



HAL
open science

Optimal control for parameter estimation in partially observed hypoelliptic stochastic differential equations

Quentin Clairon, Adeline Samson

► **To cite this version:**

Quentin Clairon, Adeline Samson. Optimal control for parameter estimation in partially observed hypoelliptic stochastic differential equations. *Computational Statistics*, 2022, 37, pp.2471-2491. 10.1007/s00180-022-01212-9 . hal-03320139v2

HAL Id: hal-03320139

<https://inria.hal.science/hal-03320139v2>

Submitted on 9 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Optimal control for parameter estimation in partially observed hypoelliptic stochastic differential equations

Quentin Clairon¹ · Adeline Samson²

Received: 7 September 2021 / Accepted: 10 February 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

We deal with the problem of parameter estimation in stochastic differential equations (SDEs) in a partially observed framework. We aim to design a method working for both elliptic and hypoelliptic SDEs, the latter being characterized by degenerate diffusion coefficients. This feature often causes the failure of contrast estimator based on Euler Maruyama discretization scheme and dramatically impairs classic stochastic filtering methods used to reconstruct the unobserved states. All of these issues make the estimation problem in hypoelliptic SDEs difficult to solve. To overcome this, we construct a well-defined cost function no matter the elliptic nature of the SDEs. We also bypass the filtering step by considering a control theory perspective. The unobserved states are estimated by solving deterministic optimal control problems using numerical methods which do not need strong assumptions on the diffusion coefficient conditioning. Numerical simulations made on different partially observed hypoelliptic SDEs reveal our method produces accurate estimate while dramatically reducing the computational price comparing to other estimation procedures.

Keywords Stochastic differential equations · Parameter estimation · Hypoellipticity · Optimal control theory

1 Introduction

Stochastic differential equations (SDEs) are widely used in several fields to model stochastic temporal dynamics. We focus on hypoelliptic SDE, that is when the diffusion matrix is not of full rank but its solution has a smooth density. More precisely, we consider a process $Z_t = (V_t, U_t) \in \mathbb{R}^d$, of dimension d on a time interval $[0, T]$,

✉ Quentin Clairon
quentin.clairon@u-bordeaux.fr

¹ Inria Bordeaux Sud-Ouest, Inserm, Bordeaux Population Health Research Center, SISTM Team, UMR1219, University of Bordeaux, 33000 Bordeaux, France

² CNRS, Inria, Grenoble Institute of Engineering, LJK, Univ. Grenoble Alpes, 38000 Grenoble, France

defined as the solution of the following hypoelliptic SDE:

$$\begin{aligned}dV_t &= g_\theta(V_t, U_t, t)dt \\dU_t &= h_\theta(V_t, U_t, t)dt + B_\sigma(Z_t, t)dW_t \\Z(0) &= Z_0\end{aligned}\tag{1}$$

where V_t is the d_V -dimensional component corresponding to the smooth part of the process, that is not directly affected by stochastic perturbations, and U_t is the d_U -dimensional rough part, W_t is a d_U dimensional Brownian motion acting on the system through the d_U squared diffusion matrix $B_\sigma(Z_t, t)$. In some applications, only the smooth component can be measured.

Let us give some examples of such models. Stochastic damping Hamiltonian systems or Langevin equations have been developed to describe a particle moving in an environment defined by a potential (Wu 2001). The system is usually two dimensional, the first equation describes the position of the particle and the other its velocity. The noise is degenerate because it acts only on the velocity and not on the position. Only the position is directly measured. We can cite applications of these models in molecular dynamics (Leimkuhler and Matthews 2015), paleoclimate research (Ditlevsen et al. 2002), neural field models (Coombes and Byrne 2019; Ditlevsen and Löcherbach 2017). Another application is neuronal models of membrane potential dynamics. Examples are the hypoelliptic FitzHugh-Nagumo (FHN) model (Gerstner and Kistler 2002; DeVille et al. 2005; Leon and Samson 2018), the hypoelliptic Hodgkin-Huxley model (Goldwyn and Shea-Brown 2011; Tuckwell and Ditlevsen 2016), or synaptic-conductance based models with stochastic channel dynamics (Paninski et al. 2012; Ditlevsen and Greenwood 2013). Only the first equation, corresponding to the membrane potential, can be measured by intra-cellular recordings. It is therefore important to develop estimation methods for this class of models.

In the rest of the paper, we assume the drift functions g_θ and h_θ depend on a d_θ dimensional parameter θ and the matrix B_σ on the d_σ dimensional parameter σ known as volatility. Partial observations of (Z_t) are defined as:

$$Y_t = CZ_t$$

where C is the $d_0 \times d$ observation matrix. We assume (Y_t) is discretely observed on the interval $[0, T]$ at times $0 = t_0 < \dots < t_n = T$ and without measurement error. Our aim is to estimate the unknown parameter $\psi = (\theta, \sigma)$ of model (1) using the discrete and partial observations (Y_0, \dots, Y_n) and possibly without knowing the initial condition Z_0 .

Parametric estimation of hypoelliptic models faces several difficulties. When the first equation reduces to $g_\theta(V_t, U_t) = U_t$, different contrasts estimators have been proposed (Genon-Catalot et al. 2000; Gloter 2000; Ditlevsen and Sørensen 2004; Gloter 2006; Samson and Thieullen 2012; Lu et al. 2016; Leon et al. 2019). This specific case allows to deal with partial observations, the second coordinate U_t being replaced by increments of V_t . When the drift g_θ is more complex, we deal with several entangled issues. First, solutions of the SDE are generally non explicit. Numerical schemes are used to approximate the discretized process and derive an estimation

criteria. However, compared to elliptic system, the degeneracy of the noise complicates the statistical estimation. Thus estimation cost function that would be derived directly from the Euler-Maruyama fails because the covariance matrix $\Gamma_\sigma^T \Gamma_\sigma$ where $\Gamma_\sigma(Z_t, t) = (0_{d_V, d_U}, B_\sigma(Z_t, t))^T$, is not full rank. Gloter and Yoshida (2021) propose a local Gaussian approximation of the transition density in the case of totally observed coordinates. Several papers propose to use a likelihood based on higher order schemes of approximation, the 1.5 order scheme (Ditlevsen and Samson 2019), the local linearization scheme (Melnykova 2020). In the case of partial observations, the unobserved coordinates U_t has to be filtered or imputed, either in a Bayesian settings with a Markov Chain Monte Carlo algorithm (Pokern et al. 2009; Graham et al. 2019) or in a frequentist settings with a particle filter or sequential Monte Carlo coupled with a stochastic approximation expectation maximization (SAEM) algorithm (Ditlevsen and Samson 2014, 2019). A recent simulation filter has been proposed by Bierkens et al. (2020) for hypoelliptic diffusions. Still the theoretical assessment and practical efficiency of such numerical methods critically rely on Γ_σ being well conditioned. These methods are also time consuming.

In this work, we propose an estimation method based on optimal control theory which applies to nonlinear hypoelliptic systems and partial observations. The idea of using optimal control theory is to treat the coordinate filtering problem as a deterministic tracking problem. Then, numerical devices coming from control theory allow us to estimate the *optimal control* bringing the observed system states the closest possible to the observations. This was proposed for ODE system (Clairon and Brunel 2019; Clairon 2021; Iolov et al. 2017). For SDE, the only reference is Clairon and Samson (2020) which uses linear-quadratic theory, a particular case of the Pontryagin maximum principle, to solve the tracking problem. But this limits the application of the method to linear systems. Also, this method separately estimates σ and θ via a nested procedure because of identifiability issues. Volatility σ was estimated via the minimization of an outer criteria demanding repeated estimation of θ based on the optimization of an inner criteria. This dramatically increases the computation time of the method when σ is unknown. Additionally, the criteria used for σ inference was properly defined only for linear SDEs.

In this work, we construct an estimation procedure which simultaneously estimate (θ, σ) by relying on a contrast function bypassing identifiability issues for σ estimation encountered in Clairon and Samson (2020). This contrast function is well defined for linear and nonlinear SDEs, both elliptics and hypoelliptics. Also, the control problem solutions required by our approach are easily obtained for non-linear SDEs via an adaptation of the method proposed by Cimen and Banks (2004a).

