightarrow \infty$ ,  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$ . Throughout this section,  $p \geq 2$  is an arbitrary integer, and we set  $K_p = \lfloor p/2 \rfloor$  as before.

### 5.1. Proof of Theorem 3.1 (Consistency)

We show the consistency of the contrast estimator via the following procedure:

**Step 1.** We show that if  $n \rightarrow \infty$ ,  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$ , then  $\hat{\beta}_{S_1,p,n} \xrightarrow{\mathbb{P}_{\theta^\dagger}} \beta_{S_1}^\dagger$ . In particular:

$$\hat{\beta}_{S_1,p,n} - \beta_{S_1}^\dagger = o_{\mathbb{P}_{\theta^\dagger}}(\Delta_n^{3/2}). \quad (18)$$

**Step 2.** Making use of the rate of convergence found in (18), we show that if  $n \rightarrow \infty$ ,  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$ , then  $\hat{\beta}_{S_2,p,n} \xrightarrow{\mathbb{P}_{\theta^\dagger}} \beta_{S_2}^\dagger$ . In particular:

$$\hat{\beta}_{S_2,p,n} - \beta_{S_2}^\dagger = o_{\mathbb{P}_{\theta^\dagger}}(\Delta_n^{1/2}). \quad (19)$$

**Step 3.** Using the rates obtained in (18)–(19), we show that if  $n \rightarrow \infty$ ,  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$ , then  $(\hat{\beta}_{R,p,n}, \hat{\sigma}_{p,n}) \xrightarrow{\mathbb{P}_{\theta^\dagger}} (\beta_R^\dagger, \sigma^\dagger)$ .

In the sequel, we express the matrix  $\mathbf{\Lambda}(x, \theta) = (\mathbf{\Sigma}^{(II)}(x, \theta))^{-1}$ ,  $(x, \theta) \in \mathbb{R}^N \times \Theta$ , as:

$$\mathbf{\Lambda}(x, \theta) = \begin{bmatrix} \mathbf{\Lambda}_{S_1 S_1}(x, \theta) & \mathbf{\Lambda}_{S_1 S_2}(x, \theta) & \mathbf{\Lambda}_{S_1 R}(x, \theta) \\ \mathbf{\Lambda}_{S_2 S_1}(x, \theta) & \mathbf{\Lambda}_{S_2 S_2}(x, \theta) & \mathbf{\Lambda}_{S_2 R}(x, \theta) \\ \mathbf{\Lambda}_{R S_1}(x, \theta) & \mathbf{\Lambda}_{R S_2}(x, \theta) & \mathbf{\Lambda}_{R R}(x, \theta) \end{bmatrix},$$

for block matrices  $\mathbf{\Lambda}_{\iota_1 \iota_2}(x, \theta) \in \mathbb{R}^{N_{\iota_1} \times N_{\iota_2}}$ ,  $\iota_1, \iota_2 \in \{S_1, S_2, R\}$ . Note that from Lemma 13 in [13] we have:

$$\mathbf{\Lambda}_{S_1 S_1}(x, \theta) = 720 a_{S_1}^{-1}(x, \theta), \quad (x, \theta) \in \mathbb{R}^N \times \Theta,$$

where  $a_{S_1}(x, \theta)$  is specified in the definition of the matrix  $\mathbf{\Sigma}^{(II)}(x, \theta)$  in (14).

**Remark 5.1.** The strategy outlined above is technically different from proofs in [18, 7], where they require that  $\Delta_n = o(n^{-1/2})$  to prove consistency. Our proof of consistency proceeds without relying on  $\Delta_n = o(n^{-1/2})$ , and such an approach then leads to a CLT under the weaker condition  $\Delta_n = o(n^{-1/p})$ ,  $p \geq 2$ . More details on this point are given in Remark 5.4.

## 5.1.1. Step 1

The consistency of the estimator  $\hat{\beta}_{S_1,p,n}$  is deduced from the following result.

**Lemma 5.2.** *Assume that conditions [H1](#), [H2](#)( $2K_p$ ), [H3](#), [H4](#) and [H5](#) hold. If  $n \rightarrow \infty$ ,  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$ , then:*

$$\sup_{\theta \in \Theta} \left| \frac{\Delta_n^3}{n} \ell_{p,n}(\theta) - \mathbf{Y}_1(\theta) \right| \xrightarrow{\mathbb{P}_{\theta^\dagger}} 0, \quad (20)$$

where, for  $\theta \in \Theta$ :

$$\mathbf{Y}_1(\theta) = \int \mathbf{q} \left( \mu_{S_1}(x_S, \beta_{S_1}) - \mu_{S_1}(x_S, \beta_{S_1}^\dagger); \mathbf{A}_{S_1 S_1}(x, \theta) \right) \nu_{\theta^\dagger}(dx).$$

The proof is given in Section B.1 of the Supplementary Material [\[12\]](#). Lemma [5.2](#) indeed implies the consistency of  $\hat{\beta}_{S_1,p,n}$  via the following arguments. We first notice that the matrix  $\mathbf{A}_{S_1 S_1}(x, \theta) = 720a_S^{-1}(x, \theta)$  is positive definite for any  $(x, \theta) \in \mathbb{R}^N \times \Theta$  under [H1](#) due to Lemma [2.5](#). From the identifiability condition [H6](#) and the positive definiteness of  $\mathbf{A}_{S_1 S_1}$ , the term  $\mathbf{Y}_1(\theta)$ ,  $\theta \in \Theta$ , should be positive if  $\beta_{S_1} \neq \beta_{S_1}^\dagger$ . Thus, for any  $\varepsilon > 0$ , there exists a constant  $\delta > 0$  so that:

$$\mathbb{P}_{\theta^\dagger}(|\hat{\beta}_{S_1,p,n} - \beta_{S_1}^\dagger| > \varepsilon) \leq \mathbb{P}_{\theta^\dagger}(\mathbf{Y}_1(\hat{\theta}_{p,n}) > \delta).$$

