# Recurrence and parameter estimation for degenerate diffusions with internal variables and randomly perturbed time-inhomogeneous deterministic input

Dissertation

zur Erlangung des Grades "Doktor der Naturwissenschaften"
am Fachbereich Physik, Mathematik und Informatik
der Johannes Gutenberg-Universität in Mainz

Simon Holbach

geboren in Adenau

Mainz, den 22. Mai 2018

Datum der mündlichen Prüfung: 11. September 2018

# Contents

# Abstract

Taking a multidimensional time-homogeneous dynamical system and adding a randomly perturbed time-dependent deterministic signal to some of its components gives rise to a high-dimensional system of stochastic differential equations which is driven by possibly very low-dimensional noise. Equations of this type are commonly used in biology for modelling neurons or in statistical mechanics for certain Hamiltonian systems.

This thesis is focused on studying two general properties of such a system. In the first part, we use methods from stability theory and control theory as well as Hörmander's condition in order to provide conditions that are sufficient for the corresponding stochastic process to be positive recurrent in the sense of Harris. Harris recurrence gives rise to Limit Theorems for a large class of functionals of the process and can thus be the foundation for applications in asymptotic statistics.

In the second part, considering a statistical model associated to a parametrised class of smooth signals, we exploit Harris recurrence in order to prove Local Asymptotic Normality in the sense of LeCam for the estimation of these parameters under continuous observation of certain components of the process.

# Zusammenfassung

Indem einem mehrdimensionalen zeitlich homogenen dynamischen System in einigen Komponenten ein in zufälliger Weise gestörtes zeitabhängiges deterministisches Signal hinzugefügt wird, erhält man ein hochdimensionales System stochastischer Differentialgleichungen, welches von möglicherweise sehr niedrigdimensionalem Rauschen angetrieben wird. Gleichungen dieser Art finden Anwendung in der Biologie bei der Modellierung von Neuronen oder in der statistischen Mechanik im Zusammenhang mit gewissen Typen Hamiltonscher Bewegungsgleichungen.

Diese Doktorarbeit widmet sich zwei allgemeinen Eigenschaften solcher Systeme. Im ersten Teil verwenden wir Methoden der Stabilitätstheorie sowie der Kontrolltheorie und die Hörmander-Bedingung, um hinreichende Bedingungen anzugeben, unter denen der entsprechende stochastische Prozess positiv Harris-rekurrent ist. Harris-Rekurrenz liefert Grenzwertsätze für große Klassen von Funktionalen des Prozesses und kann somit als Grundlage für Anwendungen in der asymptotischen Statistik dienen.

Im zweiten Teil betrachten wir ein statistisches Modell zu einer parametrischen Klasse glatter Signale und nutzen Harris-Rekurrenz, um für die Schätzung dieser Parameter unter stetiger Beobachtung gewisser Komponenten des Prozesses Lokalasymptotische Normalität im Sinne von LeCam zu beweisen.

# Chapter 1

# Introduction

In this chapter, we introduce the general model that we want to study in this thesis. Afterwards, we offer insight into the backgrounds of some particular examples which served as its motivational basis and which will be brought up repeatedly in the course of this text.

## 1.1 Degenerate diffusions with internal variables and randomly perturbed time-inhomogeneous deterministic input

Let $V \subset \mathbb{R}^L$ be a connected set with non-empty interior. Let

$$F\colon \mathbb{R}^N \times V \to \mathbb{R}^N \quad \text{and} \quad G\colon \mathbb{R}^N \times V \to \mathbb{R}^L$$

be locally Lipschitz continuous functions. Finally, let

$$S\colon [0,\infty) \to \mathbb{R}^N$$

be a continuous signal and consider the deterministic dynamical system

$$\begin{aligned} \dot{x} &= F(x,y) + S, \\ \dot{y} &= G(x,y), \end{aligned} \tag{DDS}$$

where $\dot{x}$ and $\dot{y}$ are the time-derivatives of the time-dependent variables $x\colon [0,\infty) \to \mathbb{R}^N$ and $y\colon [0,\infty) \to V$, respectively. Using integral notation, we can alternatively write (DDS) as

$$\begin{aligned} dx(t) &= F(x,y)(t)dt + S(t)dt, \\ dy(t) &= G(x,y)(t)dt. \end{aligned} \tag{DDS'}$$

If it exists, we write

$$(x,y)[x_0,y_0,S](t) \tag{1.1}$$

for its unique solution at time $t \in [0,\infty)$ with starting condition $(x,y)(0) = (x_0,y_0) \in \mathbb{R}^N \times V$.[1]

The system is divided into two groups of variables: The $N$ components of $x$ whose dynamics depend directly on the signal and the $L$ components of the *internal variables* $y$ which are affected by the signal only indirectly through the influence of $x$. Intuitively speaking, we can think of (DDS) as a dynamical system with no intrinsic time-inhomogeneity which then receives an additional time-dependent *external input* in some of its variables, while the remaining variables merely describe an interior mechanism. Note that the only source of time-inhomogeneity is in fact the signal – if the system receives no external input (i.e. $S \equiv 0_N$), it is entirely homogeneous in time. Systems of this kind frequently arise in the context of neuroscience and statistical mechanics – some particularly interesting specific examples (such as the Hodgkin-Huxley model, simplified neuron models, or a chain of coupled oscillators) will be presented and explained in Section 1.2.

The conceptual idea of our model is that the signal is not actually received in its original shape, but is subject to perturbations by external noise (i.e. noise that is independent of the rest of the system). To take account of this notion, it seems natural to substitute the signal term $S(t)dt$ in (DDS') with the increment $dZ_t$ of a time-inhomogeneous Ornstein-Uhlenbeck process carrying $S(t)$ in its time-dependent mean reversion level, i.e.

$$dZ_t = [S(t) - \gamma Z_t]dt + \sigma dW_t$$

for an $N$-dimensional standard Brownian Motion $W$, some $\gamma \in (0,\infty)$ and a suitable $\sigma \in \mathbb{R}^{N\times N}$. While this will usually be the situation we have in mind, in general nothing keeps us from considering slightly less restrictive Ornstein-Uhlenbeck type processes of the shape

$$dZ_t = [S(t) + b(Z_t)]dt + \sigma(Z_t)dW_t, \tag{1.2}$$

where $W$ is now an $M$-dimensional standard Brownian Motion, while $b\colon \mathbb{R}^N \to \mathbb{R}^N$ and $\sigma\colon \mathbb{R}^N \to \mathbb{R}^{N\times M}$ are suitable drift and volatility functions.[2] If $b$ and $\sigma$ are Lipschitz

[1] Consult the list of symbols for our notational conventions regarding column and row vectors.

[2] As usual, this notation for stochastic differential equations is short for

$$Z_t^{(i)} - Z_0^{(i)} = \int_0^t [S_i(s) - b_i(Z_s)]ds + \sum_{j=1}^M \int_0^t \sigma_{i,j}(Z_s)dW_s^{(j)} \quad \text{for all } i \in \{1,\dots,N\},$$

and we will use this abbreviation throughout this text.

continuous and $S$ is bounded, the equation (1.2) satisfies classical growth and Lipschitz conditions and hence admits a unique non-explosive strong solution (see for example [43, Theorems 5.2.5 and 5.2.9]). Note that this stochastic differential equation can be viewed as a generalised version of the classical signal in noise model (take $M = N = 1$, $b \equiv 0$, and $\sigma \equiv 1$, see for example [39, Example I.7.3, Chapter III.5]), which arises in a wide variety of fields including communication, radiolocation, seismic signal processing, or computer-aided diagnosis and has been the subject of extensive study.

Perturbing $S(t)$ randomly in this way leads to the stochastic dynamical system

$$\begin{aligned} dX_t &= F(X_t, Y_t)dt + dZ_t, \\ dY_t &= G(X_t, Y_t)dt, \\ dZ_t &= [S(t) + b(Z_t)]dt + \sigma(Z_t)dW_t, \end{aligned} \tag{SDS}$$

issued from some starting point $(x_0, y_0, z_0) \in \mathbb{R}^N \times V \times \mathbb{R}^N$. This system can be thought of as degenerate in the following sense: Firstly, the equation for $Y$ does not incorporate the driving Brownian Motion $W$ explicitly, making it rather unclear which effect noise has on these components. Secondly, the dimension $M$ of the driving Brownian Motion can (and will usually) be much lower than the dimension $N + L + N$ of the system. This is why we call a stochastic process satisfying a system of stochastic differential equations of the type (SDS) a *degenerate diffusion with internal variables and randomly perturbed time-inhomogeneous deterministic input*. The system (SDS) is a generalisation of the one introduced in equation (18) of Section 4.1 of [36], which is a probabilistic version of a class of dynamical systems that are well-known in the mathematical modelling of neurons. This connection will be discussed in more detail in Examples 1.1 and 1.2 below.

With (SDS) we now have three groups of variables: The entirely autonomous[3] external input governed by $dZ_t$ (the "noisy signal"), the components of $X$ that depend directly on the noisy signal, and the components of the internal variable $Y$ whose dynamics are only indirectly affected by noise, since the respective differential equations incorporate neither $Z$ nor the driving Brownian Motion $W$ explicitly. Note that for this reason $Y$ is conditionally deterministic given $X$ and has continuously differentiable trajectories.