Our statistical criteria exploits the hypoelliptic nature of (1) which assumes “enough” interactions between V_t and U_t such that the Wiener process W_t influences all the coordinates of the SDE, potentially with some delay. More precisely, we consider a lagged contrast function linking W_t and the observed part of the system state CZ_{t+m_B} after a delay $m_B > 0$ such that the dependence CZ_{t+m_B} with respect to W_t is regular enough to ensure a well-conditioned variance for CZ_{t+m_B} . This requires to know explicitly the dependence of CZ_{t+m_B} to W_t . For this we rely on the so-called **pseudo-linear** representation of the discretized SDE (1) (Cimen 2008). The scheme is

expressed at each time t as a function of the m_B previous lagged states. The discretized process is no more Markovian, this is a direct consequence of hypoellipticity. From this lagged formulation of the discretized process, we deduce a lagged contrast which is the pseudo-likelihood of the representation. This statistical contrast depends on the whole state variable, even the unobserved ones. We thus need to filter or predict the unobserved coordinate U_t . We propose a predictor which is a balance between fidelity to the observations and to the solution of SDE (1). This predictor is defined via an optimal control problem that can be solved easily without assumption on Γ_σ rank. The optimal control problem is solved with the linear quadratic theory for linear SDEs and for nonlinear models with an adaptation of the procedure developed by Cimen and Banks (2004b) based on the pseudo-linear representation. Our criteria is somewhat similar to the generalized profiling introduced in Ramsay et al. (2007) for ordinary differential equations but here dedicated to SDEs with potentially degenerate diffusions. We also consider a criteria making a tradeoff between data and model fidelity to regularize the inverse problem of parameter estimation. This method provides estimators for the parameters and for the unobserved coordinate U_t . The simulation study illustrates the good performances of our method.

Section 2 studies the hypoellipticity of the continuous process, provides a discrete approximation scheme and the corresponding concept of hypoellipticity. In Sect. 3, we derive the expression of our statistical criteria which defines our estimator. The construction of the state predictor and the optimal control numerical methods are presented in Sect. 4. Then, in Sect. 5 we conduct a simulation study on three hypoelliptic models to evaluate the practical accuracy of our method as well as its computational efficiency.

2 Model and its discrete approximation

This section describes the properties of hypoelliptic systems and the connexity property that allows to propagate the noise to the smooth coordinates. Then we introduce the approximate Euler-Maruyama scheme and the lagged pseudo-linear formulation, the key of our estimation method.

2.1 Models and its assumptions

Let us denote:

$$f_\theta(Z_t, t) = \begin{pmatrix} g_\theta(V_t, U_t, t) \\ h_\theta(V_t, U_t, t) \end{pmatrix}, \quad \Gamma_\sigma(Z_t, t) = \begin{pmatrix} 0_{d_V, d_U} \\ B_\sigma(Z_t, t) \end{pmatrix}$$

the drift and diffusion coefficient of the diffusion process Z_t respectively, then

$$dZ_t = f_\theta(Z_t, t)dt + \Gamma_\sigma(Z_t, t)dW_t.$$

We assume B_σ is full rank and we consider two classes of models called elliptic and hypoelliptic.

Elliptic models are defined by $d_V = 0$. In that case, the Brownian process directly acts on the whole system, the diffusion coefficient is not degenerated. Girsanov formula applies and the process has a continuous smooth density.

Hypoelliptic models are defined with $d_V > 0$ and verify the Hörmander condition detailed below. Let us first define the Lie brackets. For f a function with values in \mathbb{R}^d , f_l stands for its l -th component. For a matrix Γ , Γ_j denotes its j -th column.

Definition 1 The Lie bracket of two functions $f, g : \mathbb{R}^d \rightarrow \mathbb{R}$ is defined as

$$[f, g]_l = \frac{\partial g_l(z)}{\partial z} f(z) - \frac{\partial f_l(z)}{\partial z} g(z).$$

Hypoellipticity is then defined as follows:

Definition 2 (Hörmander condition) Let us consider the set \mathcal{L} defined iteratively as:

- Initialization step: \mathcal{L} is composed of vectors $L^0 = \Gamma_j$, for $j = 1, \dots, d_U$.
- Generalization step at iteration k , if $L^{k-1} \in \mathcal{L}$ then vectors L^k defined by $L^k = [f(Z_t, t), L^{k-1}]$ and by $L^k = [\Gamma_j(Z_t, t), L^{k-1}]$ for $j = 1, \dots, d_U$ belong to \mathcal{L} .

If at some iteration, \mathcal{L} spans \mathbb{R}^d , the weak Hörmander condition is fulfilled. This implies that the system is hypoelliptic.

2.2 Hypoellipticity properties

We derive a necessary condition on the drift g_θ respected by hypoelliptic systems (the proof is given in Appendix 1 in Supplementary Materials):

Proposition 1 (Connexity property) *If the SDE (1) is hypoelliptic then for each $l \in \llbracket 1, d_V \rrbracket$, it exists $j \in \llbracket 1, d_U \rrbracket$ and a sequence $(q_1, \dots, q_{m_l}) \in \llbracket 1, d_V \rrbracket^{m_l}$ with $m_l \leq d_V$ such that*

$$\frac{\partial g_{\theta,l}(Z_t, t)}{\partial V_{q_1}} \frac{\partial g_{\theta,q_1}(Z_t, t)}{\partial V_{q_2}} \dots \frac{\partial g_{\theta,q_{m_l-1}}(Z_t, t)}{\partial V_{q_{m_l}}} \frac{\partial g_{\theta,q_{m_l}}(Z_t, t)}{\partial U_j} \neq 0. \tag{2}$$

This can be seen as a connexity property between the set of smooth variables and rough ones. It means that each variable $V_{l,t}$ is influenced by at least one component of U_t at least undirectly through a path of auxiliary smooth variables $(V_{q_1}, \dots, V_{q_{m_l}})$. Of course, from a given sequence (q_1, \dots, q_{m_l}) , it is possible to construct bigger ones. To remove ambiguity m_l denotes the length of the shortest possible path such that the connexity property (2) holds.

Although the diffusion coefficient $\Gamma_\sigma(Z_t, t)$ is rank deficient, the solution of (1) has a smooth density with respect to the Lebesgue measure.

If the drift $f(Z_t)$ satisfies a dissipativity condition

$$\langle f(z), z \rangle \leq \alpha - \delta \|z\|^2, \forall z \in \mathbb{R}^d$$

where $\alpha, \delta > 0$, the function $L(z) = 1 + \|z\|^2$ is a Lyapounov function. It ensures the geometric ergodicity of the solution of the SDE. It means that $(Z(t))_{t \in [0, T]}$ converges exponentially fast to a unique invariant distribution.

However, usually, the solution of SDE (1) is not explicit and a numerical scheme is needed to approximate the solution.

2.3 Euler–Maruyama discretization

Let us introduce the Euler-Maruyama discretized process $(\tilde{Z}_i, \text{ for } i = 1, \dots, n)$ of SDE (1). To simplify the notations, we assume that the time points are equidistant, that is $\Delta_i = t_{i+1} - t_i = \Delta, \forall i = 1, \dots, n$. Then the discretized approximate process is defined by:

$$\begin{aligned} \tilde{Z}_{i+1} &= \tilde{Z}_i + \Delta f_\theta(\tilde{Z}_i, t_i) + \sqrt{\Delta} \Gamma_\sigma(\tilde{Z}_i, t_i) u_i \\ \tilde{Z}_0 &= Z_0 \end{aligned} \tag{3}$$

where $u_i \sim N(0, I_{d_U})$ is the normalized increment of the Brownian motion at t_i , i.e. $u_i = \frac{1}{\sqrt{\Delta}} (W_{t_{i+1}} - W_{t_i})$ and $u = (u_0, \dots, u_{n-1})$.

The connexity property can be illustrated on this Euler-Maruyama discretization. Proposition 2 proves that $m_l + 1$ iterations of (3) are required to have $\tilde{V}_{l, i+m_l+1}$ affected by some entry $\tilde{U}_{j, i}$ of the rough component.