From the definition of the estimator and Lemma [5.2](#), we have:

$$\begin{aligned} & \mathbb{P}_{\theta^\dagger}(\mathbf{Y}_1(\hat{\theta}_{p,n}) > \delta) \\ & \leq \mathbb{P}_{\theta^\dagger} \left( \frac{\Delta_n^3}{n} \ell_{p,n}(\beta_{S_1}^\dagger, \hat{\beta}_{S_2,p,n}, \hat{\beta}_{R,p,n}, \hat{\sigma}_{p,n}) - \frac{\Delta_n^3}{n} \ell_{p,n}(\hat{\theta}_{p,n}) + \mathbf{Y}_1(\hat{\theta}_{p,n}) > \delta \right) \\ & \leq \mathbb{P}_{\theta^\dagger} \left( \sup_{\theta \in \Theta} \left| \frac{\Delta_n^3}{n} \ell_{p,n}(\beta_{S_1}^\dagger, \beta_{S_2}, \beta_R, \sigma) - \frac{\Delta_n^3}{n} \ell_{p,n}(\theta) + \mathbf{Y}_1(\theta) \right| > \delta \right) \rightarrow 0, \end{aligned}$$

as  $n \rightarrow \infty$ ,  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$ , thus  $\hat{\beta}_{S_1,p,n}$  is a consistent estimator.

We now prove [\(18\)](#). A Taylor expansion of  $\partial_{\beta_{S_1}} \ell_{p,n}(\hat{\theta}_{p,n})$  gives:

$$\mathbf{A}_{p,n}(\beta_{S_1}^\dagger, \hat{\beta}_{S_2,p,n}, \hat{\beta}_{R,p,n}, \hat{\sigma}_{p,n}) = \mathbf{B}_{p,n}(\hat{\theta}_{p,n}) \times \frac{1}{\Delta_n^{3/2}}(\hat{\beta}_{S_1,p,n} - \beta_{S_1}^\dagger),$$

where we have set, for  $\theta = (\beta_{S_1}, \theta^{S_2,R}) \in \Theta$ ,  $\theta^{S_2,R} \equiv (\beta_{S_2}, \beta_R, \sigma) \in \Theta_{\beta_{S_2}} \times \Theta_{\beta_R} \times \Theta_\sigma$ :

$$\begin{aligned} \mathbf{A}_{p,n}(\theta) &= -\frac{\Delta_n^{3/2}}{n} \partial_{\beta_{S_1}} \ell_{p,n}(\theta); \\ \mathbf{B}_{p,n}(\theta) &= \frac{\Delta_n^3}{n} \int_0^1 \partial_{\beta_{S_1}}^2 \ell_{p,n}(\beta_{S_1}^\dagger + \lambda(\beta_{S_1} - \beta_{S_1}^\dagger), \theta^{S_2,R}) d\lambda. \end{aligned}$$

For simplicity of notation, we write  $\Theta_{\beta_{S_2}, \beta_R, \sigma} \equiv \Theta_{\beta_{S_2}} \times \Theta_{\beta_R} \times \Theta_\sigma$ . Limit [\(18\)](#) holds from the following result whose proof is given in Section B.2 of the Supplementary Material [\[12\]](#).

**Lemma 5.3.** Assume that conditions [H1](#), [H2](#)( $2K_p$ ), [H3](#), [H4](#), [H5](#) and [H6](#) hold. If  $n \rightarrow \infty$ ,  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$ , then:

$$\sup_{(\beta_{S_2}, \beta_R, \sigma) \in \Theta_{\beta_{S_2}, \beta_R, \sigma}} \left| \mathbf{A}_{p,n}(\beta_{S_1}^\dagger, \beta_{S_2}, \beta_R, \sigma) \right| \xrightarrow{\mathbb{P}_{\theta^\dagger}} 0; \quad (21)$$

$$\sup_{(\beta_{S_2}, \beta_R, \sigma) \in \Theta_{\beta_{S_2}, \beta_R, \sigma}} \left| \mathbf{B}_{p,n}(\hat{\beta}_{S_1,p,n}, \beta_{S_2}, \beta_R, \sigma) - 2\mathbf{B}(\beta_{S_1}^\dagger, \beta_{S_2}, \sigma) \right| \xrightarrow{\mathbb{P}_{\theta^\dagger}} 0, \quad (22)$$

where for  $(\beta_{S_2}, \beta_R, \sigma) \in \Theta_{\beta_{S_2}} \times \Theta_{\beta_R} \times \Theta_\sigma$ :

$$\begin{aligned} & \mathbf{B}(\beta_{S_1}, \beta_{S_2}, \sigma) \\ &= 720 \int \left( \partial_{\beta_{S_1}}^\top \mu_{S_1}(x_S, \beta_{S_1}) \right)^\top a_{S_1}^{-1}(x, (\beta_{S_1}, \beta_{S_2}, \sigma)) \partial_{\beta_{S_1}}^\top \mu_{S_1}(x_S, \beta_{S_1}) \nu_{\theta^\dagger}(dx). \end{aligned}$$

**Remark 5.4.** We point here to a key fact to obtain (21) leading to the rate  $\hat{\beta}_{S_1,p,n} - \beta_{S_1}^\dagger = o_{\mathbb{P}_{\theta^\dagger}}(\Delta_n^{3/2})$ . The term  $\mathbf{A}_{p,n}$  is given in the form of:

$$\begin{aligned} \mathbf{A}_{p,n}(\beta_{S_1}^\dagger, \beta_{S_2}, \beta_R, \sigma) &= \frac{1}{n\sqrt{\Delta_n}} \sum_{i=1}^n F(1, X_{t_{i-1}}, (\beta_{S_1}^\dagger, \beta_{S_2}, \beta_R, \sigma)) \\ &\quad + R(\Delta_n, (\beta_{S_1}^\dagger, \beta_{S_2}, \beta_R, \sigma)), \end{aligned} \quad (23)$$

where  $F = [F^k]_{1 \leq k \leq N_{\beta_{S_1}}}$  with  $F^k \in \mathcal{S}$ , and the second term is a residual such that:

$$\sup_{(\beta_{S_2}, \beta_R, \sigma) \in \Theta_{\beta_{S_2}} \times \Theta_{\beta_R} \times \Theta_\sigma} \left| R(\Delta_n, (\beta_{S_1}^\dagger, \beta_{S_2}, \beta_R, \sigma)) \right| \xrightarrow{\mathbb{P}_{\theta^\dagger}} 0.$$