Let us write

$$\mathbb{X} := (X, Y, Z)$$

[3] We use the term autonomous in the sense of "independent of the rest of the system" and not – as is common in the theory of ordinary differential equations – in the sense of "homogeneous with respect to time".

for the entire diffusion process on the state space

$$\mathtt{E} := \mathbb{R}^N \times V \times \mathbb{R}^N.$$

We can then express the diffusion equation (SDS) as

$$d\mathbb{X}_t = B(t, \mathbb{X}_t)dt + \Sigma(\mathbb{X}_t)dW_t$$

with drift

$$B\colon [0,\infty) \times \mathtt{E} \to \mathbb{R}^{N+L+N}, \quad (t,x,y,z) \mapsto \begin{pmatrix} F(x,y) + S(t) + b(z) \\ G(x,y) \\ S(t) + b(z) \end{pmatrix}, \tag{1.3}$$

and volatility

$$\Sigma\colon \mathtt{E} \to \mathbb{R}^{(N+L+N)\times M}, \quad (x,y,z) \mapsto \begin{pmatrix} \sigma(z) \\ 0_{L\times M} \\ \sigma(z) \end{pmatrix}. \tag{1.4}$$

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space on which the $M$-dimensional Brownian Motion $W$ is defined. We always assume that, for all deterministic starting points $\mathbb{X}_0 \in \mathtt{E}$, the equation (SDS) has a unique non-explosive strong solution $\mathbb{X}$ on $(\Omega, \mathcal{A})$ under $\mathbb{P}$.

This thesis pursues two objectives. The first one (treated in Chapter 2) is to find verifiable conditions on (DDS) and the way in which the external equation of (SDS) perturbs the signal, under which positive Harris recurrence of $\mathbb{X}$ can be established in spite of its apparent degeneracy (the main results being Theorems 2.11, 2.13, 2.25, 2.46, and 2.56 in combination with Theorem 2.3). Of course, in order to even talk about Harris recurrence, one has to look for some time-homogeneous substructure of $\mathbb{X}$, which suggests the basic assumption that the signal function $S$ should be periodic. As we will see in Section 1.2, this assumption does not hurt at all, as it is commonly fulfilled in the main applications we have in mind. This thesis' second objective (treated in Chapter 3) revolves around the following statistical problem: Suppose that $S$ has an unknown periodicity and also depends on some unknown finite-dimensional shape parameter. What kind of quality can we expect to achieve for an estimator for both of these parameters, if we cannot observe the entire process $\mathbb{X}$, but only its $X$-components? We answer this question by first arguing that, from a statistical point of view, it does not matter if we observe $\mathbb{X}$, $X$, or even $Z$ (see Remark 3.1 and Proposition 3.3) and then proving Local Asymptotic Normality for the sequence of statistical experiments corresponding to continuous observation of $Z$ (Theorem 3.11). Positive Harris recurrence plays a crucial role in making this work, but here we only require it for the external component $Z$ and not for the entire process $\mathbb{X}$.

The respective tools for these two objectives will be introduced and explained whenever it is the most convenient for the flow of this thesis. Similarly, we will comment on relevant existing literature in the respective spots.

## 1.2 Examples

In this Section, we introduce the main examples that inspired this work and will accompany us throughout it.

**Example 1.1.** Let $N = 1$, $L = 3$, $V = [0,1]^3$ and consider the system

$$\begin{aligned} \dot{x} &= F(x,y) + S, \\ \dot{y}_i &= \alpha_i(x)(1-y_i) - \beta_i(x)y_i \quad \text{for all } i \in \{1,2,3\}, \end{aligned} \tag{HH}$$

where for all $(x,y) = (x, y_1, y_2, y_3)^\top \in \mathbb{R} \times V$ we set

$$F(x,y) = -36y_1^4(x+12) - 120y_2^3y_3(x-120) - 0.3(x-10.6) \tag{1.5}$$

and

$$\begin{aligned} \alpha_1(x) &= \begin{cases} \frac{0.1-0.01x}{\exp(1-0.1x)-1}, & x \neq 10, \\ 0.1, & \text{else}, \end{cases} & \beta_1(x) &= 0.125\exp(-x/80), \\ \alpha_2(x) &= \begin{cases} \frac{2.5-0.1x}{\exp(2.5-0.1x)-1}, & x \neq 25, \\ 1, & \text{else}, \end{cases} & \beta_2(x) &= 4\exp(-x/18), \\ \alpha_3(x) &= 0.07\exp(-x/20), & \beta_3(x) &= \frac{1}{\exp(3-0.1x)+1}. \end{aligned} \tag{1.6}$$

As outlined in [8, pp. 156-157] the dynamical system (HH) possesses a unique global solution living in $\mathbb{R} \times V$ for every starting point. It is known as the *Hodgkin-Huxley system* and it was first introduced by Hodgkin and Huxley in 1952 (see [27], note however that we use the slightly different model constants from [41]) with the aim of describing the initiation and propagation of action potentials in the cell membrane of a neuron in response to an external stimulus. While $x$ is the membrane potential itself (usually labelled $V$ in the literature), the internal variables $y_1$, $y_2$, and $y_3$ (commonly denoted by $n$, $m$, and $h$) correspond to the ionic mechanism underlying its evolution. The two predominant ion currents in the cell membrane are import of sodium $Na^+$ and export of potassium $K^+$ through the membrane. Each of the internal variables signifies the probability that a specific type of voltage-gated ion channel is open at a given time. It is for this reason that $n$, $m$, and $h$ are often called gating variables. This notion of probability has to be understood in the sense of a law of large numbers, of course.

On a microscopic level, the variable for each gate should be able to occupy only two states: the respective ion channel being either open (equal to 1) or closed (equal to 0). The transition rates between these two states depend only on the membrane potential (as is also reflected in the internal equations of (HH)). Taking a suitable average over large numbers of these $\{0,1\}$-valued representations of ion channels leads to the $[0,1]$-valued gating variables. This idea is given a rigorous meaning by the fact that the Hodgkin-Huxley system can indeed be obtained as the limit of a sequence of piecewise deterministic Markov processes (compare [9], [20], [55], [57]). In the context of this model, the signal $S$ represents the dendritic input which the neuron receives from a large number of other neurons, transported by an even larger number of synapses located on the respective dendritic tree. The resulting "total dendritic input" can then be thought of as an average of interdependent and repeating similar currents, which is why $S$ is usually assumed to be periodic (or even constant). When modelling neurons, particular interest lies in the typical spiking behaviour of the membrane potential, a feature that is commonly agreed upon to be adequately described by the Hodgkin-Huxley model. For a more detailed modern introduction, interpretation, and an in-depth comparison with other neuron models, see for example [41] and [15].

Adding noise to (HH) in the same way we derived (SDS) from (DDS), we introduce the degenerate diffusion with internal variables and randomly perturbed time-inhomogeneous deterministic input that is given by

$$\begin{aligned} dX_t &= F(X_t, Y_t)dt + dZ_t, \\ dY_t^{(i)} &= \big[\alpha_i(X_t)\big(1 - Y_t^{(i)}\big) - \beta_i(X_t)Y_t^{(i)}\big]dt, \quad i \in \{1,2,3\}, \\ dZ_t &= [S(t) - \gamma Z_t]dt + \sigma(Z_t)dW_t, \end{aligned} \tag{SHH}$$

with $\gamma \in (0,\infty)$ and $\sigma \in C^\infty(\mathbb{R})$, starting in some $(x_0, y_0, z_0) \in \mathbb{R} \times [0,1]^3 \times \mathbb{R}$. We will call (SHH) the *stochastic Hodgkin-Huxley system (with mean reverting Ornstein-Uhlenbeck type input)* in the sequel. It was first introduced and studied by Höpfner, Löcherbach, and Thieullen in the series of the three papers [35], [36], and [37] which were published in 2016 and 2017. The constant $\gamma$ is determined by the so-called time constant of the membrane which represents spontaneous voltage decay not related to the input. For many types of neurons, the time constant is known from experiments (see [16]). A degree of freedom lies in the choice of the volatility $\sigma$ which reflects the nature of the influence of noise. In the past, mean reverting Ornstein-Uhlenbeck type equations with various volatilities have been used to model the membrane potential itself (see for example [46] or [30]), and in a sense (SHH) can be viewed as a refinement of this kind of model. If $\sigma$ is Lipschitz continuous, existence of a unique non-exploding strong solution taking values in $\mathtt{E} = \mathbb{R} \times [0,1]^3 \times \mathbb{R}$ follows from the same arguments as in [35, Proposition 1] and [36, Proposition 2].

In the afore-mentioned articles [35], [36], and [37], the authors prove that if the external equation is of classical Ornstein-Uhlenbeck type

$$dZ_t = [S(t) - \gamma Z_t]dt + \sigma dW_t$$

with constant $\sigma \in (0, \infty)$ or (when the signal $S$ is non-negative) of Cox-Ingersoll-Ross type[4]

$$dZ_t = [a + S(t) - \gamma Z_t]dt + \sqrt{Z_t}dW_t$$

with $2a \in (1, \infty)$, the solution to the stochastic Hodgkin-Huxley system (SHH) is positive Harris recurrent (see [37, Theorem 2.7]). Moreover, they quantify the typical spiking behaviour: Almost surely, there are infinitely many spikes but also infinitely many periods of the signal in which no spike occurs (see [37, Theorem 2.8]). Harris recurrence then enables them to prove a Glivenko-Cantelli type Theorem for the interspike intervals (see [37, Theorem 2.9]).