Proposition 2 *Let us assume that for each $l \in \llbracket 1, d_V \rrbracket$ it exists a unique minimal length sequence (q_1, \dots, q_{m_l}) such that the connexity property holds. Then the Euler-Maruyama approximation $(\tilde{Z}_i, \text{ for } i = 1, \dots, n)$ is such that for each $l \in \llbracket 1, d_V \rrbracket$ it exists $j \in \llbracket 1, d_U \rrbracket$ such that:*

$$\frac{\partial \tilde{V}_{l, i+k}}{\partial \tilde{U}_{j, i}} = 0, \text{ for } k = 1, \dots, m_l \text{ and } \frac{\partial \tilde{V}_{l, i+m_l+1}}{\partial \tilde{U}_{j, i}} \neq 0$$

So, in model (3), each coordinate of \tilde{V}_{i+m_B+1} with $m_B := \max_l m_l$ is influenced by \tilde{U}_i . It also means that $m_B + 1$ is the minimal delay to propagate the noise u_i from the rough components to all the smooth ones in the discretized model (3).

2.4 Lagged pseudo-linear formulation of the discretized process

Proposition 2 proves the existence of a minimal delay m_B . For estimation purpose, we need the explicit link between \tilde{Z}_{i+m_B+1} and u_i . For this, we rely on what we call later the *lagged pseudo-linear formulation* of (3).

Let us introduce the matrix $A_\theta(\tilde{Z}_i, t_i)$ and the vector $r_\theta(t_i)$ such that $A_\theta(\tilde{Z}_i, t_i) \tilde{Z}_i + r_\theta(t_i) = f_\theta(\tilde{Z}_i, t_i)$. This allows to reformulate (3) with a pseudo-linear expression:

$$\begin{cases} \tilde{Z}_{i+m_B+1} = \mathbf{A}_{\theta, i+m_B} \tilde{Z}_{i+m_B} + \Delta r_\theta(t_{i+m_B}) \\ \quad \quad \quad + \sqrt{\Delta} \Gamma_\sigma(\tilde{Z}_{i+m_B}, t_{i+m_B}) u_{i+m_B} \\ \tilde{Z}_0 = Z_0 \end{cases} \tag{4}$$

with $\mathbf{A}_{\theta,i} := I_d + \Delta A_{\theta}(\tilde{Z}_i, t_i)$. We can replace \tilde{Z}_{i+m_B} by its state-space expression (3) and iterate to obtain the lagged pseudo-linear formulation.

Proposition 3 *The lagged pseudo-linear formulation of the discretized process (3) is defined as follows. For $i = 0, \dots, n - m_B - 1$:*

$$\begin{aligned} \tilde{Z}_{i+m_B+1} &= \prod_{r=0}^{m_B} \mathbf{A}_{\theta,i+r} \tilde{Z}_i \\ &+ \sum_{r=0}^{m_B} \left(\prod_{l=r+1}^{m_B} \mathbf{A}_{\theta,i+l} \right) \Delta r_{\theta}(t_{i+r}) \\ &+ \sum_{r=0}^{m_B} \left(\prod_{l=r+1}^{m_B} \mathbf{A}_{\theta,i+l} \right) \sqrt{\Delta} \Gamma_{\sigma}(\tilde{Z}_{i+r}, t_{i+r}) u_{i+r}. \end{aligned} \tag{5}$$

In the last expression the multiplication symbol $\prod_{l=0}^{m_B-1}$ denotes the sequential left multiplication i.e. $\prod_{j=a}^b \mathbf{A}_{\theta,j} = \mathbf{A}_{\theta,b} \times \mathbf{A}_{\theta,b-1} \times \dots \times \mathbf{A}_{\theta,a}$ with the convention $\prod_{j=b+1}^b \mathbf{A}_{\theta,j} = I_d$.

The pseudo-linear formulation gives the explicit link between \tilde{Z}_{i+m_B+1} and u_i . The Markovian property is no more fulfilled for the discretized process: the smooth density is obtained only if we express \tilde{Z}_{i+m_B+1} as a function of the m_B previous states. This is a direct consequence of hypoellipticity.

Let us introduce a new observability assumption that reinforces hypoellipticity for a discrete process. It requires that the observed variables give enough information about the diffusion process, or equivalently that the Y_t are influenced by all the rough components.

Assumption 1 (H1): The $d_o \times d_U$ matrix

$$C \left(\prod_{l=1}^{m_B} \mathbf{A}_{\theta,i+l} \right) \Gamma_{\sigma}(z_i, t_i)$$

is full rank for all $1 \leq i \leq n - m_B$ and for all possible values $\{z_i\}_{0 \leq l \leq m_B} \in \mathbb{R}^{d \times m_B}$.

Let us detail this assumption for elliptic and some hypoelliptic models.

- Elliptic models verify $m_B = 0$. Assumption (H1) imposes that $C \Gamma_{\sigma}(\tilde{Z}_i, t_i)$ is full rank. It ensures a smooth transition density for the discrete process (Y_i) . This is similar to the continuous process.
- We focus on hypoelliptic SDE with $\frac{\partial g_{\theta}}{\partial V}(V_t, U_t, t)$ full rank and $Y_t = V_t$. This corresponds to the case $m_B = 1$, that is all the components of V_t have a direct interaction with at least one component of U_t and stochasticity only acts on the unobserved variables. This corresponds to the 1-step hypoellipticity defined by Buckwar et al. (2021).

3 Cost function construction

Our aim is to estimate the unknown parameter $\psi = (\theta, \sigma)$ of model (1) using the discrete and partial observations (Y_0, \dots, Y_n) and possibly without knowing the initial condition Z_0 .

We first assume that the process Z is observed and construct an estimation contrast based on the lagged pseudo-linear representation. Then, we propose to filter Z by optimal control.

3.1 Statistical criteria assuming Z observed

Let us consider in this section that the variables Z are observed. We can use formula (5) to build a statistical criteria. Let us denote $t_{i+m_B}^i = (t_i, \dots, t_{i+m_B})$, $\tilde{Z}_{i:i+m_B} = (\tilde{Z}_i, \dots, \tilde{Z}_{i+m_B})$ and define

$$\begin{aligned} X_{i:i+m_B+1} &= Y_{i+m_B+1} - C \left(\prod_{r=0}^{m_B} \mathbf{A}_{\theta,i+r} \right) \tilde{Z}_i - \sum_{r=0}^{m_B} F_{i+r:i+m_B} \\ F_{i+r:i+m_B} &= \Delta C \left(\prod_{l=r+1}^{m_B} \mathbf{A}_{\theta,i+l} \right) r_{\theta}(t_{i+r}) \\ G_{i+r:i+m_B} &= \sqrt{\Delta} C \left(\prod_{l=r+1}^{m_B} \mathbf{A}_{\theta,i+l} \right) \Gamma_{\sigma}(\tilde{Z}_{i+r}, t_{i+r}). \end{aligned}$$

Then we have

$$X_{i:i+m_B} = \sum_{r=0}^{m_B} G_{i+r:i+m_B} u_{i+r}. \tag{6}$$

Under assumption (H1), the matrix

$$\Sigma_{i:i+m_B} = \sum_{r=0}^{m_B} G_{i+r:i+m_B} G_{i+r:i+m_B}^T$$

is nonsingular. Then, conditionally on $\tilde{Z}_{i:i+m_B}$, we have

$$X_{i:i+m_B+1} \mid \tilde{Z}_{i:i+m_B} \sim N(0_{d_o,1}, \Sigma_{i:i+m_B})$$

and we can define the following cost function based on the log likelihood:

$$H^{m_B}(\psi \mid Y, Z) = -\log \left(\mathbb{P}_{\psi} \left\{ X_{i:i+m_B+1} \mid \tilde{Z}_{i:i+m_B} \right\}_{i \in \llbracket 1, n-m_B-1 \rrbracket} \right) \tag{7}$$

and the associated estimator:

$$\hat{\psi} = \arg \min_{\psi} H^{m_B}(\psi \mid Y, Z). \tag{8}$$

From equation (6), we see that $X_{i:i+m_B+1}$ depends at most of (u_i, \dots, u_{i+m_B}) . Thus the sequence $(X_{i:i+m_B+1})$ for varying i can be composed of dependant terms which

makes H^{m_B} not easy to calculate in this general setting. We detail in the next subsection some cases where H^{m_B} has an explicit form.