The first term of the right side of (23) includes  $\Delta_n^{-1/2}$ , however  $F^k(1, x, \theta)$  is identically 0 for any  $(x, \theta) \in \mathbb{R}^N \times \Theta$  following some matrix algebra (as indicated in Lemma B.1 of the Supplementary Material [12]) and then (21) holds. A similar argument related with matrix algebra (see, e.g., Lemmas B.2, B.3 [12]) is also used in the proofs of other technical lemmas below to deal with terms of size  $\mathcal{O}(\Delta_n^{-1/2})$ , and then the proof of consistency proceeds without requiring that  $\Delta_n = o(n^{-1/2})$ .

### 5.1.2. Step 2

Making use of limit (18), we obtain the following result whose proof is postponed to Section B.3 of the Supplementary Material [12].

**Lemma 5.5.** Assume that conditions [H1](#), [H2](#)( $2K_p$ ), [H3](#), [H4](#), [H5](#) and [H6](#) hold. If  $n \rightarrow \infty$ ,  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$ , then:

$$\sup_{(\beta_{S_2}, \beta_R, \sigma) \in \Theta_{\beta_{S_2}} \times \Theta_{\beta_R} \times \Theta_\sigma} \left| \frac{\Delta_n}{n} \ell_{p,n}(\hat{\beta}_{S_1,p,n}, \beta_{S_2}, \beta_R, \sigma) - \mathbf{Y}_2(\beta_{S_2}, \beta_R, \sigma) \right| \xrightarrow{\mathbb{P}_{\theta^\dagger}} 0,$$



where:

$$\begin{aligned} \mathbf{Y}_2(\beta_{S_2}, \beta_R, \sigma) \\ = 12 \int \mathbf{q}(\mu_{S_2}(x, \beta_{S_2}) - \mu_{S_2}(x, \beta_{S_2}^\dagger); a_{S_2}^{-1}(x, (\beta_{S_1}^\dagger, \beta_{S_2}, \sigma))) \nu_{\theta^\dagger}(dx). \end{aligned}$$

Lemma 5.5 leads to the consistency of  $\hat{\beta}_{S_2,p,n}$  following arguments similar to the ones used in Section 5.1.1 to show the consistency of  $\hat{\beta}_{S_1,p,n}$ .

To obtain the rate of convergence in (19), we consider the Taylor expansion of  $\partial_{\beta_S} \ell_{p,n}(\hat{\theta}_{p,n})$  at  $(\beta_S^\dagger, \hat{\beta}_{R,p,n}, \hat{\sigma}_{p,n})$ :

$$\tilde{\mathbf{A}}_{p,n}(\beta_{S_1}^\dagger, \beta_{S_2}^\dagger, \hat{\beta}_{R,p,n}, \hat{\sigma}_{p,n}) = \tilde{\mathbf{B}}_{p,n}(\hat{\theta}_{p,n}) \begin{bmatrix} \frac{1}{\sqrt{\Delta_n^3}}(\hat{\beta}_{S_1,p,n} - \beta_{S_1}^\dagger) \\ \frac{1}{\sqrt{\Delta_n}}(\hat{\beta}_{S_2,p,n} - \beta_{S_2}^\dagger) \end{bmatrix}, \quad (24)$$

where we have set for  $\theta = (\beta_S, \beta_R, \sigma) \in \Theta$  with  $\beta_S = (\beta_{S_1}, \beta_{S_2}) \in \Theta_{\beta_S}$ :

$$\begin{aligned} \tilde{\mathbf{A}}_{p,n}(\theta) &= \begin{bmatrix} -\frac{\sqrt{\Delta_n^3}}{n} \partial_{\beta_{S_1}} \ell_{p,n}(\theta) \\ -\frac{\sqrt{\Delta_n}}{n} \partial_{\beta_{S_2}} \ell_{p,n}(\theta) \end{bmatrix}; \\ \tilde{\mathbf{B}}_{p,n}(\theta) &= \tilde{\mathbf{M}}_n \left( \int_0^1 \partial_{\beta_S}^2 \ell_{p,n}(\beta_S^\dagger + \lambda(\beta_S - \beta_S^\dagger), \beta_R, \sigma) d\lambda \right) \tilde{\mathbf{M}}_n^\top, \end{aligned}$$

for the matrix  $\tilde{\mathbf{M}}_n = \text{Diag}(\tilde{v}_n)$ , where:

$$\tilde{v}_n = \left[ \underbrace{\sqrt{\frac{\Delta_n^3}{n}}, \dots, \sqrt{\frac{\Delta_n^3}{n}}}_{N_{\beta_{S_1}}}, \underbrace{\sqrt{\frac{\Delta_n}{n}}, \dots, \sqrt{\frac{\Delta_n}{n}}}_{N_{\beta_{S_2}}} \right]^\top.$$

The convergence (19) is immediately deduced from (24) given the following result whose proof is provided in Section B.4 of the Supplementary Material [12].