**Example 1.2.** Considerable downsides of the deterministic Hodgkin-Huxley system are its analytical difficulty and complex stability structures (see for example [26], [22], [19]). Therefore, several similar but simplified neuron models have been proposed which provide a convenient approximation – at the price of inaccurate and biologically questionable modelling of the internal mechanism. Due to its simplicity, the most prominent and most widely used such example is the two-dimensional *FitzHugh-Nagumo model*

$$\begin{aligned}\dot{x} &= -x(x-a)(x-1) - y + S,\\ \dot{y} &= bx - cy,\end{aligned}$$

where $b \in (0, \infty)$, $c \in [0, \infty)$, and $a \in \mathbb{R}$ (see [41, equations (4.11) and (4.12)]). Here, $x$ is still the membrane potential, while $y$ is a so-called recovery variable. In contrast to the gating variables of the Hodgkin-Huxley model, this recovery variable lacks a clear biological interpretation. The *Morris-Lecar model* – as originally introduced in [52] – is a three-dimensional hybrid of the Hodgkin-Huxley model and the FitzHugh-Nagumo model that incorporates two internal variables. However, modern presentations (see for example [61]) also work with a version with only one internal variable which again plays the role of a recovery variable. Another three-dimensional reduced Hodgkin-Huxley system is looked into in [14] and [62].

Stochastic versions of such simplified systems have been studied for example in [4] and [5]. Noise is added in a different way than in (SDS), but the much more important

[4] Note that a Cox-Ingersoll-Ross type equation for $Z$ is not contained in our model, since we require its state space to be the full euclidean space and the volatility to be defined everywhere on it. This is only of major importance for Section 2.3, though.

difference to our model lies in the basic spirit: in these articles, the noise is assumed to be moderate such that essential (possibly chaotic) features of the deterministic system are preserved, while we show positive Harris recurrence which in a sense means that stochastic influences actually smoothen the dynamics.

**Example 1.3.** Systems of coupled oscillators are particularly intuitive Hamiltonian systems and several different stochastic models have been subject to research in the past (see e.g [24], [10], [60]). The special case that we will focus on is inspired by the model studied by Cuneo, Eckmann, and Poquet in [11].

Let us think of three rotors, each given by their angle $q_i(t) \in \mathbb{R}$ and momentum $p_i(t) \in \mathbb{R}$ at the time $t \in [0, \infty)$ for each $i \in \{1, 2, 3\}$. Assuming their respective masses to be all equal to 1 and not taking into account units, the laws of classical mechanics imply

$$\dot{q}_i = p_i \quad \text{for all } i \in \{1, 2, 3\}. \tag{1.7}$$

We suppose that these rotors are coupled in row, i.e.

$$\begin{aligned} \dot{p}_1 &= w_1(q_2 - q_1) - u_1(q_1), \\ \dot{p}_2 &= -[w_1(q_2 - q_1) + w_3(q_2 - q_3)] - u_2(q_2), \\ \dot{p}_3 &= w_3(q_2 - q_3) - u_3(q_3), \end{aligned} \tag{1.8}$$

where $w_i$ and $u_i$ are the derivatives of given interaction potentials $W_i \colon \mathbb{R} \to \mathbb{R}$ and pinning potentials $U_i \colon \mathbb{R} \to \mathbb{R}$ for all $i \in \{1, 2, 3\}$. A classical model is the one that arises if we let one or both of the outer rotors receive external torques and interact with Langevin type heat baths. In order to give a mathematical description of this, we fix

$$i \in \{1, 3\}$$

for the remainder of this paragraph. Applying an external time-dependent torque $S_i \colon [0, \infty) \to \mathbb{R}$ to the $i$-th rotor means expanding the equation for $p_i$ to

$$dp_i = [w_i(q_2 - q_i) - u_i(q_i)]\, dt + S_i dt,$$

which turns (1.7) and (1.8) into a system like (DDS). On top of that, we want to add interaction with a heat bath, i.e. for a temperature $\tau_i \in (0, \infty)$ and a dissipation constant $\delta_i \in (0, \infty)$, the equation for $p_i$ is further expanded to

$$\begin{aligned} dp_i &= [w_i(q_2 - q_i) - u_i(q_i)]\, dt + S_i dt - \delta_i p_i dt + \sqrt{2\delta_i \tau_i} dW_t^{(i)} \\ &= [w_i(q_2 - q_i) - u_i(q_i) - \delta_i p_i]\, dt + \left[ S_i dt + \sqrt{2\delta_i \tau_i} dW_t^{(i)} \right], \end{aligned} \tag{1.9}$$

where the last term in parentheses is the total sum of external influences. Following the spirit of the general model we study in this thesis, we may replace this term with the increments of a more general random perturbation of the torque: We take

$$dp_i = [w_i(q_2 - q_i) - u_i(q_i) - \delta_i p_i]\, dt + dZ_t^{(i)},$$

with

$$dZ_t^{(i)} = \left[S_i(t) + b_i(Z_t^{(i)})\right] dt + \sigma_i(Z_t^{(i)}) dW_t^{(i)}$$

for some volatility $\sigma_i\colon \mathbb{R} \to \mathbb{R}$ and a drift $b_i\colon \mathbb{R} \to \mathbb{R}$. What we end up with is indeed a degenerate diffusion with internal variables and randomly perturbed time-inhomogeneous deterministic input as in (SDS).

If only the first rotor in the chain receives an external input, the dimensions are $M = N = 1$ and $L = 5$, and we have to write

$$X = p_1, \quad Y = \begin{pmatrix} p_2 \\ p_3 \\ q_1 \\ q_2 \\ q_3 \end{pmatrix},$$

while the corresponding coefficient functions (which also define the exact pertaining deterministic system (DDS)) are to be written as

$$F(x, y) = F(x, y_3, y_4) = w_1(y_4 - y_3) - u_1(y_3) - \delta_1 x \tag{1.10}$$

and

$$G(x, y) = \begin{pmatrix} G_1(y_3, y_4, y_5) \\ G_2(y_4, y_5) \\ x \\ y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} -w_1(y_4 - y_3) - w_3(y_4 - y_5) - u_2(y_4) \\ w_3(y_4 - y_5) - u_3(y_5) \\ x \\ y_1 \\ y_2 \end{pmatrix}, \tag{1.11}$$

where $x \in \mathbb{R}$ and $y = (y_1, \ldots, y_5)^\top \in V = \mathbb{R}^5$.

If both of the outer rotors receive an external input, the dimensions are $M = N = 2$ and $L = 4$, and the corresponding variables of (SDS) become

$$X = \begin{pmatrix} p_1 \\ p_3 \end{pmatrix}, \quad Y = \begin{pmatrix} p_2 \\ q_1 \\ q_2 \\ q_3 \end{pmatrix},$$

and the respective coefficient functions are

$$F(x,y) = \begin{pmatrix} F_1(x_1,y_2,y_3) \\ F_2(x_2,y_3,y_4) \end{pmatrix} = \begin{pmatrix} w_1(y_3-y_2) - u_1(y_2) - \delta_1 x_1 \\ w_3(y_3-y_4) - u_3(y_4) - \delta_3 x_2 \end{pmatrix} \tag{1.12}$$

and

$$G(x,y) = \begin{pmatrix} G_1(y_2,y_3,y_4) \\ x_1 \\ y_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -w_1(y_3-y_2) - w_3(y_3-y_4) - u_2(y_3) \\ x_1 \\ y_1 \\ x_2 \end{pmatrix}, \tag{1.13}$$

where $x = (x_1,x_2)^\top \in \mathbb{R}^2$ and $y = (y_1,\dots,y_4)^\top \in V = \mathbb{R}^4$. The external influence can then be written as one variable $Z = (Z^{(1)}, Z^{(2)})^\top$ which solves

$$dZ_t = [S(t) + b(Z_t)]\,dt + \sigma(Z_t)dW_t,$$

where

$$S\colon [0,\infty) \to \mathbb{R}^2, \quad t \mapsto \begin{pmatrix} S_1(t) \\ S_2(t) \end{pmatrix},$$

and

$$b\colon \mathbb{R}^2 \to \mathbb{R}^2, \quad z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \mapsto \begin{pmatrix} b_1(z_1) \\ b_2(z_2) \end{pmatrix},$$

and

$$\sigma\colon \mathbb{R}^2 \to \mathbb{R}^{2\times 2}, \quad z \mapsto \begin{pmatrix} \sigma_1(z_1) & 0 \\ 0 & \sigma_2(z_2) \end{pmatrix}.$$

Admittedly, the numeration can be a bit confusing, but we present it here explicitly in order to demonstrate how changing from one-sided to two-sided input rearranges the roles of the equations from (1.7) and (1.8) (or (1.9)) when turning them into a system of the type (SDS). We always implicitly assume that the choice of the coefficient functions allows the respective system to have a unique non-exploding solution, which for example is guaranteed if $u_1$, $u_2$, $u_3$, $w_1$, $w_2$, $w_3$, $b$ and $\sigma$ are Lipschitz-continuous and the signal $S$ is bounded.

In the present text, this example will mainly serve as a hint at the potential of studying the general system (SDS) beyond the realm of neuroscience – which the theory displayed in this thesis is predominantly inspired by. While it serves as a nice illustrating example in Chapter 3, only certain parts of the results from Chapter 2 are applicable to this stochastic rotor system, but there are enough connections to spawn optimism for future work that may unify some of the approaches used in different settings.

Let us stress that the differences between the eight-dimensional system above with input on both ends of the chain and the six-dimensional system that is discussed in

[11] mainly lie in the role of the external equations and in time-inhomogeneity: Forcing the latter into our notational framework, the equations for $X$ and $Y$ would remain unaltered, while the external equations would read

$$dZ_t^{(i)} = S_i dt + \sqrt{2\delta_i \tau_i} dW_t^{(i)}, \quad i \in \{1, 2\},$$

where the coefficient functions do not depend on $Z_t^{(i)}$. Therefore, including these variables separately into the system is rendered obsolete. In addition to that, with the external torque in [11] being constant, their entire system is homogeneous in time – in contrast to ours. Another important difference is that while the system in [11] lets the angle take values in the torus $\mathbb{R} \bmod 2\pi$, the state space for (SDS) has to be a proper subset of $\mathbb{R}^{N+L+N}$ (without any topological identifications), and therefore we have the angles live in $\mathbb{R}$. This means that our model actually sees if a rotor has rotated multiple times.