3.2 Simplification of H^{m_B} for some models

Note first that when the sequence of $(X_{i:i+m_B+1})$ is composed of independent terms, the expression of H^{m_B} simplifies to

$$H^{m_B}(\psi | Y, Z) = \sum_{i=1}^{n-m_B-1} \left(X_{i:i+m_B+1}^T \Sigma_{i:i+m_B}^{-1} X_{i:i+m_B+1} + \log(\text{Det}(\Sigma_{i:i+m_B})) \right).$$

This is the case for elliptic and 1-step hypoelliptic systems. Let us give more detailed expressions for $\Sigma_{i:i+m_B}$ for these two cases.

Elliptic system Here, we have $m_B = 0$:

$$X_{i:i+1} = Y_{i+1} - C \left(\mathbf{A}_{\theta,i} \tilde{Z}_i + \Delta r_{\theta}(t_i) \right)$$

and $X_{i:i+1} | \tilde{Z}_i \sim N(0_{d_o,1}, \Sigma_{i:i})$ with $\Sigma_{i:i} = \Delta C \Gamma_{\sigma}(\tilde{Z}_i, t_i) \Gamma_{\sigma}(\tilde{Z}_i, t_i)^T C^T$. Thus, we fall back to the classic contrast estimator of elliptic SDE.

1-step hypoelliptic system In the case where $m_B = 1$ and $V_t = CZ_t$, we end up with $C \Gamma_{\sigma}(\tilde{Z}_i, t_i) = 0$. So condition (H1) is respected if $CA_{\theta}(\tilde{Z}_{i+1}, t_{i+1}) \Gamma_{\sigma}(\tilde{Z}_i, t_i)$ is of full rank. Then $X_{i:i+m_B+1}$ is given by:

$$X_{i:i+2} = Y_{i+2} - C \mathbf{A}_{\theta,i+1} \mathbf{A}_{\theta,i} \tilde{Z}_i - C \left(\Delta \mathbf{A}_{\theta,i+1} r_{\theta}(t_i) + \Delta r_{\theta}(t_{i+1}) \right)$$

and follows the Gaussian law $X_{i:i+2} | \tilde{Z}_{i:i+1} \sim N(0_{d_o,1}, \Sigma_{i:i+1})$ with:

$$\Sigma_{i:i+1} = \Delta C A_{\theta}(\tilde{Z}_{i+1}, t_{i+1}) \Gamma_{\sigma}(\tilde{Z}_i, t_i) \Gamma_{\sigma}(\tilde{Z}_i, t_i)^T A_{\theta}(\tilde{Z}_{i+1}, t_{i+1}) C^T.$$

4 State predictor $\bar{Z}_{\theta,\sigma}$

Now, we propose a state predictor of the hidden process Z using optimal control theory and then we plug it in the statistical criteria.

4.1 Formalization of the related optimal control problem

In this section, we denote $(\tilde{Z}(Z_0, u))$ the solution of (4) for a given sequence $u = (u_0, \dots, u_{n-1})$ and a given initial condition Z_0 . Similarly to trajectory fitting estimators (Kutoyants 1991; Dietz 2001), a first predictor could be the solution \bar{Z} of (4) which is the closest to the observations. This solution corresponds to the sequence

$u_{Z_0}^M$ such that:

$$u_{Z_0}^M = \arg \min_u \left\{ \sum_{i=0}^n \left\| C \tilde{Z}_i(Z_0, u) - Y_i \right\|_2^2 \right\} \tag{9}$$

when the initial condition Z_0 is fixed. However, this optimization problem is ill-posed. The solution is not necessarily unique and would not always be continuous as a function of the parameters and observations.

We thus propose to introduce a penalization term into the optimization problem which will lead to a Tikhonov regularized version of it. This is known to remove sources of ill-posedness (Engl et al. 2009), in particular it re-establishes uniqueness and the continuity of the solution for linear SDEs. The penalized problem is the following:

$$\bar{u}_{Z_0} := \arg \min_u \left\{ \sum_{i=0}^n \left\| C \tilde{Z}_i(Z_0, u) - Y_i \right\|_2^2 - \frac{2}{w} \log P(u) \right\}$$

where $P(u)$ is the density of the increment u . The term $-\frac{2}{w} \log P(u)$ penalizes the sequences which are unlikely to be a realization of the Brownian motion and $w > 0$ makes the balance between the model and data fidelity. We have

$$\log P(u) = \log \prod_{i=0}^{n-1} P(u_i) \propto \log \prod_{i=0}^{n-1} e^{-\frac{1}{2}u_i^T u_i} = - \sum_{i=0}^{n-1} \frac{1}{2} u_i^T u_i$$

thus

$$\bar{u}_{Z_0} = \arg \min_u \left\{ \sum_{j=1}^{n-1} \left(\left\| C \tilde{Z}_j(Z_0, u) - Y_j \right\|_2^2 + \frac{1}{w} u_j^T u_j \right) + \left\| C \tilde{Z}_n(Z_0, u) - Y_n \right\|_2^2 \right\}.$$

The predictor corresponds to the solution \bar{u}_{Z_0} of the following deterministic discrete optimal control problem:

$$\begin{aligned} \text{Minimize} \quad & C_w(u \mid Y; Z_0) = \sum_{j=1}^{n-1} \left(\left\| C \tilde{Z}_j(Z_0, u) - Y_j \right\|_2^2 + \frac{1}{w} u_j^T u_j \right) \\ & + \left\| C \tilde{Z}_n(Z_0, u) - Y_n \right\|_2^2 \\ \text{Subject to:} \quad & \begin{cases} \tilde{Z}_{i+1}(Z_0, u) = \mathbf{A}_{\theta, i} \tilde{Z}_i(Z_0, u) + \Delta r_{\theta}(t_i) \\ \quad \quad \quad + \sqrt{\Delta} \Gamma_{\sigma}(\tilde{Z}_i(Z_0, u), t_i) u_i \\ \tilde{Z}_0 = Z_0 \end{cases} \end{aligned} \tag{10}$$

Thus, if we are able to solve problem (10) to derive \bar{u}_{Z_0} , we can solve (4) to obtain the related state predictor $\bar{Z}_{Z_0} := \tilde{Z}(Z_0, \bar{u}_{Z_0})$. However, our optimal control still depends

on Z_0 which is potentially unknown. If required, its estimation is bypassed by profiling the cost C_w and defining as initial condition estimator:

$$\widehat{Z}_0 = \arg \min_{Z_0} \left\{ \min_u C_w(u \mid Y; Z_0) \right\}.$$

Let us now denote $\bar{u} := \bar{u}_{\widehat{Z}_0}$ the corresponding optimal control and $\bar{Z} := \bar{Z}_{\widehat{Z}_0}$ the related state-space predictor.

We propose to solve (10) with numerical control theory methods. For linear models, (10) is a Linear-Quadratic problem. The control theory then ensures:

- existence and uniqueness of the solutions \bar{u} and \bar{Z} , which are continuous functions of (θ, σ) and Y , under mild regularity hypothesis on A_θ and Γ_σ (in particular, Γ_σ is not required to be of full rank),
- computation of \bar{u} , \bar{Z} and, if required, \widehat{Z}_0 by solving a finite difference equation.

For nonlinear models, we use an adaptation of Cimen and Banks (2004b) which substitutes the original problem with a finite sequence of Linear-Quadratic ones to benefit of the points listed above. The linear and nonlinear cases are detailed in the next subsection.

4.2 Numerical methods for solving (10)

We have defined our state predictor as solution of a deterministic tracking problem. Several algorithms or tools have been developed to solve this kind of control problem. In this section, θ is fixed at a given value. Then to simplify the notations, we omit it.