**Lemma 5.6.** Assume that conditions H1, H2(2K<sub>p</sub>), H3, H4, H5 and H6 hold. If  $n \rightarrow \infty$ ,  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$ , then:

$$\sup_{(\beta_R, \sigma) \in \Theta_{\beta_R} \times \Theta_\sigma} \left| \tilde{\mathbf{A}}_{p,n}(\beta_S^\dagger, \beta_R, \sigma) \right| \xrightarrow{\mathbb{P}_{\theta^\dagger}} 0; \quad (25)$$

$$\sup_{(\beta_R, \sigma) \in \Theta_{\beta_R} \times \Theta_\sigma} \left| \tilde{\mathbf{B}}_{p,n}(\beta_S^\dagger, \beta_R, \sigma) - 2\tilde{\mathbf{B}}(\beta_S^\dagger, \beta_R, \sigma) \right| \xrightarrow{\mathbb{P}_{\theta^\dagger}} 0; \quad (26)$$

where we have set for  $\theta \in \Theta$ ,  $\tilde{\mathbf{B}}(\theta) = \text{Diag}[\tilde{\mathbf{B}}^{S_1}(\theta), \tilde{\mathbf{B}}^{S_2}(\theta)]$  with  $\tilde{\mathbf{B}}^{S_1}(\theta) \in \mathbb{R}^{N_{\beta_{S_1}} \times N_{\beta_{S_1}}}$  and  $\tilde{\mathbf{B}}^{S_2}(\theta) \in \mathbb{R}^{N_{\beta_{S_2}} \times N_{\beta_{S_2}}}$  defined as:

$$\tilde{\mathbf{B}}^{S_1}(\theta) = 720 \int (\partial_{\beta_{S_1}}^\top \mu_{S_1}(x_S, \beta_{S_1}))^\top a_{S_1}^{-1}(x, \theta) \partial_{\beta_{S_1}}^\top \mu_{S_1}(x_S, \beta_{S_1}) \nu_{\theta^\dagger}(dx); \quad (27)$$

$$\tilde{\mathbf{B}}^{S_2}(\theta) = 12 \int (\partial_{\beta_{S_2}}^\top \mu_{S_2}(x, \beta_{S_2}))^\top a_{S_2}^{-1}(x, \theta) \partial_{\beta_{S_2}}^\top \mu_{S_2}(x, \beta_{S_2}) \nu_{\theta^\dagger}(dx). \quad (28)$$

### 5.1.3. Step 3

Finally, we show the consistency of estimators  $(\hat{\beta}_{R,p,n}, \hat{\sigma}_{p,n})$ . Working with the rates of convergence obtained in (18) and (19), we prove the following result leading to the consistency of  $\hat{\sigma}_{p,n}$ .

**Lemma 5.7.** *Assume that conditions H1, H2(2K<sub>p</sub>), H3, H4, H5 and H6 hold. If  $n \rightarrow \infty$ ,  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$ , then:*

$$\sup_{(\beta_R, \sigma) \in \Theta_{\beta_R} \times \Theta_\sigma} \left| \frac{1}{n} \ell_{p,n}(\hat{\beta}_{S,p,n}, \beta_R, \sigma) - \mathbf{Y}_3(\sigma) \right| \xrightarrow{\mathbb{P}_{\theta^\dagger}} 0,$$

where we have set for  $\sigma \in \Theta_\sigma$ :

$$\mathbf{Y}_3(\sigma) = \int \left\{ \text{Tr}(\mathbf{\Lambda}(x, (\beta_S^\dagger, \sigma)) \mathbf{\Sigma}(x, (\beta_S^\dagger, \sigma^\dagger))) + \log \det \mathbf{\Sigma}(x, (\beta_S^\dagger, \sigma)) \right\} \nu_{\theta^\dagger}(dx).$$

We provide the proof in Section B.5 of the Supplementary Material [12]. We will show that Lemma 5.7 indeed leads to the consistency of  $\hat{\sigma}_{p,n}$ . We have:

$$\mathbf{Y}_3(\sigma) - \mathbf{Y}_3(\sigma^\dagger) = 2 \int \left( \int_{\mathbb{R}^N} \varphi(y; x, \sigma^\dagger) \log \frac{\varphi(y; x, \sigma^\dagger)}{\varphi(y; x, \sigma)} dy \right) \nu_{\theta^\dagger}(dx), \quad \sigma \in \Theta,$$

with  $y \mapsto \varphi(y; x, \sigma)$ ,  $(x, \sigma) \in \mathbb{R}^N \times \Theta$  being the density of the distribution  $\mathcal{N}(\mathbf{0}_N, \mathbf{\Sigma}(x, (\beta_S^\dagger, \sigma)))$ . Thus, under condition H6,  $\mathbf{Y}_3(\sigma) - \mathbf{Y}_3(\sigma^\dagger)$  should be positive if  $\sigma \neq \sigma^\dagger$ . Hence, for every  $\varepsilon > 0$ , there exists a constant  $\delta > 0$  so that:

$$\begin{aligned} \mathbb{P}_{\theta^\dagger}(|\hat{\sigma}_{p,n} - \sigma^\dagger| > \varepsilon) &\leq \mathbb{P}_{\theta^\dagger}(\mathbf{Y}_3(\hat{\sigma}_{p,n}) - \mathbf{Y}_3(\sigma^\dagger) > \delta) \\ &\leq \mathbb{P}_{\theta^\dagger}\left(\frac{1}{n} \ell_{p,n}(\hat{\beta}_{p,n}, \sigma^\dagger) - \frac{1}{n} \ell_{p,n}(\hat{\theta}_{p,n}) + \mathbf{Y}_3(\hat{\sigma}_{p,n}) - \mathbf{Y}_3(\sigma^\dagger) > \delta\right) \rightarrow 0, \end{aligned}$$

as  $n \rightarrow \infty$ ,  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$ .

To show the consistency of  $\hat{\beta}_{R,p,n}$ , we consider, for  $\theta = (\beta_{S_1}, \beta_{S_2}, \beta_R, \sigma) \in \Theta$ :

$$\mathbf{K}(\theta) := \frac{1}{n\Delta_n} \ell_{p,n}(\theta) - \frac{1}{n\Delta_n} \ell_{p,n}(\beta_{S_1}, \beta_{S_2}, \beta_R^\dagger, \sigma). \quad (29)$$

The consistency of  $\hat{\beta}_{R,p,n}$  follows from the result below whose proof is given in Section B.6 of the Supplementary Material [12].