For the six-dimensional system from [11], the authors' main result (Theorem 1.3) states the existence of smooth transition densities, unique existence of an invariant measure which again has a smooth density, and finally establishes a subgeometric convergence rate of the semi-group to the invariant measure.

**Example 1.4.** Let us also establish a custom-built toy example that will help us illustrate our results in a simple context. While the dimensions $N, L \in \mathbb{N}$ of the variables are arbitrary, the state space of the internal variables is chosen as $V = [0, \infty)^L$. We will now describe our choice of the coefficient functions

$$F\colon \mathbb{R}^N \times V \to \mathbb{R}^N \quad \text{and} \quad G\colon \mathbb{R}^N \times V \to \mathbb{R}^L,$$

such that

$$\begin{aligned} \dot{x} &= F(x, y) + S, \\ \dot{y} &= G(x, y), \end{aligned} \tag{1.14}$$

becomes a convenient toy example for a system of the type (DDS), which we can then turn into a degenerate diffusion with internal variables and randomly perturbed time-inhomogeneous deterministic input in the sense of (SDS).

Let

$$F\colon \mathbb{R}^N \times V \to \mathbb{R}^N, \quad (x, y) \mapsto -f(y)\big(x + h(x)\big) + j(x, y), \tag{1.15}$$

where $f \in C^\infty(V)$ obeys

$$\inf_{y \in V} f(y) \geq 1 \quad \text{and} \quad \sup_{y \in V} f(y) < \infty, \tag{1.16}$$

while $h \in C^\infty\big(\mathbb{R}^N;\mathbb{R}^N\big)$ satisfies

$$\begin{aligned} h(0_N) = 0_N, \quad \operatorname{supp} h \subset \overline{B_1(0_N)}, \\ x^\top h(x) \geq -h_0 \left|x\right|^2 \text{ for some } h_0 \in [0,1) \text{ and all } x \in \mathbb{R}^N, \end{aligned} \tag{1.17}$$

and $j \in C_b^\infty\big(\mathbb{R}^N \times V;\mathbb{R}^N\big)$ has the property

$$j(x,y) = 0_N \quad \text{for all } (x,y) \in B_1(0_N) \times V. \tag{1.18}$$

For $x \in B_1(0_N)$ the function $F$ basically drives the $x$-variable back to the origin with a force that is quantified by $f(y)$. The function $h$ is simply a small perturbation around zero that we will need for technical reasons later on,[5] but that has no significant influence on the basic interpretation of $F$. One possible example of such a function is given by

$$h\colon \mathbb{R}^N \to \mathbb{R}^N, \quad x \mapsto -h_0 e^{-\frac{1}{1-|x|^2}} 1_{B_1(0_N)}(x) \cdot x.$$

When $x \notin B_1(0_N)$, the function $j$ can cause $F$ to induce a multitude of different local behaviours.

For the internal mechanism, we consider

$$G\colon \mathbb{R}^N \times V \to \mathbb{R}^L, \quad (x,y) \mapsto -y + g(x), \tag{1.19}$$

with some Lipschitz continuous and bounded $g \in C^\infty\big(\mathbb{R}^N;(0,\infty)^L\big)$. If $g$ was entirely absent, $G$ would simply let the internal variables decay exponentially. Hence, we can interpret the term $g(x)$ as a source that is determined by the current state of the $x$-variables.

Let us briefly comment on existence and uniqueness of a solution to (1.4) with $F$ and $G$ as in (1.15) and (1.19). In this situation, $F$ and $G$ are obviously locally Lipschitz continuous. Furthermore, for any signal function $S \in C\big([0,\infty);\mathbb{R}^N\big)$ we have linear growth

$$\begin{aligned} |F(x,y) + S(t)| + |G(x,y)| &\leq \|f\|_\infty \big(1 + \operatorname{Lip}(h)\big) |x| + \operatorname{Lip}(j) \left|(x,y)\right| + |S(t)| \\ &\quad + |y| + \operatorname{Lip}(g) |x| \\ &\leq \mathrm{cst} \left|(x,y)\right| + |S(t)| \end{aligned}$$

for all $(x,y) \in \mathbb{R}^N \times V$ and $t \in [0,\infty)$. Thus, with these choices for $F$ and $G$ the system (1.14) has a unique global solution – provided the $y$-component cannot leave $V$. This is indeed the case, as variation of constants yields

$$y(t) = e^{-t} y(0) + \int_0^t e^{-(t-s)} g(x(s)) ds \quad \text{for all } t \in [0,\infty),$$

[5]To be precise, $h$ will be needed in Example 2.53 in order to provide non-vanishing second derivatives of $F$ with respect to $x$ in $x^* = 0_N$.

which obviously stays in $V$ whenever it starts there, since all of the components of the function $g$ are positive. Existence and uniqueness of a strong non-explosive solution of the corresponding stochastic system in the sense of (SDS) can be checked in the same way as for the stochastic Hodgkin-Huxley system (provided the drift and volatility of the external equation are Lipschitz continuous).

Taking $N = 1$, we may also replace the internal mechanism by the one described by

$$G \colon \mathbb{R} \times V \to \mathbb{R}^L, \quad (x, y) \mapsto -y + g(x, y), \tag{1.20}$$

with

$$g(x, y) = \begin{pmatrix} g_1(x) \\ g_2(y_1) \\ \vdots \\ g_L(y_{L-1}) \end{pmatrix},$$

where $g_1 \in C^\infty(\mathbb{R})$ and $g_2, \dots, g_L \in C^\infty([0, \infty))$ are positive Lipschitz continuous bounded functions. Unique global existence of a solution with values in $\mathbb{R} \times V$ remains valid and is shown similarly to the above.

# Chapter 2

# Positive Harris recurrence of $\mathbb{X}$

The aim of this chapter is to find appropriate conditions under which we can apply Theorem 2.2 from [37] (see Theorem 2.3 below) in order to establish positive Harris recurrence for the Markov process $\mathbb{X}$ that solves the diffusion equation (SDS). Section 2.1 is devoted to presenting and explaining this Theorem, while Sections 2.2 through 2.4 will each deal with one of its main assumptions. Along the way, we will discuss applications to the examples introduced in Section 1.2.

## 2.1 General strategy

Let us briefly recall some basic notions about recurrence. A time-homogeneous Markov process is called *Harris recurrent*, if for some $\sigma$-finite measure $\varphi$ on its state space, every $\varphi$-positive set has infinite occupation time (almost surely for every starting point). If this is the case, there is a unique (up to a multiplicative constant) invariant measure $\mu$, and it can replace $\varphi$ in the above property. If $\mu$ is finite, it can be normalised to a unique invariant probability measure, and the process is termed *positive* Harris recurrent. Otherwise, it is called null-recurrent. Positive Harris recurrence basically means that returning times of $\mu$-positive sets have finite expectation.

Positive Harris recurrence is the foundation of a wide class of Limit Theorems for Markov processes, most prominently the Ratio Limit Theorem. Classical references for the theory of recurrence in the sense of Harris include [25], [58], and [54] in the discrete-time setting and [1] in the continuous-time case. Almost all of the most important results feature the invariant distribution $\mu$ as a central object.

As the notion of an invariant measure becomes far more complicated in the time-inhomogeneous setting, one is inclined to get rid of time-dependent dynamics first, if one is interested in studying recurrence properties of a Markov process. Therefore, it is important to notice once again that the only time-dependence of (SDS) is contained in the signal $S$. Let us assume that

**(P) Periodic signal:** $S$ is periodic with periodicity $T \in (0, \infty)$.

Then the entire drift of the diffusion $\mathbb{X}$ solving (SDS) is $T$-periodic in time, and consequently its transition semi-group $(P_{s,t})_{t>s\geq 0}$ under $\mathbb{P}$ that is defined by

$$P_{s,t}(x, A) := \mathbb{P}(\mathbb{X}_t \in A \,|\, \mathbb{X}_s = x) \quad \text{for all } x \in \mathtt{E},\ A \in \mathcal{B}(\mathtt{E}), \text{ and } t > s \geq 0, \tag{2.1}$$

has the property

$$P_{s+kT,t+kT} = P_{s,t} \quad \text{for all } t > s \geq 0 \text{ and } k \in \mathbb{N}. \tag{2.2}$$

This allows us to derive several other Markov processes from this semi-group which are homogeneous in time and whose ergodic properties we can study with classical methods.

**Lemma 2.1.** *Let (P) hold.*

1. *The* grid chain
$$\mathbb{X}^{\mathbf{gr}} := (\mathbb{X}^{\mathbf{gr}}_k)_{k\in\mathbb{N}_0},$$
*defined by*
$$\mathbb{X}^{\mathbf{gr}}_k := \mathbb{X}_{kT} \quad \textit{for all } k \in \mathbb{N}_0,$$
*is an* $\mathtt{E}$*-valued time-homogeneous discrete-time Markov process with one-step transition kernel* $P_{0,T}$.