4.2.1 Linear models

We consider the case where $\mathbf{A}_i := \mathbf{A}_{\theta,i} = I_d + \Delta A_\theta(\tilde{Z}_i, t_i) := I_d + \Delta A_\theta(t_i)$ and $\Gamma(t_i) := \Gamma_\sigma(\tilde{Z}_i, t_i)$ for $i = 1, \dots, n$. In this framework \bar{u}_{Z_0} is derived by solving a finite difference equation known as Riccati equation defined by

$$\begin{aligned} E_n &= C^T C \\ h_n &= -C^T Y_n \end{aligned}$$

and for $i = n - 1, \dots, 1$:

$$\begin{aligned} E_i &= \mathbf{A}_i^T E_{i+1} \mathbf{A}_i + C^T C - \Delta \mathbf{A}_i^T E_{i+1} \Gamma(t_i) G(E_{i+1}) \Gamma(t_i)^T E_{i+1} \mathbf{A}_i \\ h_i &= \Delta_i \mathbf{A}_i^T E_{i+1} r(t_i) + \mathbf{A}_i^T h_{i+1} - C^T Y_i \\ &\quad - \Delta \mathbf{A}_i^T E_{i+1} \Gamma(t_i) G(E_{i+1}) \Gamma(t_i)^T (h_{i+1} + \Delta_i E_{i+1} r(t_i)) \end{aligned} \tag{11}$$

with $G(E_{i+1}) = \left[\frac{1}{w} \times I_d + \Delta \Gamma(t_i)^T E_{i+1} \Gamma(t_i) \right]^{-1}$. Then $\bar{u}_{Z_0,i}$ is given by:

$$\bar{u}_{Z_0,i} = -\sqrt{\Delta} G(R_{k+1}) \Gamma(t_i)^T \left(E_{i+1} (\mathbf{A}_i \bar{Z}_i + \Delta_i r(t_i)) + h_{i+1} \right) \tag{12}$$

and $\bar{u}_{Z_0,i}$ is used in (1) to obtain \bar{Z}_{Z_0} . Moreover, the initial condition estimator \widehat{Z}_0 is given by $\widehat{Z}_0 = -(E_0)^{-1} h_0$. Computational details are given in Appendix 2.

4.2.2 Nonlinear models

For non-linear model, we propose to apply the previous algorithm thanks to the pseudo-linear representation. We replace the original problem by a sequence of Linear-Quadratic control problems solved iteratively until a convergence criteria is verified. This method is an adaptation of Cimen and Banks (2004b) to the case of discrete models. We propose the following algorithm to solve (10).

1. Initialisation $\forall i \in \llbracket 0, n \rrbracket, \bar{Z}_i^0 = Z_i^l$ where Z_i^l is an arbitrary starting point if Z_0 is unknown or $\bar{Z}_0^0 = Z_0$ otherwise.
2. At iteration l , compute (\bar{Z}^l, \bar{u}^l) by solving the Linear-Quadratic optimal control problem derived from the pseudo-linear representation of the SDE:

$$\begin{aligned} \text{Minimize} \quad & C_w^l(u \mid Y; Z_0) = \sum_{j=1}^{n-1} \left(\left\| C \tilde{Z}_j(Z_0, u) - Y_j \right\|_2^2 + \frac{1}{w} u_j^T u_j \right) \\ & + \left\| C \tilde{Z}_n(Z_0, u) - Y_n \right\|_2^2 \\ \text{Subject to: } \quad & \begin{cases} \tilde{Z}_{i+1}(Z_0, u) = \mathbf{A}_{\theta,i}^l \tilde{Z}_i(Z_0, u) + \Delta r_\theta(t_i) + \sqrt{\Delta} \Gamma_\sigma^l(t_i) u_i \\ \tilde{Z}_0 = Z_0 \end{cases} \end{aligned}$$

with $\mathbf{A}_i^l = I_d + \Delta A(\bar{Z}_i^{l-1}, t_i)$, $\Gamma^l(t_i) := \Gamma(\bar{Z}_i^{l-1}, t_i)$ and \bar{Z}^{l-1} the state variable corresponding to the optimal control \bar{u}^{l-1} , the minimizer of the cost $C_w^{l-1}(u \mid Y; Z_0)$.

3. If $\sum_{i=0}^n \left\| \bar{Z}_i^l - \bar{Z}_i^{l-1} \right\|_2^2 < \varepsilon$ then stop otherwise go back to step 2.
4. Use $\bar{Z} \simeq \bar{Z}^l$ as state variable predictor.

The interest of this algorithm is that at each iteration l , the optimal control problem of step 2 can be solved using Linear-Quadratic theory which ensures:

- the existence and uniqueness of the solution problem \bar{u}^l ,
- that \bar{u}^l and the corresponding state predictor \bar{Z}^l can be computed by solving the Riccati equation (11) with $\mathbf{A}_i^l := I_d + \Delta A_\theta(\bar{Z}_i^{l-1}, t_i)$ and $\Gamma^l(t_i) := \Gamma(\bar{Z}_i^{l-1}, t_i)$.

4.3 Selection of weights w

A data driven selection method for w needs to be specified. Let us denote \bar{u}_w the solution of the control problem (10) obtained for a given weight value w and the corresponding estimator $\widehat{\psi} = (\widehat{\theta}, \widehat{\sigma})$. The sequence $\bar{u}_w = (\bar{u}_{w,0}, \dots, \bar{u}_{w,n-1})$ is supposed to mimic increments of a Brownian motion. So ideally $\left\| \bar{u}_{w,i} \right\|_2^2 \sim \chi^2(d_U)$ with $\chi^2(d_U)$ the χ^2 distribution with d_U degrees of freedom. Thus the

i.i.d sequence $(\|\bar{u}_{w,0}\|_2^2, \dots, \|\bar{u}_{w,n-1}\|_2^2)$ has ideally a density proportional to $\prod_i \|\bar{u}_{w,i}\|_2^{2\left(\frac{d_U}{2}-1\right)} e^{-\|\bar{u}_{w,i}\|_2^2/2}$. Based on that, we choose the optimal weight \hat{w} among a set of values W which maximizes the external criteria:

$$K(w) = \prod_i \|\bar{u}_{w,i}\|_2^{2\left(\frac{d_U}{2}-1\right)} e^{-\|\bar{u}_{w,i}\|_2^2/2}. \tag{13}$$

There are some similarities between our method and generalized Profiling (Ramsay et al. 2007). In both cases, estimation is based on a nested optimization procedure. 1/ The hyperparameter w balancing the model and data fidelity is chosen via the minimization of an outer criteria K given by (13). 2/ For a given w , the structural parameter ψ is estimated by minimizing the middle criteria H^{mB} given by (8). 3/ For a given set (w, ψ) , a state variable predictor is computed by optimizing the inner criteria C_w given by (10).

4.4 Summary of the estimation procedure

We summarize the method with a pseudo-algorithm formalism:

- Outer criteria: estimation of optimal weight \hat{w} defined by:

$$\hat{w} := \arg \min_{w \in W} K(w).$$

- Middle criteria: estimate $\hat{\psi}$ of ψ defined by:

$$\hat{\psi} := \arg \min_{\psi} H^{mB}(\psi \mid Y, \bar{Z}_{\psi}).$$

- Inner criteria: compute state predictor \bar{Z}_{ψ} via algorithm presented in Sect. 4.2.1 for linear SDEs or Sect. 4.2.2 for nonlinear ones.

5 Simulation study

5.1 Experimental design for the simulations

A simulation study is conducted to analyze the practical accuracy of our method and its computational cost on three partially observed hypoelliptic systems, one linear and two non-linear ones.

Each system is observed on an interval $[0, T]$ and is sampled at n times uniformly at every $\frac{T}{(n-1)}$ time points with Euler-Maruyama scheme. Three different values for the set (T, n) are tested to quantify the effects of the interval length T and the sample size n on estimation accuracy. Estimation results are given in terms of empirical bias and variance computed after Monte- Carlo simulations based on $N_{MC} = 1000$ trials. The

sample size has an important impact on the computational efficiency of the method. Thus, we also give the mean computational time for a given w .

5.2 Examples

5.2.1 Monotone cyclic feedback system

We consider a neuronal monotone cyclic feedback system proposed in Ditlevsen and Löcherbach (2017). This model describes the oscillatory behavior of a system of three populations of neurons in interaction:

$$\begin{aligned} dX_1(t) &= (-\nu X_1(t) + X_2(t)) dt \\ dX_2(t) &= (-\nu X_2(t) + X_3(t)) dt \\ dX_3(t) &= -\nu X_3(t) dt + cdW_t \end{aligned} \tag{14}$$

where $X_1(t)$, $X_2(t)$ and $X_3(t)$ are the limit dynamics of each population.

We consider partial observations with $C = (100)$. The model is defined by $g_\theta(x_1, x_2, x_3) = \begin{pmatrix} -\nu x_1 + x_2 \\ -\nu x_2 + x_3 \end{pmatrix}$ and $h_\theta(x_1, x_2, x_3) = -\nu x_3$, $A_\theta(t) = \begin{pmatrix} -\nu & 1 & 0 \\ 0 & -\nu & 1 \\ 0 & 0 & -\nu \end{pmatrix}$, $r_\theta(t) = (000)^T$ and $\Gamma_\sigma(t) = (00c)^T$. The system is hypoelliptic.