**Lemma 5.8.** *Assume that conditions H1, H2(2K<sub>p</sub>), H3, H4, H5 and H6 hold. If  $n \rightarrow \infty$ ,  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$ , then:*

$$\sup_{\beta_R \in \Theta_{\beta_R}} \left| \mathbf{K}(\hat{\beta}_{S_1,p,n}, \hat{\beta}_{S_2,p,n}, \beta_R, \hat{\sigma}_{p,n}) - \mathbf{Y}_4(\beta_R) \right| \xrightarrow{\mathbb{P}_{\theta^\dagger}} 0,$$

where we have set for  $\beta_R \in \Theta_{\beta_R}$ :

$$\mathbf{Y}_4(\beta_R) = \int \mathbf{q}(\mu_R(x, \beta_R) - \mu_R(x, \beta_R^\dagger); a_R^{-1}(x, \sigma^\dagger)) \nu_{\theta^\dagger}(dx).$$

Thus, the proof of consistency for the proposed estimator is now complete.

### 5.2. Proof of Theorem 3.2 (CLT)

We define the  $N_\theta \times N_\theta$  matrix  $\mathbf{M}_n$  as  $\mathbf{M}_n = \text{Diag}(v_n)$ , where:

$$v_n = \left[ \underbrace{\sqrt{\frac{n}{\Delta_n^3}}, \dots, \sqrt{\frac{n}{\Delta_n^3}}}_{N_{\beta_{S_1}}}, \underbrace{\sqrt{\frac{n}{\Delta_n}}, \dots, \sqrt{\frac{n}{\Delta_n}}}_{N_{\beta_{S_2}}}, \underbrace{\sqrt{n\Delta_n}, \dots, \sqrt{n\Delta_n}}_{N_{\beta_R}}, \underbrace{\sqrt{n}, \dots, \sqrt{n}}_{N_\sigma} \right].$$

Noting that  $\partial_\theta \ell_{p,n}(\hat{\theta}_{p,n}) = \mathbf{0}_{N_\theta}$ , a Taylor expansion of  $\partial_\theta \ell_{p,n}(\hat{\theta}_{p,n})$  at the true parameter  $\theta^\dagger$  gives:

$$\mathbf{I}_{p,n}(\theta^\dagger) = \int_0^1 \mathbf{J}_{p,n}(\theta^\dagger + \lambda(\hat{\theta}_{p,n} - \theta^\dagger)) d\lambda \times \mathbf{M}_n(\hat{\theta}_{p,n} - \theta^\dagger), \quad (30)$$

where we have set for  $\theta \in \Theta$ :

$$\mathbf{I}_{p,n}(\theta) \equiv -\mathbf{M}_n^{-1} \partial_\theta \ell_{p,n}(\theta), \quad \mathbf{J}_{p,n}(\theta) \equiv \mathbf{M}_n^{-1} \partial_\theta^2 \ell_{p,n}(\theta) \mathbf{M}_n^{-1}.$$

To prove the asymptotic normality, we exploit the following two results.

**Lemma 5.9.** *Assume that conditions [H1](#), [H2](#)( $2K_p$ ), [H3](#), [H4](#), [H5](#) and [H6](#) hold. If  $n \rightarrow \infty$ ,  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$ , then:*

$$\sup_{\lambda \in [0,1]} \left| \mathbf{J}_{p,n}(\theta^\dagger + \lambda(\hat{\theta}_{p,n} - \theta^\dagger)) - 2\Gamma(\theta^\dagger) \right| \xrightarrow{\mathbb{P}_{\theta^\dagger}} 0.$$

**Proposition 5.10.** *Assume that conditions [H1](#), [H2](#)( $2K_p$ ), [H3](#), [H4](#) and [H5](#) hold. If  $n \rightarrow \infty$ ,  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$ , with the additional design condition  $\Delta_n = o(n^{-1/p})$ , then:*

$$\mathbf{I}_{p,n}(\theta^\dagger) \xrightarrow{\mathcal{L}_{\theta^\dagger}} \mathcal{N}(\mathbf{0}_{N_\theta}, 4\Gamma(\theta^\dagger)).$$

We provide the proof of Lemma 5.9 in Section C.1 of the Supplementary Material [12]. We show the proof of Proposition 5.10 in the next subsection where we highlight the manner in which we make use of the condition  $\Delta_n = o(n^{-1/p})$ ,  $p \geq 2$ . By applying these two results to equation (30), the proof of Theorem 3.2 is complete.

#### 5.2.1. Proof of Proposition 5.10

For simplicity of notation, we write  $\partial_{\theta,k} \ell_{p,n}(\theta) = \sum_{i=1}^n \xi_i^k(\theta)$ ,  $\theta \in \Theta$ ,  $1 \leq k \leq N_\theta$ , where:

$$\xi_i^k(\theta) = [\mathbf{M}_n^{-1}]_{kk} \times \sum_{j=0}^{K_p} \Delta_n^j \cdot \partial_{\theta,k} \left\{ \mathbf{m}_{p,i}(\Delta_n, \theta)^\top \mathbf{G}_{i-1,j}(\theta) \mathbf{m}_{p,i}(\Delta_n, \theta) + \mathbf{H}_{i-1,j}(\theta) \right\}. \quad (31)$$

Due to Theorems 4.2 & 4.4 in [10], Proposition 5.10 holds once we prove the following convergences:

(i) If  $n \rightarrow \infty$ ,  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$  with  $\Delta_n = o(n^{-1/p})$ , then:

$$\sum_{i=1}^n \mathbb{E}_{\theta^\dagger} [\xi_i^k(\theta^\dagger) | \mathcal{F}_{t_{i-1}}] \xrightarrow{\mathbb{P}_{\theta^\dagger}} 0, \quad 1 \leq k \leq N_\theta. \quad (32)$$

(ii) If  $n \rightarrow \infty$ ,  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$ , then:

$$\sum_{i=1}^n \mathbb{E}_{\theta^\dagger} [\xi_i^{k_1}(\theta^\dagger) \xi_i^{k_2}(\theta^\dagger) | \mathcal{F}_{t_{i-1}}] \xrightarrow{\mathbb{P}_{\theta^\dagger}} 4[\Gamma(\theta^\dagger)]_{k_1 k_2}, \quad 1 \leq k_1, k_2 \leq N_\theta; \quad (33)$$

$$\sum_{i=1}^n \mathbb{E}_{\theta^\dagger} [(\xi_i^{k_1}(\theta^\dagger) \xi_i^{k_2}(\theta^\dagger))^2 | \mathcal{F}_{t_{i-1}}] \xrightarrow{\mathbb{P}_{\theta^\dagger}} 0, \quad 1 \leq k_1, k_2 \leq N_\theta. \quad (34)$$

Notice that the design condition  $\Delta_n = o(n^{-1/p})$  is used in the proof that the expectation of the score function converges to 0, i.e., for convergence (32). We provide the proof of the two convergences (33) and (34) in Section C.2 of the Supplementary Material [12] and focus on the proof of (32) in this section.