2. *The* path segment chain
$$\mathbb{X}^{\mathbf{ps}} := (\mathbb{X}^{\mathbf{ps}}_k)_{k\in\mathbb{N}_0},$$
*defined by*
$$\begin{aligned} \mathbb{X}^{\mathbf{ps}}_k &:= \big([0,T] \ni t \mapsto \mathbb{X}_{(k-1)T+t}\big) \quad \textit{for all } k \in \mathbb{N}, \\ \mathbb{X}^{\mathbf{ps}}_0 &\in C([0,T];\mathtt{E}) \ \textit{arbitrary with } \mathbb{X}^{\mathbf{ps}}_0(T) = \mathbb{X}_0, \end{aligned}$$
*is a* $C([0,T];\mathtt{E})$*-valued time-homogeneous discrete-time Markov process.*

3. *The* time-space process
$$\mathbb{X}^{\mathbf{ts}} := \big(\mathbb{X}^{\mathbf{ts}}_t\big)_{t\in[0,\infty)},$$
*defined by*
$$\mathbb{X}^{\mathbf{ts}}_t := (t \bmod T, \mathbb{X}_t) \quad \textit{for all } t \in [0,\infty),$$
*is a* $[0,T) \times \mathtt{E}$*-valued time-homogeneous continuous-time Markov process.*

*Proof.* Since (P) implies (2.2), these assertions are evident. □

**Remark 2.2.** Of course, if the drift was not periodic, we could still define the time-space process as $(t, \mathbb{X}_t)_{t\in[0,\infty)}$ on $[0, \infty) \times \mathtt{E}$, and it would still be homogeneous in time (we actually do so in Section 2.3, and this is in fact consistent with the definition above, if we think of aperiodicity as the case $T = \infty$). However, with $t$ being its first component, there is obviously no chance of it being Harris recurrent.

Our aim is to apply the following Theorem which is a slight reformulation of Theorem 2.2 from [37] and the remark immediately below it.

**Theorem 2.3** (Höpfner, Löcherbach, Thieullen (2016))**.** *Let $\mathbb{X}$ be the unique strong solution of* (SDS) *with deterministic starting point $\mathbb{X}_0 \in \mathtt{E}$. Assume that the following conditions are satisfied:*

1. *There is a strictly increasing sequence $(G_n)_{n\in\mathbb{N}}$ of bounded open convex subsets of $\mathtt{E}$ such that:*

   *(a) The sequence of compacts $C_n := \overline{G_n}$ increases to the full space $\mathtt{E}$.*

   *(b) If $\mathbb{X}_0 \in \partial\mathtt{E}$, the process $\mathbb{X}$ immediately enters* $\mathrm{int}(\mathtt{E})$.

   *(c) If $\mathbb{X}_0 \in C_n \setminus G_{n+1}$ for some $n \in \mathbb{N}$, the process $\mathbb{X}$ immediately enters $G_{n+1}$.*

   *(d) For any $\mathbb{X}_0 \in \mathtt{E}$ we have*

   $$\inf\{t > 0 \,|\, \mathbb{X}_t \notin C_n\} \xrightarrow{n\to\infty} \infty$$

   *$\mathbb{P}$-almost surely.*

2. *Condition (P) holds: The signal $S$ is $T$-periodic for some $T \in (0, \infty)$.*

3. *All of the coefficient functions $F$, $G$, $S$, $b$, and $\sigma$ have derivatives of any order with respect to any of their variables.*

4. *The process $\mathbb{X}$ possesses a Lyapunov function.*

5. *There is a point in* $\mathrm{int}(\mathtt{E})$ *which is attainable in a sense of deterministic control and at which the local weak Hörmander condition holds.*

*Then the grid chain $\mathbb{X}^{\mathbf{gr}}$, the path segment chain $\mathbb{X}^{\mathbf{ps}}$, and the time-space process $\mathbb{X}^{\mathbf{ts}}$ are all positive Harris recurrent.*

*Proof.* A detailed proof can be found in [37]. In order to convey its central ideas and to illustrate the distinct role of each assumption, we will explain its basic route along the way whenever we discuss the pertaining condition (see Remarks 2.8, 2.20, and 2.36). $\square$

Assumption 1 of this Theorem is basically non-explosiveness and sufficiently nice boundary behaviour which together avoid problems with localisation arguments. Assumption 2 is of course needed for the conclusion to even make sense at all, while assumption 3 is more or less inevitable if we want to speak about Hörmander's condition in assumption 5. Therefore, one has to think of assumptions 4 and 5 as the crucial ones.

Let us discuss assumption 1 in the context of the examples that were introduced in Section 1.2. Non-explosiveness has already been commented on there.

**Example 2.4.** For the stochastic Hodgkin-Huxley system (SHH), assumption 1 of Theorem 2.3 can be fulfilled by choosing

$$G_n := (-n, n) \times (0, 1)^3 \times (-n, n) \quad \text{for all } n \in \mathbb{N}.$$

Strict positivity of the functions $\alpha_1, \alpha_2, \alpha_3$ and $\beta_1, \beta_2, \beta_3$ ensures that the internal variables immediately enter $(0, 1)^3$ when started in $[0, 1]^3$ (compare [37, bottom of page 533]).

**Example 2.5.** In the toy example (1.14), we can verify assumption 1 of Theorem 2.3 in a similar way: The sequence defined by

$$G_n := (-n, n)^N \times (0, n)^L \times (-n, n)^N \quad \text{for all } n \in \mathbb{N}$$

does the job, since the functions $g, g_1, \ldots g_L$ from (1.19) and (1.20) are strictly positive.

**Example 2.6.** Since the chain of coupled oscillators from Example 1.3 has the entire $\mathbb{R}^{N+L+N}$ as its state space $\mathtt{E}$, assumption 1 of Theorem 2.3 is trivially fulfilled with

$$G_n := (-n, n)^{N+L+N} \quad \text{for all } n \in \mathbb{N}.$$

Of course, we still owe the reader an explanation of what exactly we mean by a Lyapunov function, by attainability in a sense of deterministic control and by the local weak Hörmander condition. This will be the content of the following three sections which will each deal with one of these topics independently: Section 2.2 presents simple criteria under which a Lyapunov function can be constructed, Section 2.3 is devoted to finding points that are attainable in a sense of deterministic control, and finally Section 2.4 plays through different scenarios in which verifiable and handy sufficient conditions for the local weak Hörmander condition can be formulated. Each of these sections can be read independently, and we try to work under as little assumptions as necessary (within reason). For instance, when dealing with deterministic control in Section 2.3, we will work without infinitely differentiable coefficient functions and even without periodicity of the signal – even though both are needed anyway for Theorem 2.3.

## 2.2 Lyapunov function

The foundation of Theorem 2.3 is laid by the idea that is outlined for example in [50]. The authors discuss conditions that imply regular returning of a Markov chain into certain sets, provided the existence of a so-called Lyapunov function. As their technique is rooted in the theory of time-homogeneous discrete-time Markov processes, the periodic nature of the underlying deterministic signal becomes crucial in this section. Throughout it, we will assume that (P) holds, i.e. $S$ is periodic with periodicity $T \in (0,\infty)$. Note that since $S$ is also continuous, this necessarily means that $S$ is bounded.

First, let us introduce and explain our notion of a Lyapunov function, which is essentially taken from [50]. Recall from Lemma 2.1 that if $(P_{s,t})_{t>s\geq 0}$ is the transition semi-group of $\mathbb{X}$, the grid chain $\mathbb{X}^{\mathbf{gr}} = (\mathbb{X}_{kT})_{k\in\mathbb{N}_0}$ evolves according to the transition kernel $P_{0,T}$.

**Definition 2.7.** Let $\Phi\colon \mathtt{E} \to [1,\infty)$ be a measurable function. If there is a compact set $K \subset \mathtt{E}$ such that

1. $\inf_{\mathtt{E}\setminus K}(1 - P_{0,T})\Phi > 0$,
2. $\Phi(\omega_n) \xrightarrow{n\to\infty} \infty$ for each sequence $(\omega_n)_{n\in\mathbb{N}} \subset \mathtt{E}$ with $|\omega_n| \xrightarrow{n\to\infty} \infty$,
3. $\sup_K P_{0,T}\Phi < \infty$,

we call $\Phi$ a *Lyapunov function for* $\mathbb{X}$.

**Remark 2.8.** We want to interpret this definition and explain its role in the proof of Theorem 2.3. The first condition in Definition 2.7 yields that outside of $K$, the process $\Phi(\mathbb{X}^{\mathbf{gr}})$ behaves like a non-negative strict super-martingale (which has to converge almost surely by [59, Corollary II.2.11]) which, together with the second condition, implies that $\mathbb{X}^{\mathbf{gr}}$ will almost surely enter $K$ in finite time and, thanks to the Markov property, will visit it infinitely often. The third condition in Definition 2.7 can then be used to establish a bound for the expected duration of excursions away from $K$. In order to extend these properties to other sets, we will use the methods from Sections 2.3 and 2.4.