Applying the generation step of the weak Hörmander condition gives $L_0 = (00, c)^T$, $L_1 = [f_\theta(Z), L_0]^T = (0 - c \nu c)^T$ and $L_2 = [f_\theta(Z), L_1]^T = (c 0 \nu c^2)^T$. Matrix $(L_0 L_1 L_2)$ spans \mathbb{R}^3 .

Interestingly the need for a second iteration to ensure hypoellipticity is mimicked by our connexity condition (2) which needs the auxiliary variable x_2 to verify $\frac{\partial g_{\theta,1}}{\partial x_2} \frac{\partial g_{\theta,2}}{\partial x_3} = 1 \neq 0$.

Let us now detail assumption (H1). For $m_B = 2$, we have $G_{i+2:i+m_B} = 0$, $G_{i+1:i+m_B} = 0$ and $G_{i:i+m_B} = \sqrt{\Delta}c$. Thus $m_B = 2$ implies hypothesis (H1) and $X_{i:i+m_B}$ only depends on u_i . So the estimator is defined as the minimizer of H^{m_B} given by equation (8).

A thousand simulations are performed with initial conditions $(X_1(0), X_2(0), X_3(0)) = (0, 0, 0)$ and true parameter values set to $\nu = 0.2$ and $c = 0.15$. Figure 1 illustrates a simulation on the observation interval $[0, T] = [0, 20]$.

As said before, the estimation of (ν, c) is made from the observation of X_1 only. The initial condition are considered known. We choose w among $W = \{10^{15}, 10^{20}, 10^{25}, 10^{30}\}$. Results are given in Table 1 for different values of T and n : $(T, n) = (10, 10^3), (100, 10^3), (10, 10^4)$. The estimators have a good precision which increases with T and mesh refinement. Also, we point out the computation time is expressed in terms of second for the case $n = 10^3$ where it was expressed in terms of hours in Clairon and Samson Clairon and Samson (2020) for an equivalently complex linear model (see Harmonic Oscillator model, section 6.2.1). This is due to the complexity of our method which repeatedly solves the control problem to obtain our state variable predictor. It requires to integrate a sequence of finite difference equations of length n . This explains why the computational time of our method increases with n .

Fig. 1 Simulated trajectory of the neuronal monotone cyclic feedback system (14) with parameters $\nu = 0.2$ and $c = 0.15$. X_1 : Solid line, X_2 : dashed line, X_3 : dotted line

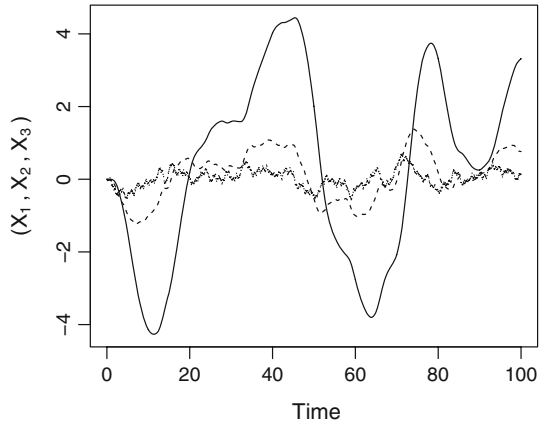


Table 1 Cyclic feedback model

	ν	c	Comp time
True value	0.2	0.15	
$T = 10, n = 10^3$	0.23 (5e-3)	0.14 (1e-5)	9 s
$T = 100, n = 10^3$	0.21 (6e-4)	0.15 (1e-6)	19 s
$T = 10, n = 10^4$	0.22 (2e-3)	0.15 (1e-6)	2 min 50 s

Estimation of parameters from 1000 simulated trajectories (mean and variance) and mean computational time for a given w

5.2.2 Hypoelliptic FitzHugh–Nagumo model

We consider now the hypoelliptic neuronal model, which is a minimal representation of a spiking neuron model such as the Hodgkin-Huxley model (Leon and Samson 2018). It is defined as:

$$\begin{aligned} dV_t &= \frac{1}{\varepsilon}(V_t - V_t^3 - U_t + s)dt \\ dU_t &= (\gamma V_t - U_t + \beta)dt + \sigma dW_t \end{aligned} \tag{15}$$

where the variable V_t represents the membrane potential of a neuron at time t , and U_t is a recovery variable, which could represent channel kinetics. Parameter s is the magnitude of the stimulus current. Often s represents injected current and is thus controlled in a given experiment. It is therefore assumed known and set to $s = 0$.

We consider partial observations with $C = (1\ 0)$. The model is non-linear. Several choices for $A_\theta(Z_t, t)$ are possible, we take $A_\theta(Z_t, t) = \begin{pmatrix} (1 - V_t^2)/\varepsilon & -1/\varepsilon \\ \gamma & -1 \end{pmatrix}$ and $r_\theta(t) = \begin{pmatrix} 0 \\ \beta \end{pmatrix}$. Since $\Gamma_\sigma(t) = (0\ \sigma)^T$, we got $C\Gamma_\sigma = 0$, $\frac{\partial}{\partial U}(\frac{1}{\varepsilon}(V_t - V_t^3 - U_t)) = -\frac{1}{\varepsilon} \neq 0$ and $C A_\theta(\tilde{Z}_{i+1}, t_{i+1})\Gamma_\sigma(\tilde{Z}_i, t_i) = -\sigma/\varepsilon \neq 0$. As described in Sect. 3.2, the

Fig. 2 Simulated trajectory of the neuronal hypoelliptic FitzHugh–Nagumo model (15) with parameters $(\epsilon, \gamma, \beta) = (0.1, 1.5, 0.8)$ and $\sigma = 0.3$. V : solid line, U : dashed line

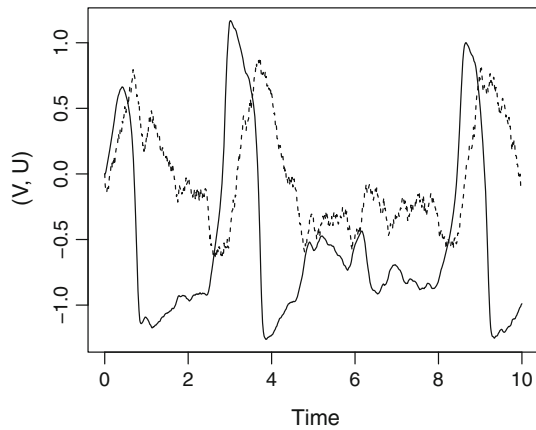


Table 2 Hypoelliptic FitzHugh–Nagumo model

	ϵ	γ	β	σ	Comp time
True value	0.1	1.5	0.8	0.3	
$T = 1, n = 10^3$	0.11 (2e-3)	2.13 (1e-1)	1.18 (3e-1)	0.32 (2e-2)	1 min 10 s
$T = 10, n = 10^3$	0.09 (2e-5)	1.58 (6e-2)	0.87 (5e-2)	0.29 (2e-4)	3 min 15 s
$T = 1, n = 10^4$	0.10 (4e-4)	1.90 (8e-2)	1.10 (2e-1)	0.31 (3e-3)	8 min 30 s

Estimation of parameters from 1000 simulated trajectories (mean and variance) and mean computational time for a given w

model is a 1-step hypoelliptic system and we can use the corresponding simplified expression of H^{m_B} for the parameter estimation.

A thousand simulations are performed with initial conditions set to $(V_0, U_0) = (0, 0)$ and true parameter values $(\epsilon, \gamma, \beta) = (0.1, 1.5, 0.8)$ and $\sigma = 0.3$. Figure 2 illustrates a simulation on the observation interval $[0, T] = [0, 20]$.

The estimation is made from the observation of V only and the initial conditions are considered unknown and need to be estimated. We choose w among $W = \{10^{16}, 10^{18}, 10^{20}, 10^{25}\}$.