*Proof of convergence (32).* Let  $1 \leq k \leq N_\theta$ . We write for  $\theta \in \Theta$  and  $1 \leq i \leq n$ :

$$\partial_{\theta,k} \mathbf{m}_{p,i}(\Delta_n, \theta) \equiv v_{p,k}(\Delta_n, X_{t_{i-1}}, \theta).$$

Note that the  $\mathbb{R}^N$ -valued function  $v_{p,k}(\Delta_n, X_{t_{i-1}}, \theta)$  is independent of  $X_{t_i}$ . It then follows that:

$$\sum_{i=1}^n \mathbb{E}_{\theta^\dagger} [\xi_i^k(\theta^\dagger) | \mathcal{F}_{t_{i-1}}] = \sum_{i=1}^n \{F_{i-1}^{(1),k}(\theta^\dagger) + F_{i-1}^{(2),k}(\theta^\dagger)\},$$

where we have set:

$$\begin{aligned} F_{i-1}^{(1),k}(\theta^\dagger) &= 2[\mathbf{M}_n^{-1}]_{kk} \\ &\times \sum_{j=0}^{K_p} \sum_{l_1, l_2=1}^N \Delta_n^j v_{p,k}^{l_1}(\Delta_n, X_{t_{i-1}}, \theta^\dagger) [\mathbf{G}_{i-1,j}(\theta^\dagger)]_{l_1 l_2} \mathbb{E}_{\theta^\dagger} [\mathbf{m}_{p,i}^{l_2}(\Delta_n, \theta^\dagger) | \mathcal{F}_{t_{i-1}}]; \\ F_{i-1}^{(2),k}(\theta^\dagger) &= [\mathbf{M}_n^{-1}]_{kk} \\ &\times \sum_{j=0}^{K_p} \Delta_n^j \left\{ \sum_{l_1, l_2=1}^N [\partial_{\theta,k} \mathbf{G}_{i-1,j}(\theta^\dagger)]_{l_1 l_2} \mathbb{E}_{\theta^\dagger} [\mathbf{m}_{p,i}^{l_1}(\Delta_n, \theta^\dagger) \mathbf{m}_{p,i}^{l_2}(\Delta_n, \theta^\dagger) | \mathcal{F}_{t_{i-1}}] \right. \\ &\quad \left. + \partial_{\theta,k} \mathbf{H}_{i-1,j}(\theta^\dagger) \right\}. \end{aligned}$$

For the first term, it follows from (15) that:

$$F_{i-1}^{(1),k}(\theta^\dagger) = \begin{cases} \frac{1}{n} R^{(1),k}(\sqrt{n\Delta_n^{2K_p+1}}, X_{t_{i-1}}, \theta^\dagger), & 1 \leq k \leq N_\beta; \\ \frac{1}{n} R^{(1),k}(\sqrt{n\Delta_n^{2K_p+2}}, X_{t_{i-1}}, \theta^\dagger), & N_\beta + 1 \leq k \leq N_\theta, \end{cases}$$

for some  $R^{(1),k} \in \mathcal{S}$ , under condition [H2](#) ( $2K_p$ ). Thus, it follows from Lemmas A.1–A.2 in the Supplementary Material [\[12\]](#) that if  $n \rightarrow \infty$ ,  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$  with  $\Delta_n = o(n^{-1/p})$ , then:

$$\sum_{i=1}^n F_{i-1}^{(1),k}(\theta^\dagger) \xrightarrow{\mathbb{P}_{\theta^\dagger}} 0, \quad 1 \leq k \leq N_\theta.$$

We consider the convergence of the term  $\sum_{i=1}^n F_{i-1}^{(2),k}(\theta^\dagger)$ . We introduce the following subsets in the event space  $\Omega$ :

$$\begin{aligned} \Omega_\varepsilon &:= \left\{ \omega \in \Omega \mid \left| \sum_{i=1}^n F_{i-1}^{(2),k}(\theta^\dagger) \right| > \varepsilon \right\}, \quad \varepsilon > 0; \\ D_{\Delta_n, i} &:= \left\{ \omega \in \Omega \mid \det \Xi_{K_p, i}(\Delta_n, \theta^\dagger) > 0 \right\}, \quad 0 \leq i \leq n-1. \end{aligned}$$

We also set  $D_{\Delta_n} = \bigcap_{i=1}^n D_{\Delta_n, i-1}$ . We then have for any  $\varepsilon > 0$ :

$$\mathbb{P}_{\theta^\dagger}(\Omega_\varepsilon) \leq \mathbb{P}_{\theta^\dagger}(\Omega_\varepsilon \cap D_{\Delta_n}) + \mathbb{P}_{\theta^\dagger}(\Omega_\varepsilon \cap D_{\Delta_n}^c). \quad (35)$$

For the first term of the right hand side of [\(35\)](#), we have the following result whose proof is provided in Section C.3 of the Supplementary Material [\[12\]](#).