In order to construct a Lyapunov function for the process $\mathbb{X}$, we will have to make sure that the coefficient functions in its diffusion equation (SDS) behave favourably. We consider the following conditions:

**(L1) Backwards drift condition for $X$:** There are a function $\varphi \in C^2\big(\mathbb{R}^N; [1,\infty)\big)$ and an $r(\varphi) \in (1,\infty)$ with

$$\varphi(x) = |x| \quad \text{for all } x \in \mathbb{R}^N \setminus B_{r(\varphi)}(0_N),$$

and there are positive constants $c_1, \tilde{c}_1 \in (0, \infty)$ such that for all $(x, y) \in \mathbb{R}^N \times V$ we have

$$F(x, y)^\top \nabla\varphi(x) \leq -c_1\varphi(x) + \tilde{c}_1.$$

**(L2) Backwards drift condition for $Y$:** There are positive constants $c_2, \tilde{c}_2 \in (0, \infty)$ such that for all $(x, y) \in \mathbb{R}^N \times V$ we have

$$G(x, y)^\top y \leq -c_2 |y|^2 + \tilde{c}_2.$$

**(L3) Backwards drift condition for $Z$:** There are positive constants $c_3, \tilde{c}_3 \in (0, \infty)$ such that for all $z \in \mathbb{R}^N$ we have

$$b(z)^\top z \leq -c_3 |z|^2 + \tilde{c}_3.$$

**(L4) Subquadratic growth of the drift of $Z$:** We have

$$\frac{|b(z)|}{|z|^2} \xrightarrow{|z| \to \infty} 0.$$

**(L5) Sublinear growth of the volatility of $Z$:** We have

$$\frac{\operatorname{tr}(\sigma\sigma^\top(z))}{|z|^2} \xrightarrow{|z| \to \infty} 0.$$

**(L6) Bounded volatility of $Z$:** We have

$$\|\sigma_{i,j}\|_\infty < \infty \quad \text{for all } i \in \{1, \ldots, N\} \text{ and } j \in \{1, \ldots, M\}.$$

If condition (L1) holds and $(x, y) \in \big(\mathbb{R}^N \setminus B_{r(\varphi)}(0_N)\big) \times V$, the respective bound can be rewritten as

$$F(x, y)^\top \frac{x}{|x|} \leq -c_1 |x| + \tilde{c}_1.$$

This means that up to some modification around the origin $0_N$ and up to an additive constant, the vector field $F(\cdot, y)$ acts on $X$ as a back-driving force pointing towards the origin. This force is at least of linear order with respect to the position in $\mathbb{R}^N$, where the constant factor is uniform with respect to $y \in V$. Similarly, (L2) signifies that $G(x, \cdot)$ uniformly drives $Y$ back to the origin $0_L$. Finally, (L3) does the same for the drift of the $Z$-component, where no uniformity is present, since $b$ does not depend on any variables other than $z \in \mathbb{R}^N$ itself. Conditions that are very similar to (L1) - (L3) are used, for example, in [64]. Our variants are very strong, but they are simple and can be verified in Examples 2.16 and 2.17 below. The conditions (L4) - (L6) are growth conditions for the external equation, where (L6) is obviously much stronger than (L5). Note also that even though (L3) and (L4) seem very similar, (L3) makes a statement about the effect of the drift $b$ in specific directions, while (L4) is a purely asymptotic growth condition.

**Remark 2.9.** It may appear a bit counter-intuitive that we allow additive constants in (L1) - (L3) but require the respective inequalities to hold globally and not only away from the origin. So, for the moment, let us assume that

$$\begin{gathered}\text{there are } c_1, r_1 \in (0,\infty) \text{ such that for all } (x,y) \in (\mathbb{R}^N \times V) \setminus B_{r_1}(0_{N+L})\text{:}\\ F(x,y)^\top \nabla\varphi(x) \le -c_1\varphi(x).\end{gathered} \tag{2.3}$$

Since $F$, $\varphi$, and $\nabla\varphi$ are continuous and hence locally bounded, we have

$$C := \sup_{(x,y)\in B_{r_1}(0_{N+L})} \big(|F(x,y)| + |\varphi(x)| + |\nabla\varphi(x)|\big) < \infty.$$

Turning to the global situation, we see that for all $(x,y) \in \mathbb{R}^N \times V$ the property (2.3) implies

$$\begin{aligned} F(x,y)^\top \nabla\varphi(x) &= F(x,y)^\top \nabla\varphi(x) \cdot 1_{B_{r_1}(0_{N+L})^c}(x,y) + F(x,y)^\top \nabla\varphi(x) \cdot 1_{B_{r_1}(0_{N+L})}(x,y) \\ &\le -c_1\varphi(x) 1_{B_{r_1}(0_{N+L})^c}(x,y) + C^2 \\ &= -c_1\varphi(x)\big(1 - 1_{B_{r_1}(0_{N+L})}(x,y)\big) + C^2 \\ &\le -c_1\varphi(x) + c_1 C + C^2 \end{aligned}$$

which is (L1) with $\tilde{c}_1 = c_1 C + C^2$. Therefore, (2.3) in fact implies (L1). Analogously, we see that the existence of $c_2, r_2 \in (0,\infty)$ such that

$$G(x,y)^\top y \le -c_2 |y|^2 \quad \text{for all } (x,y) \in (\mathbb{R}^N \times V) \setminus B_{r_2}(0_{N+L}) \tag{2.4}$$

is sufficient for (L2). Note that for $x \in \mathbb{R}^N \setminus B_{r(\varphi)}(0_N)$ the condition (2.3) is entirely analogous to (2.4).

As above, we also see that the existence of $c_3, r_3 \in (0,\infty)$ with

$$b(z)^\top z \le -c_3 |z|^2 \quad \text{for all } z \in \mathbb{R}^N \setminus B_{r_3}(0_N) \tag{2.5}$$

is sufficient for (L3). The property (2.5) is a (very strong) variant of what is often referred to as Veretennikov's drift condition (see [64, Equation (6)]).

**Remark 2.10.** Consider, as in (L1), a function $\varphi \in C^2\big(\mathbb{R}^N; [1,\infty)\big)$ such that for some $r(\varphi) \in (1,\infty)$ it coincides with $|\cdot|$ outside of $B_{r(\varphi)}(0_N)$. Then all of its first and second order partial derivatives are bounded globally: Locally, this is trivial, and for $x \in \mathbb{R}^N \setminus B_{r(\varphi)}(0_N)$ we have

$$\partial_{x_i}\varphi(x) = \frac{x_i}{|x|}, \quad \partial_{x_j}\partial_{x_i}\varphi(x) = \frac{1}{|x|}\left(1_{i=j} - \frac{x_i x_j}{|x|^2}\right) \quad \text{for all } i,j \in \{1,\dots,N\},$$

which are bounded when $x$ is separated from the origin.

In the sequel, we use the notation

$$K_r := \overline{B_r(0_{N+L+N})} \cap \mathtt{E} \quad \text{for any } r \in (0, \infty). \tag{2.6}$$

Note that if $V$ is closed, the same is true for $\mathtt{E} = \mathbb{R}^N \times V \times \mathbb{R}^N$ and hence also for $K_r$ which is therefore compact. Let $\Phi \in C^2(\mathtt{E}; [1, \infty))$ such that

$$\Phi(x, y, z) := \varphi(x) + |y|^2 + |z|^2 \quad \text{for all } (x, y, z) \in \mathtt{E} \setminus K_2, \tag{2.7}$$

where $\varphi$ is the function from (L1). In Theorem 2.11 below, we will show that $\Phi$ is a Lyapunov function for $\mathbb{X}$.

**Theorem 2.11.** *Assume that the state space $V$ of the internal variables is closed and that (P) and (L1) - (L4) hold. Beyond that, assume that one of the following conditions is fulfilled:*

*(i) (L6) holds.*

*(ii) (L5) holds and $\sigma\sigma^\top(z)$ is a diagonal matrix for all $z \in \mathbb{R}^N$.*

*Then the function $\Phi$ from* (2.7) *is a Lyapunov function for $\mathbb{X}$.*

*Proof.* 1.) For the function $\Phi$ from (2.7), we trivially have

$$\Phi(x, y, z) \xrightarrow{|(x,y,z)| \to \infty} \infty, \tag{2.8}$$

so the second property from Definition 2.7 is fulfilled. The rest of this proof is therefore devoted to finding a suitable compact set $K \subset \mathtt{E}$ such that the remaining conditions

$$\inf_{\mathtt{E} \setminus K} (1 - P_{0,T})\Phi > 0 \tag{2.9}$$

and

$$\sup_K P_{0,T}\Phi < \infty \tag{2.10}$$

are satisfied as well. We will do this by following a well-known general route (compare [49, (2.2) - (2.4)]).

2.) We see directly from (SDS) that the generator of the time inhomogeneous Markov process $\mathbb{X}$ satisfies

$$\begin{aligned} L_t f(x, y, z) = {} & F(x, y)^\top \nabla_x f(x, y, z) + (S(t) + b(z))^\top (\nabla_x f(x, y, z) + \nabla_z f(x, y, z)) \\ & + G(x, y)^\top \nabla_y f(x, y, z) \\ & + \frac{1}{2} \sum_{i,j=1}^{N} (\sigma\sigma^\top)_{i,j}(z) \left( \partial_{x_i}\partial_{x_j} + \partial_{x_i}\partial_{z_j} + \partial_{z_i}\partial_{x_j} + \partial_{z_i}\partial_{z_j} \right) f(x, y, z) \end{aligned}$$

for all $f \in C^2(\mathtt{E})$ and $(t, x, y, z) \in [0, \infty) \times \mathtt{E}$. If we let $(t, x, y, z) \in [0, \infty) \times (\mathtt{E} \setminus K_2)$, plugging in the function $\Phi$ from (2.7) yields

$$\begin{aligned} L_t\Phi(x, y, z) = F(x, y)^\top \nabla\varphi(x) + (S(t) + b(z))^\top (\nabla\varphi(x) + 2z) \\ + 2G(x, y)^\top y + \frac{1}{2} \sum_{i,j=1}^{N} (\sigma\sigma^\top)_{i,j}(z)\partial_{x_i}\partial_{x_j}\varphi(x) + \operatorname{tr}(\sigma\sigma^\top(z)). \end{aligned} \tag{2.11}$$

Our intermediate goal is to prove that

$$\limsup_{|(x,y,z)|\to\infty} \sup_{t\in[0,\infty)} \frac{L_t\Phi(x, y, z)}{\Phi(x, y, z)} < 0. \tag{2.12}$$