Results are given in Table 2 for $(T, n) = (1, 10^3), (10, 10^3), (1, 10^4)$. As in the previous example, we observe the bias and variance decreasing with T and n . Regarding the computation time and comparative accuracy with other methods, we recall in Table 3, the results obtained for the case $T = 10, n = 10^3$ by Clairon and Samson (2020) and Ditlevsen and Samson (2019) (they are originally presented in Clairon and Samson (2020) table 4, section 6.2.2). We obtain estimation with equivalent or higher accuracy and always with significantly reduced computational cost.

5.2.3 Synaptic-conductance model

We consider the conductance-based model with diffusion synaptic input defined in Ditlevsen and Samson (2019). It describes the voltage dynamics across the membrane

Table 3 Hypoelliptic FitzHugh–Nagumo model

	ε	γ	β	σ	Comp time
True value	0.1	1.5	0.8	0.3	
$\hat{\psi}$	0.09 (2e-5)	1.58 (6e-2)	0.87 (5e-2)	0.29 (2e-4)	3 min 15 s
Clairon and Samson	0.09 (1e-4)	1.27 (2e-2)	0.65 (1e-2)	0.31 (4e-3)	45 min
Ditlevsen and Samson	0.10 (1e-4)	1.59 (2e-2)	0.87 (2e-2)	0.31 (4e-4)	6 h 00 min

Estimation results for the case $T = 10, n = 10^3$ with our method (first row), Clairon and Samson (2020) (second row), Ditlevsen and Samson (2019) third row

of a neuron:

$$\begin{aligned}
 C_c dV_t &= (-G_L(V_t - V_L) - G_{E,t}(V_t - V_E) - G_{I,t}(V_t - V_I) + I_{inj})dt \\
 dG_{E,t} &= -\frac{1}{\tau_E}(G_{E,t} - g_E) + \sigma_E \sqrt{G_{E,t}} dW_{E,t} \\
 dG_{I,t} &= -\frac{1}{\tau_I}(G_{I,t} - g_I) + \sigma_I \sqrt{G_{I,t}} dW_{I,t}
 \end{aligned}
 \tag{16}$$

where C_c is the total capacitance, G_L, G_E and G_I are the leak, excitation and inhibition conductances, V_L, V_E and V_I are their respective reversal potentials, and I_{inj} is the injected current. The conductances $G_{E,t}$ and $G_{I,t}$ are assumed to be stochastic functions of time, where $W_{E,t}$ and $W_{I,t}$ are two independent Brownian motions. The square root in the diffusion coefficient ensures that the conductances stay positive. Parameters τ_E, τ_I are time constants, g_E, g_I are the mean conductances and σ_E, σ_I the diffusion coefficients. Here $U_t = (G_{E,t}, G_{I,t})$. We assume to know the capacitance, the reversal potentials, and the injected current.

We consider partial observations with only component V_t observed which corresponds to the observation matrix $C = (1\ 0\ 0)$. For the pseudo-linear representation, we choose $A_\theta(Z_t, t) = \begin{pmatrix} -G_L/C_c - (V_t - V_E)/C_c - (V_t - V_I)/C_c \\ 0 & -1/\tau_E & 0 \\ 0 & 0 & -1/\tau_I \end{pmatrix}$, $r_\theta(t) = \begin{pmatrix} (G_L V_L + I_{inj})/C_c \\ g_E/\tau_E \\ g_I/\tau_I \end{pmatrix}$. Since $g_\theta(V, G_E, G_I) = (-G_L(V - V_L) - G_E(V - V_E) - G_I(V - V_I) + I_{inj})/C_c$, $C = (1\ 0\ 0)$ and $\Gamma_\sigma(Z_t, t) = \begin{pmatrix} 0 & \sigma_E \sqrt{G_{E,t}} & 0 \\ 0 & 0 & \sigma_I \sqrt{G_{I,t}} \end{pmatrix}^T$, we got $C\Gamma_\sigma = (0\ 0)$, $\frac{\partial g_\theta}{\partial (G_E, G_I)}(V, G_E, G_I)$ is of full rank and $C A_\theta(\tilde{Z}_{i+1}, t_{i+1}) \Gamma_\sigma(\tilde{Z}_i, t_i) \neq 0$. It is a 1-step hypoelliptic system and enters into the framework described in Sect. (3.2).

A thousand simulations are performed with initial conditions set to $(V_0, G_{E,0}, G_{I,0}) = (-60, 10, 1)$ and true parameters values set to $(G_L, V_L, V_E, V_I, I_{inj}, g_E) = (50, -70, 0, -80, -60, 17.8)$, $(\tau_E, \tau_I, g_I) = (0.5, 1, 9.4)$ and $(\sigma_E, \sigma_I) = (0.1, 0.1)$. The estimation of $(\tau_E, \tau_I, g_I, \sigma_E, \sigma_I)$ is made from the observation of V only. For the sake of identifiability, the initial conditions are assumed known. We choose w among $W = \{10^8, 5 \times 10^8, 10^9, 5 \times 10^9\}$.

Table 4 Synaptic-conductance model

	τ_E	τ_I	g_I	σ_E	σ_I	Comp time
True value	0.5	1	9.4	0.1	0.1	
$T = 20, n = 10^3$	0.51 (2e-3)	1.11 (0.04)	9.45 (0.02)	0.05 (3e-4)	0.15 (1e-3)	1 min 40 s
$T = 200, n = 10^3$	0.47 (4e-4)	1.11 (0.01)	9.41 (2e-3)	0.09 (7e-6)	0.14 (8e-5)	2 min 00 s
$T = 20, n = 10^4$	0.52 (2e-3)	1.08 (0.04)	9.40 (0.02)	0.11 (4e-5)	0.07 (1e-3)	1 h 40 min

Estimation of parameters from 1000 simulated trajectories (mean and variance) and mean computational time for a given w

Results are given in Table 4 for $(T, n) = (20, 10^3), (200, 10^3), (20, 10^4)$. We observe a difference in terms of accuracy between the estimation of $\theta = (\tau_E, \tau_I, g_I)$ and $\sigma = (\sigma_E, \sigma_I)$, the latter being more biased than the former. This is understandable and already noticed in Ditlevsen and Samson (2019). Indeed we have to estimate the diffusion of two sources of stochastic disturbance from only one resulting signal. This estimation problem is more complex than the previous ones as the number of parameters to estimate is higher and the ratio of observed/unobserved states is also higher. This explains the higher computational time for the case $T = 20, n = 10^4$.

6 Discussion

In this work, we propose a new estimation method which gives an unifying framework for elliptic and hypoelliptic systems, partially or fully observed. For this, we rely on a lagged discretization of the original SDE which lets enough time to the stochastic perturbations to affect all the state variables. By doing so, we have constructed a statistical criteria based on a well-defined density even for partially observed hypoelliptic SDEs. This criteria requires a state variable predictor obtained by solving a control problem balancing data and model fidelity. The numerical procedures used to solve it do not require $B_\sigma(Z_t, t)$ to be of full rank and so are well adapted to hypoelliptic systems. Because of this, the deterministic control perspective constitutes a relevant alternative to MCMC approaches and explain the reasonable computational cost of our method. It only provides the pointwise estimator \bar{Z} we need for ψ estimation and does not aim to reconstruct its whole distribution. Now, we conclude this work by presenting refinements and extensions of the presented method which will be investigated in the future.

Our method requires to select an hyperparameter w . To bypass this, we aim to define in future works our predictor \bar{u} as a maximum a posteriori estimator (MAP) in a functional space. For this, we will rely on the work of Dashti et al. (2013) in which the MAP is defined as the solution of a new deterministic optimal control problem where our regularization term $\frac{2}{w} \ln P(u)$ is replaced by $\|u\|_E$. Here, E denotes the so-called Cameron-Martin space which is totally determined by the covariance operator of the Brownian motion u and does not involve a nuisance parameter.

Also, we think our method is well suited for generalization to SDEs driven by processes W_t different from the Wiener ones. All it requires would be to know their

densities to modify H^{m_B} accordingly and the penalization term appearing in the control problem. Interestingly in this general setting, to a given type of SDE will be associated a given type of deterministic optimal control problem.