**Lemma 5.11.** *Assume that conditions [H1](#), [H2](#) ( $2K_p$ ), [H3](#), [H4](#), [H5](#) and [H6](#) hold. For any  $\varepsilon > 0$ , if  $n \rightarrow \infty$ ,  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$  with  $\Delta_n = o(n^{-1/p})$ , then  $\mathbb{P}_{\theta^\dagger}(\Omega_\varepsilon \cap D_{\Delta_n}) \rightarrow 0$ .*

We next consider the second term of the right hand side of [\(35\)](#). We have:

$$\begin{aligned} \mathbb{P}_{\theta^\dagger}(\Omega_\varepsilon \cap D_{\Delta_n}^c) &\leq \sum_{i=1}^n \mathbb{P}_{\theta^\dagger}(D_{\Delta_n, i-1}^c) \\ &= \sum_{i=1}^n \mathbb{P}_{\theta^\dagger}(\{\omega \in \Omega \mid \det \Xi_{K_p, i-1}(\Delta_n, \theta^\dagger) = 0\}). \end{aligned}$$

Since it follows that:

$$\det \Xi_{K_p, i-1}(\Delta_n, \theta^\dagger) - \det \Sigma_{i-1}(\theta^\dagger) = R(\Delta_n, X_{t_{i-1}}, \theta^\dagger)$$

for  $R \in \mathcal{S}$ , and  $\det \Sigma_{i-1}(\theta^\dagger) > 0$ ,  $1 \leq i \leq n$ , under condition [H1](#)-(ii) due to Lemma [2.5](#), we have that  $\mathbb{P}_{\theta^\dagger}(\Omega_\varepsilon \cap D_{\Delta_n}^c) \rightarrow 0$  as  $\Delta_n \rightarrow 0$ . Thus, we obtain that  $\sum_{i=1}^n F_{i-1}^{(2),k}(\theta^\dagger) \xrightarrow{\mathbb{P}_{\theta^\dagger}} 0$  if  $n \rightarrow \infty$ ,  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$  with  $\Delta_n = o(n^{-1/p})$ , and the proof of convergence [\(32\)](#) is complete.

## 6. Conclusions

We have proposed general contrast estimators for a wide class of hypo-elliptic diffusions specified in ([Hypo-I](#)) and ([Hypo-II](#)), and showed that the estimators

achieve consistency and asymptotic normality in a high-frequency complete observation regime under the weakened design condition  $\Delta_n = o(n^{-1/p})$ ,  $p \geq 2$ . For elliptic diffusions, contrast estimators under such a general weak design condition have been investigated in earlier works, e.g. [16, 28], however in the context of hypo-elliptic SDEs, established results (e.g. [6, 8, 18, 13]) before our work typically relied on a condition of ‘rapidly increasing experimental design’, i.e. the requirement that  $\Delta_n = o(n^{-1/2})$ . The numerical experiments in Section 4 illustrated cases where for a given step-size  $\Delta_n$  in-between observations and given  $n$ , estimators requiring the condition  $\Delta_n = o(n^{-1/2})$  can induce unreliable biased parametric inference procedures. In contrast, the bias was removed upon consideration of estimators which required the weaker design condition  $\Delta_n = o(n^{-1/p})$ ,  $p \geq 3$ . We stress that computations w.r.t. the contrast function that delivers estimators under the weakened design condition can be automated via use of symbolic programming and automatic differentiation, so that that users need only specify the drift and diffusion coefficients of the given SDE model.

Our research opens up several potential future directions.

1. *Development and analytical study of estimators under alternative observation designs, such as partially observed coordinates and/or low-frequency observation settings.* Such designs are important from a practical perspective but have yet to be investigated analytically even for elliptic diffusions. The results in this paper are already relevant for such different observation designs as they can be regarded as providing a minimum requirement on the step-size  $\Delta_n$  (for given  $n$ ) that needs to be used when considering smaller amounts of data compared to the complete observation regime treated in this work. E.g., in a low-frequency setting where *data augmentation* procedures (within an Expectation-Maximisation or MCMC setting) will introduce latent SDE values, the user-specified step-size  $\Delta_n$  will need to satisfy the weakened design conditions obtained in our work (i.e. if the likelihood-based method is not supported in the case where imputed variables were indeed observations, there is absolutely no basis for the method that uses imputed values to provide reliable estimates). In general, it is of interest to explore the type of convergence rates and step-size conditions obtained and required, respectively, in CLTs under such different observation designs.
2. *Parametric inference for McKean-Vlasov type hypo-elliptic SDEs and related interacting particle system models.* Recently, parametric inference for SDEs with coefficients depending on the (empirical) law of the process has become an active research area in Statistics, e.g. [2, 24], though these works focus on elliptic models. Law-dependent hypo-elliptic SDEs appear in several applications [3], and it is of high interest to establish estimators and analytic results that cover such hypo-elliptic models.
3. *Non-parametric estimation for hypo-elliptic SDEs.* Established results in the literature have focused on elliptic diffusions. In the scalar case, [11] studied non-parametric estimation of drift and diffusion functions from high-frequency observations and obtained optimal convergence rates, with ex-

tensions to low-frequency observations regimes given in [9]. [1] studied a Bayesian non-parametric approach using high-frequency data. For multivariate elliptic models, see [19] for a work in a Bayesian setting and [21] which proposed a non-parametric drift estimator based on deep-learning and obtained a convergence rate under the condition of  $\Delta_n = o(n^{-1/2})$ . Providing results for hypo-elliptic models along the above lines makes up a very interesting research direction.

## Funding

Yuga Iguchi is supported by the Additional Funding Programme for Mathematical Sciences, delivered by EPSRC (EP/V521917/1) and the Heilbronn Institute for Mathematical Research.

## Supplementary Material

### Supplementary Material for “Parameter inference for hypo-elliptic diffusions under a weak design condition”

(doi: [10.1214/25-EJS2365SUPP](https://doi.org/10.1214/25-EJS2365SUPP); .pdf). The document contains proofs of the technical results introduced in Section 5 and derivation of the marginal likelihood used in the numerical experiment of Section 4.2.2.