Setting

$$C := \min\{c_1, 2c_2, 2c_3\} \quad \text{and} \quad \tilde{C} := \max\{\tilde{c}_1, 2\tilde{c}_2, 2\tilde{c}_3\}$$

with the constants from (L1), (L2), and (L3), we obtain from (2.11) that

$$\begin{aligned} L_t\Phi(x, y, z) &\le -C\Phi(x, y, z) + \tilde{C} + (S(t) + b(z))^\top \nabla\varphi(x) + 2S(t)^\top z \\ &\quad + \frac{1}{2} \sum_{i,j=1}^{N} (\sigma\sigma^\top)_{i,j}(z)\partial_{x_i}\partial_{x_j}\varphi(x) + \operatorname{tr}(\sigma\sigma^\top(z)) \\ &\le -C\Phi(x, y, z) + \tilde{C} + \|S\|_\infty (\|\nabla\varphi\|_\infty + 2\,|z|) + |b(z)|\,\|\nabla\varphi\|_\infty \\ &\quad + \frac{1}{2} \sum_{i,j=1}^{N} (\sigma\sigma^\top)_{i,j}(z)\partial_{x_i}\partial_{x_j}\varphi(x) + \operatorname{tr}(\sigma\sigma^\top(z)). \end{aligned}$$

Thanks to (2.8), this implies

$$\limsup_{|(x,y,z)|\to\infty} \sup_{t\in[0,\infty)} \frac{L_t\Phi(x, y, z)}{\Phi(x, y, z)} \le -C + \limsup_{|(x,y,z)|\to\infty} R(x, y, z)$$

with

$$R(x, y, z) = \frac{2\,\|S\|_\infty\,|z| + |b(z)|\,\|\nabla\varphi\|_\infty + \frac{1}{2}\sum_{i,j=1}^{N} (\sigma\sigma^\top)_{i,j}(z)\partial_{x_i}\partial_{x_j}\varphi(x) + \operatorname{tr}(\sigma\sigma^\top(z))}{\Phi(x, y, z)}.$$

Showing that this remainder vanishes for $|(x, y, z)| \to \infty$ will prove (2.12). First of all,

$$\frac{|z|}{\Phi(x, y, z)} \quad \text{and} \quad \frac{|b(z)|}{\Phi(x, y, z)} \tag{2.13}$$

both go to zero when $|(x, y, z)| \to \infty$, which is a consequence of (2.7) and (L4). Indeed: If $\big((x_n, y_n, z_n)\big)_{n\in\mathbb{N}} \subset \mathtt{E}$ is a sequence with $|(x_n, y_n, z_n)| \xrightarrow{n\to\infty} \infty$ and $|z_n| \xrightarrow{n\to\infty} \infty$, the numerators in these fractions are dominated by the denominator thanks to the definition of $\Phi$ or thanks to (L4), respectively. If $(|z_n|)_{n\in\mathbb{N}}$ is bounded, (L4) is of no help, but then the numerators have to remain bounded as well, as they depend continuously on

$z$ (and only on $z$). Because of (2.8), the denominator explodes either way, so this case is also covered. Treating the remaining terms

$$\frac{\sum_{i,j=1}^{N}(\sigma\sigma^\top)_{i,j}(z)\partial_{x_i}\partial_{x_j}\varphi(x)}{\Phi(x,y,z)} \quad \text{and} \quad \frac{\mathrm{tr}(\sigma\sigma^\top(z))}{\Phi(x,y,z)} \tag{2.14}$$

is the only part of this proof where the different assumptions (i) and (ii) actually make a difference. In the setting (i), condition (L6) holds, i.e. the functions $\sigma_{i,j}$ are bounded. The partial derivatives of $\varphi$ are bounded as well (see Remark 2.10), and consequently both of the fractions in (2.14) have bounded numerators. As a result, the exploding denominator makes sure that these terms vanish in the limit. In the setting (ii), the matrix $\sigma\sigma^\top(z)$ is diagonal, and hence the numerator of the first term from (2.14) can be bounded by some constant (arising from the bounds for the derivatives of $\varphi$) times $\mathrm{tr}(\sigma\sigma^\top(z))$. It therefore remains to prove that the second fraction from (2.14) goes to zero, but this follows from (L5) by the same argument we used for the terms in (2.13).

3.) Now that we have proved (2.12) for any of our assumptions, we can find positive real numbers $C_1, r \in (0,\infty)$ such that

$$L_t\Phi(x,y,z) \leq -C_1\Phi(x,y,z) \quad \text{for all } (x,y,z) \in \mathtt{E}\setminus K_r \text{ and } t \in [0,\infty).$$

As $L_\cdot\Phi$ is obviously continuous and depends on time only through the bounded signal, we have

$$C_2 := \sup_{(x,y,z)\in K_r}\ \sup_{t\in[0,\infty)} |C_1\Phi(x,y,z) + L_t\Phi(x,y,z)| < \infty,$$

and we finally arrive at the global property

$$L_t\Phi(x,y,z) \leq -C_1\Phi(x,y,z) + C_2 \quad \text{for all } (x,y,z) \in \mathtt{E} \text{ and } t \in [0,\infty). \tag{2.15}$$

Applying the time dependent version of Itō's formula to the process $\big(e^{C_1 t}\Phi(\mathbb{X}_t)\big)_{t\in[0,\infty)}$ and then using (2.15) yields

$$\begin{aligned}
e^{C_1 t}\Phi(\mathbb{X}_t) - \Phi(\mathbb{X}_0) &= \int_0^t C_1 e^{C_1 s}\Phi(\mathbb{X}_s)ds + \int_0^t e^{C_1 s}L_s\Phi(\mathbb{X}_s)ds \\
&\quad + \sum_{i=1}^{N}\sum_{j=1}^{M}\int_0^t e^{C_1 s}\left(\partial_{x_i}\Phi(\mathbb{X}_s) + \partial_{z_i}\Phi(\mathbb{X}_s)\right)\sigma_{i,j}(Z_s)dW_s^{(j)} \\
&\leq C_2\int_0^t e^{C_1 s}ds + \sum_{i=1}^{N}\sum_{j=1}^{M}\int_0^t e^{C_1 s}\big(\partial_{x_i}\varphi(X_s) + 2Z_s^{(i)}\big)\sigma_{i,j}(Z_s)dW_s^{(j)} \\
&\leq \frac{C_2}{C_1}e^{C_1 t} + \sum_{i=1}^{N}\sum_{j=1}^{M}\int_0^t e^{C_1 s}\big(\partial_{x_i}\varphi(X_s) + 2Z_s^{(i)}\big)\sigma_{i,j}(Z_s)dW_s^{(j)}.
\end{aligned}$$

Choosing $\mathbb{X}_0 = (x,y,z) \in \mathtt{E}$, $t = T$, and taking the expected value (after localisation with level crossing times of $|Z|$) leads to

$$e^{C_1 T}P_{0,T}\Phi(x,y,z) - \Phi(x,y,z) = \mathbb{E}_{\mathbb{P}}\big[e^{C_1 T}\Phi(\mathbb{X}_T) - \Phi(\mathbb{X}_0)\big] \leq \frac{C_2}{C_1}e^{C_1 T}.$$

Rearranging the terms yields

$$(1 - P_{0,T})\Phi(x, y, z) \geq (1 - e^{-C_1 T})\Phi(x, y, z) - \frac{C_2}{C_1}. \tag{2.16}$$

Thanks to (2.8) and the strict positivity of $1 - e^{-C_1 T}$, there is some $R \in (0, \infty)$ such that for $(x, y, z)$ outside of the compact set $K_R$ (see (2.6)) the right hand side of (2.16) is positive and bounded away from zero. We have thus acquired the desired strict inequality (2.9) with $K := K_R$. Rearranging (2.16) once again and using that $\Phi$ is continuous and therefore locally bounded, we also obtain

$$\sup_K P_{0,T}\Phi \leq \frac{C_2}{C_1} + e^{-C_1 T} \sup_K \Phi < \infty$$

which takes care of (2.10). Thus, the proof is completed. □

**Remark 2.12.** At first glance, one might wonder why we do not take

$$\Phi(x, y, z) := |x|^2 + |y|^2 + |z|^2 \quad \text{for all } (x, y, z) \in \mathtt{E} \setminus K_2, \tag{2.17}$$

and use a simpler variant of (L1) that is directly analogous to (L2) and does not involve the auxiliary function $\varphi$. The reason is that in the second step of the proof of Theorem 2.11 (when showing that $R(x, y, z)$ vanishes in the limit) it is crucial that the first order partial derivatives of $\Phi$ with respect to the $x$-variables remain bounded (compare Remark 2.10), which is obviously not the case for $\Phi$ as defined in (2.17). This problem could be overcome by taking

$$\Phi(x, y, z) := |x| + |y|^2 + |z|^2 \quad \text{for all } (x, y, z) \in \mathtt{E} \setminus K_2,$$

but then $\Phi \notin C^2(\mathtt{E})$, since this expression is not differentiable with respect to the $x$-variables in any $(0_N, y, z) \in \mathtt{E} \setminus K_2$. We circumvent this problem by introducing the function $\varphi \in C^2\big(\mathbb{R}^N; [1, \infty)\big)$ which is basically a smoothened version of the euclidean norm.

We can also write down the following variant of Theorem 2.11 which states that we can simplify the Lyapunov function and omit the condition (L2), provided that the state space $V$ of the internal variables is not only closed but also bounded.