We end this work by going one step further about the connection between our estimation problem and control theory by pointing out its similarity with the issue of structural assessment of a system controllability as exposed in Daoutidis and Kravaris (1992). Their problem is the following: given an ODE $\dot{x} = f(x) + Bu$, can we control the behavior of some system outputs $y_i = h_i(x)$ knowing that $u = \{u_j\}_{j \in \llbracket 1, d_U \rrbracket}$ only affects directly a subset of x ? This leads to define in a similar way as our m_l 's the integers r_{ij} , named relative orders, quantifying the sluggishness of y_i response to u_j variation. Then, conditions for controllability are formulated via the non-singularity of a so-called characteristic matrix $C(x)$ constructed from the r_{ij} 's and Lie derivative in a way mirroring the matrix M appearing in the proof of proposition 1. From this, we hope to derive m_B in a less exploratory manner directly from a given model structure and make explicit its link with the required number of iteration of generalization step 2 to fulfill Hörmander condition.

Supplementary information A supplementary file containing the proofs of propositions 1 and 2 and the derivation of the algorithm presented in Sect. 4.2.1 is available alongside this article.

Acknowledgements Quentin Clairon gratefully acknowledges the Innovative Medicines Initiative 2 Joint Undertaking under projects EBOVAC1 and EBOVAC3 (respectively Grant Agreement Nos. 115854 and 800176). The IMI2 Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and the European Federation of Pharmaceutical Industries and Association. Adeline Samson gratefully acknowledges the LabEx PERSYVAL-Lab (ANR-11-LABX-0025-01), MIAI@Grenoble Alpes, (ANR-19-P3IA-0003) and the Grant ANR ChaMaNe, ANR-19-CE40-0024.

Code availability Our estimation method is implemented in R and a code reproducing the examples of Sect. 5 is available on a GitHub repository located https://github.com/QuentinClairon/SDE_estimation_via_optimal_control.git here.

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

- Bierkens J, van der Meulen F, Schauer M (2020) Simulation of elliptic and hypo-elliptic conditional diffusions. *Adv Appl Probab* 52:173–212
- Buckwar E, Samson A, Tamborrino M, et al (2021) Splitting methods for SDEs with locally lipschitz drift. An illustration on the FitzHugh-Nagumo model. [arXiv:2101.01027](https://arxiv.org/abs/2101.01027)
- Cimen T (2008) State-dependent Riccati equation (SDRE) control: a survey. *IFAC Proc* 41:3761–3775
- Cimen T, Banks S (2004) Global optimal feedback control for general nonlinear systems with nonquadratic performance criteria. *Syst Control Lett* 53:327–346
- Cimen T, Banks S (2004) Nonlinear optimal tracking control with application to super-tankers for autopilot design. *Automatica* 40:1845–1863
- Clairon Q (2021) A regularization method for the parameter estimation problem in ordinary differential equations via discrete optimal control theory. *J Stat Plann Inference* 210:1–19

- Clairon Q, Brunel N (2019) Tracking for parameter and state estimation in possibly misspecified partially observed linear ordinary differential equations. *J Stat Plann Inference* 199:188–206
- Clairon Q, Samson A (2020) Optimal control for estimation in partially observed elliptic and hypoelliptic linear stochastic differential equations. *Stat Infer Stoch Process* 23(1):105–127
- Coombes S, Byrne A (2019) Lecture notes in nonlinear dynamics in computational neuroscience: from physics and biology to ICT, Springer, chap Next generation neural mass models. *PoliTO Springer Series*
- Daoutidis P, Kravaris C (1992) Structural evaluation of control configurations for multivariable nonlinear processes. *Chem Eng Sci* 47(5):1091–1107
- Dashti M, Law KJ, Stuart AM et al (2013) Map estimators and their consistency in Bayesian nonparametric inverse problems. *Inverse Probl* 29(9):095017
- DeVille R, Vanden-Eijnden E, Muratov C (2005) Two distinct mechanisms of coherence in randomly perturbed dynamical systems. *Phys Rev E* 72(3, 1):031105
- Dietz H (2001) Asymptotic behaviour of trajectory fitting estimators for certain non-ergodic SDE. *Stat Infer Stoch Process* 4:249–258
- Ditlevsen S, Greenwood P (2013) The Morris–Lecar neuron model embeds a leaky integrate-and-fire model. *J Math Biol* 67(2):239–259
- Ditlevsen S, Löcherbach E (2017) Multi-class oscillating systems of interacting neurons. *Stoch Process Appl* 127:1840–1869
- Ditlevsen S, Samson A (2014) Estimation in the partially observed stochastic Morris–Lecar neuronal model with particle filter and stochastic approximation methods. *Ann Appl Stat* 2:674–702
- Ditlevsen S, Samson A (2019) Hypoelliptic diffusions: discretization, filtering and inference from complete and partial observations. *J R Stat Soc B* 81:361–384
- Ditlevsen S, Sørensen M (2004) Inference for observations of integrated diffusion processes. *Scand J Stat* 31(3):417–429
- Engl HW, Flamm C, Kügler P, et al (2009) Inverse problems in systems biology. *Inverse Probl* 25(12)
- Ditlevsen P, Ditlevsen S, Andersen K (2002) The fast climate fluctuations during the stadial and interstadial climate states. *Ann Glaciol* 35:457–462
- Genon-Catalot V, Jeantheau T, Larédo C (2000) Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli* 6(6):1051–1079
- Gerstner W, Kistler W (2002) Spiking neuron models. Cambridge University Press, Cambridge
- Gloter A (2000) Discrete sampling of an integrated diffusion process and parameter estimation of the diffusion coefficient. *ESAIM Probab Stat* 4:205–227
- Gloter A (2006) Parameter estimation for a discretely observed integrated diffusion process. *Scand J Stat* 33(1):83–104
- Gloter A, Yoshida N (2021) Adaptive estimation for degenerate diffusion process. *Electron J Stat* 15:1424–1472
- Goldwyn JH, Shea-Brown E (2011) The what and where of adding channel noise to the Hodgkin–Huxley equations. *PLOS Comput Biol* 7(11)
- Graham MM, Thiery AH, Beskos A (2019) Manifold Markov chain Monte Carlo methods for Bayesian inference in a wide class of diffusion models. [arXiv:1912.02982](https://arxiv.org/abs/1912.02982)
- Iolov A, Ditlevsen S, Longtin A (2017) Optimal design for estimation in diffusion processes from first hitting times. *SIAM/ASA J Uncertainty Quant* 5:88–110
- Kutoyants Y (1991) Minimum distance parameter estimation for diffusion type observation. *Comptes rendus de l Academie des sciences* 312:637
- Leimkuhler B, Matthews C (2015) Molecular Dynamics with deterministic and stochastic numerical methods. *Interdiscip Appl Math*, vol 39. Springer, Berlin
- Leon J, Samson A (2018) Hypoelliptic stochastic FitzHugh–Nagumo neuronal model: mixing, up-crossing and estimation of the spike rate. *Ann Appl Probab* 28(4):2243–2274
- Leon J, Rodriguez L, Ruggiero R (2019) Consistency of a likelihood estimator for stochastic damping hamiltonian systems. *Totally observed data. ESAIM Probab Stat* 23:1–36
- Lu F, Lin K, Chorin A (2016) Comparison of continuous and discrete-time data-based modeling for hypoelliptic systems. *Commun Appl Math Comput* 11:187–216
- Melynkova A (2020) Parametric inference for hypoelliptic ergodic diffusions with full observations. *Stat Infer Stoch Process* 23:595–635
- Paninski L, Vidne M, DePasquale B et al (2012) Inferring synaptic inputs given a noisy voltage trace via sequential Monte Carlo methods. *J Comput Neurosci* 33(1):1–19

- Pokern Y, Stuart A, Wiberg P (2009) Parameter estimation for partially observed hypoelliptic diffusions. *J R Stat Soc B* 71(1):49–73
- Ramsay J, Hooker G, Cao J et al (2007) Parameter estimation for differential equations: a generalized smoothing approach. *J R Stat Soc B* 69:741–796
- Samson A, Thieullen M (2012) Contrast estimator for completely or partially observed hypoelliptic diffusion. *Stoch Process Appl* 122:2521–2552
- Tuckwell HC, Ditlevsen S (2016) The space-clamped Hodgkin–Huxley system with random synaptic input: inhibition of spiking by weak noise and analysis with moment equations. *Neural Comput* 28(10):2129–2161
- Wu L (2001) Large and moderate deviations and exponential convergence for stochastic damping Hamiltonian systems. *Stoch Process Appl* 91:205–238

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.