## References

- [1] ABRAHAM, K. (2019). Nonparametric Bayesian posterior contraction rates for scalar diffusions with high-frequency data. *Bernoulli* **25** 2696–2728. [MR4003562](#)
- [2] AMORINO, C., HEIDARI, A., PILIPAIUSKAITĖ, V. and PODOLSKIJ, M. (2023). Parameter estimation of discretely observed interacting particle systems. *Stochastic Processes and their Applications* **163** 350–386. [MR4612302](#)
- [3] BALADRON, J., FASOLI, D., FAUGERAS, O. and TOUBOUL, J. (2012). Mean-field description and propagation of chaos in networks of Hodgkin-Huxley and FitzHugh-Nagumo neurons. *The Journal of Mathematical Neuroscience* **2** 1–50. [MR2974499](#)
- [4] BISCAY, R., JIMENEZ, J., RIERA, J. and VALDES, P. (1996). Local linearization method for the numerical solution of stochastic differential equations. *Annals of the Institute of Statistical Mathematics* **48** 631–644. [MR1440261](#)
- [5] BUCKWAR, E., TAMBORRINO, M. and TUBIKANEC, I. (2020). Spectral density-based and measure-preserving ABC for partially observed diffusion processes. An illustration on Hamiltonian SDEs. *Statistics and Computing* **30** 627–648. [MR4065223](#)

- [6] DITLEVSEN, S. and SAMSON, A. (2019). Hypoelliptic diffusions: filtering and inference from complete and partial observations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **81** 361–384. [MR3928146](#)
- [7] GLOTER, A. and YOSHIDA, N. (2020). Adaptive and non-adaptive estimation for degenerate diffusion processes. *arXiv preprint* [arXiv:2002.10164](#). [MR4255288](#)
- [8] GLOTER, A. and YOSHIDA, N. (2021). Adaptive estimation for degenerate diffusion processes. *Electronic Journal of Statistics* **15** 1424–1472. [MR4255288](#)
- [9] GOBET, E., HOFFMANN, M. and REISS, M. (2004). Nonparametric estimation of scalar diffusions based on low frequency data. *Annals of Statistics* **32** 2223–2253. [MR2102509](#)
- [10] HALL, P. and HEYDE, C. C. (1980). *Martingale limit theory and its application*. Academic press. [MR0624435](#)
- [11] HOFFMANN, M. (1999). Adaptive estimation in diffusion processes. *Stochastic Processes and their Applications* **79** 135–163. [MR1670522](#)
- [12] IGUCHI, Y. and BESKOS, A. (2025). Supplementary Material for “Parameter inference for hypo-elliptic diffusions under a weak design condition”. DOI: [10.1214/25-EJS2365SUPP](#).
- [13] IGUCHI, Y., BESKOS, A. and GRAHAM, M. M. (2024). Parameter inference for degenerate diffusion processes. *Stochastic Processes and their Applications* **174** 104384. [MR4750236](#)
- [14] IGUCHI, Y., BESKOS, A. and GRAHAM, M. M. (2025). Parameter estimation with increased precision for elliptic and hypo-elliptic diffusions. *Bernoulli* **31** 333–358. [MR4815987](#)
- [15] KALLIADASIS, S., KRUMSCHEID, S. and PAVLIOTIS, G. A. (2015). A new framework for extracting coarse-grained models from time series with multiscale structure. *Journal of Computational Physics* **296** 314–328. [MR3348504](#)
- [16] KESSLER, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scandinavian Journal of Statistics* **24** 211–229. [MR1455868](#)
- [17] LEIMKUHLE, B. and MATTHEWS, C. (2015). Molecular Dynamics. *Interdisciplinary applied mathematics* **39** 443. [MR3362507](#)
- [18] MELNYKOVA, A. (2020). Parametric inference for hypoelliptic ergodic diffusions with full observations. *Statistical Inference for Stochastic Processes* **23** 595–635. [MR4136706](#)
- [19] NICKL, R. and RAY, K. (2020). Nonparametric statistical inference for drift vector fields of multi-dimensional diffusions. *The Annals of Statistics* **48** 1383–1408. [MR4124327](#)
- [20] NUALART, D. (2006). *The Malliavin Calculus and Related Topics* **1995**. Springer. [MR1344217](#)
- [21] OGA, A. and KOIKE, Y. (2024). Drift estimation for a multi-dimensional diffusion process using deep neural networks. *Stochastic Processes and their Applications* **170** 104240. [MR4682669](#)
- [22] POKERN, Y., STUART, A. M. and WIBERG, P. (2009). Parameter estima-



- tion for partially observed hypoelliptic diffusions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **71** 49–73. [MR2655523](#)
- [23] PRAKASA-RAO, B. (1988). Statistical inference from sampled data for stochastic processes. *Contemporary Mathematics* **80** 249–284. [MR0999016](#)
- [24] SHARROCK, L., KANTAS, N., PARPAS, P. and PAVLIOTIS, G. A. (2023). Online parameter estimation for the McKean–Vlasov stochastic differential equation. *Stochastic Processes and their Applications* **162** 481–546. [MR4597535](#)
- [25] SPANNAUS, A., PAPAMARKOU, T., ERWIN, S. and CHRISTIAN, J. B. (2022). Inferring the spread of COVID-19: the role of time-varying reporting rate in epidemiological modelling. *Scientific Reports* **12** 1–12.
- [26] STRAMER, O. and ROBERTS, G. (2007). On Bayesian analysis of nonlinear continuous-time autoregression models. *Journal of Time Series Analysis* **28** 744–762. [MR2395912](#)
- [27] TSAI, H. and CHAN, K. (2000). Testing for nonlinearity with partially observed time series. *Biometrika* **87** 805–821. [MR1813976](#)
- [28] UCHIDA, M. and YOSHIDA, N. (2012). Adaptive estimation of an ergodic diffusion process based on sampled data. *Stochastic Processes and their Applications* **122** 2885–2924. [MR2931346](#)
- [29] VROYLANDT, H., GOUDENÈGE, L., MONMARCHÉ, P., PIETRUCCI, F. and ROTENBERG, B. (2022). Likelihood-based non-Markovian models from molecular dynamics. *Proceedings of the National Academy of Sciences* **119** e2117586119.