**Theorem 2.13.** *Assume that the state space $V$ of the internal variables is compact and that (P), (L1), (L3), and (L4) hold. Beyond that, assume that one of the following conditions is fulfilled:*

*(i) (L6) holds.*

*(ii) (L5) holds and $\sigma\sigma^\top(z)$ is a diagonal matrix for all $z \in \mathbb{R}^N$.*

*Let* $\Psi \in C^2(\mathtt{E}; [1,\infty))$ *such that*

$$\Psi(x,y,z) := \varphi(x) + |z|^2 \quad \textit{for all } (x,y,z) \in \mathtt{E} \setminus K_2,$$

*where* $\varphi$ *is the function from (L1). Then* $\Psi$ *is a Lyapunov function for* $\mathbb{X}$.

*Proof.* Going into the proof of Theorem 2.11 and replacing $\Phi$ with $\Psi$, we see that we can use the exact same arguments as before – with no need for any condition on $G$ due to the absence of any $y$-dependence in $\Psi$. Note that the second property from Definition 2.7 remains true in spite of this, as with $V$ being compact, the $y$-component cannot grow indefinitely anyway. □

**Remark 2.14.** We want to comment on and contextualise the conditions that are imposed on the coefficients of the external equation in Theorems 2.11 and 2.13. Using the same general approach as in the proof of Theorem 2.11, it is easy to see that $\Phi(z) = |z|^2$ is a Lyapunov function for $Z$, if

$$\sup_{|z|>r} \left( 2\frac{(S(t)+b(z))^\top z}{|z|^2} + \frac{\operatorname{tr}(\sigma\sigma^\top(z))}{|z|^2} \right) < 0 \tag{2.18}$$

for some sufficiently large $r \in (0,\infty)$ (compare [44, Example 3.9]). This is obviously covered by the conditions (P), (L3), and (L5). Turning to the entire process $\mathbb{X}$, the conditions (L4) and (L6) or diagonality of $\sigma\sigma^\top$ are only needed in order to decouple the mutual influences of the $x$- and $z$-variables in the generator, as becomes apparent in the second step of the proof of Theorem 2.11.

We observe that the requirements of Theorems 2.11 and 2.13 can be divided into assumptions that are specific to the deterministic model (conditions (L1) and (L2) and the topological assumption on $V$) and assumptions on the external noise (everything else). Therefore, in Example 2.15 we first give some examples of the kind of external noise these Theorems can handle. In the line of Examples 2.16 to 2.18, we discuss applications to the examples that were introduced in Section 1.2 by checking (L1) and (L2) in the respective context.

**Example 2.15.** Concerning the drift $b$ of the external equation, our main motivation and interest lies in a mean reverting mechanism, i.e.

$$b\colon \mathbb{R}^N \to \mathbb{R}^N, \quad z \mapsto -\gamma z + \beta,$$

with $\gamma \in (0,\infty)$ and $\beta \in \mathbb{R}^N$. In this case, the conditions (L3) and (L4) are obviously fulfilled.

Turning to the volatility $\sigma$ of the external equation, the very simple yet interesting case of $\sigma \in \mathbb{R}^{N\times M}$ being constant is obviously covered by (L6). For $M = N = 1$, the

$1 \times 1$-matrix $\sigma\sigma^\top(z) = \sigma^2(z)$ is always diagonal, and (L5) just means that $\sigma$ attains no more than sublinear growth. This includes smooth functions of the type

$$\sigma\colon \mathbb{R} \to (0,\infty), \quad z \mapsto \left(z^2 + \varepsilon\right)^{\alpha/4},$$

with some truncation level $\varepsilon > 0$ and an exponent $\alpha \in (0,2)$. For $\alpha = 1$ this is a smoothly truncated Cox-Ingersoll-Ross type volatility $\sigma(z) \approx \sqrt{|z|}$. For general $M = N$ and a diagonal volatility

$$\sigma\colon \mathbb{R}^N \to \mathbb{R}^{N\times N}, \quad z \mapsto \begin{pmatrix} \sigma_1(z) & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_N(z) \end{pmatrix},$$

the matrix $\sigma\sigma^\top(z)$ is also diagonal, and (L5) is satisfied if and only if

$$\frac{\sigma_i^2(z)}{|z|^2} \xrightarrow{|z|\to\infty} 0 \quad \text{for all } i \in \{1,\dots,N\}.$$

In particular, if each $\sigma_i$ only depends on $z_i$, (L5) is equivalent to

$$\frac{\sigma_i^2(z_i)}{|z_i|^2} \xrightarrow{|z_i|\to\infty} 0 \quad \text{for all } i \in \{1,\dots,N\}.$$

This covers the simple yet interesting special case of $N$ independent sources of noise each acting on one of the $N$ components of the signal $S$. However, note that assumption (ii) from Theorems 2.11 and 2.13 can also be satisfied for $M > N$, since diagonality of $\sigma\sigma^\top(z)$ just means that the rows of $\sigma(z)$ are orthogonal.

**Example 2.16.** In the case of the stochastic Hodgkin-Huxley system (SHH), the state space $V = [0,1]^3$ of the internal variables is compact. Thanks to Theorem 2.13, it is therefore enough to check (L1), while (L2) is irrelevant. This example is handled essentially in the same way as in the proof of [35, Proposition 4].

The function $F$ from (1.5) can be rewritten as

$$F(x,y) = -x\big(0.3 + u(y)\big) + v(y) \quad \text{for all } (x,y) \in \mathbb{R} \times V,$$

with suitable functions $u \in C(V;[0,\infty))$ and $v \in C(V)$ which are bounded due to the compactness of $V$. Let $\varphi \in C^2(\mathbb{R};[1,\infty))$ be any function such that for some $r(\varphi) \in (1,\infty)$ it coincides with $|\cdot|$ outside of $B_{r(\varphi)}(0)$. Let $y \in V$. On the one hand, for all $x \in \mathbb{R} \setminus B_{r(\varphi)}(0)$ we have

$$\begin{aligned} F(x,y) \cdot \varphi'(x) = F(x,y) \cdot \frac{x}{|x|} &= -\,|x|\,\big(0.3 + u(y)\big) + \frac{x}{|x|} v(y) \\ &\le -0.3\,|x| + \|v\|_\infty = -0.3\varphi(x) + \|v\|_\infty\,. \end{aligned}$$

On the other hand, for $x \in B_{r(\varphi)}(0)$ we have

$$F(x,y) \cdot \varphi'(x) \le r(\varphi) \, \|\varphi'\|_\infty \big(0.3 + \|u\|_\infty\big) + \|\varphi'\|_\infty \, \|v\|_\infty =: c \in (0,\infty).$$

Together, for all $x \in \mathbb{R}$ we obtain

$$F(x,y) \cdot \varphi'(x) \le -0.3\varphi(x) + \max\left\{ \|v\|_\infty, c + 0.3 \sup_{\tilde{x} \in B_{r(\varphi)}(0)} |\varphi(\tilde{x})| \right\},$$

and hence (L1) holds.

**Example 2.17.** In the toy example (1.14), the state space $V = [0,\infty)^L$ of the internal variables is closed but not compact. Hence, we are in the setting of Theorem 2.11 and we have to check both (L1) and (L2). The function $F$ as in (1.15) can be treated in the same way as in Example 2.16. Just take any $\varphi \in C^2\big(\mathbb{R}^N; [1,\infty)\big)$ that for some $r(\varphi) \in (1,\infty)$ coincides with $|\cdot|$ outside of $B_{r(\varphi)}(0_N)$ and let $y \in V$. If $x \in B_{r(\varphi)}(0_N)$, we get the bound

$$F(x,y)^\top \nabla\varphi(x) \le \|f\|_\infty \big(r(\varphi) + \|h\|_\infty\big) \|\nabla\varphi\|_\infty + \|j\|_\infty \|\nabla\varphi\|_\infty,$$

and for $x \in \mathbb{R}^N \setminus B_{r(\varphi)}(0_N)$ (which is not in the support of $h$) we have

$$\begin{aligned} F(x,y)^\top \nabla\varphi(x) &= F(x,y)^\top \frac{x}{|x|} = -f(y)\,|x| + j(x,y)^\top \frac{x}{|x|} \\ &\le -|x| + \|j\|_\infty = -\varphi(x) + \|j\|_\infty. \end{aligned}$$

As in Example 2.16, these two properties together imply that (L1) is fulfilled. In order to treat the coefficient functions for the internal variables, let $(x,y) \in \mathbb{R}^N \times V$. For $G$ as in (1.19), we have

$$\begin{aligned} G(x,y)^\top y &= -|y|^2 + g(x)^\top y \le -|y|^2 + \|g\|_\infty |y| \\ &= -|y|^2 + \big(1_{|y| \le 2\|g\|_\infty} + 1_{|y| > 2\|g\|_\infty}\big) \|g\|_\infty |y| \\ &\le -|y|^2 + 2\|g\|_\infty^2 + \frac{1}{2}|y|^2 = -\frac{1}{2}|y|^2 + 2\|g\|_\infty^2. \end{aligned}$$

For $G$ as in (1.20), we calculate analogously

$$G(x,y)^\top y = -\left|y^2\right| + g(x,y)^\top y \le -\frac{1}{2}\left|y^2\right| + 2\|g\|_\infty^2.$$

In other words, (L2) is fulfilled in both situations.

**Example 2.18.** Let us discuss to which extend the rotor models from Example 1.3 can be treated with the technique from this section. Using similar arguments as in Examples 2.16 and 2.17, we see that both the function $F$ from (1.10) for one-sided input and $F$ from (1.12) for two-sided input satisfy (L1), provided that $u_1$, $w_1$ (and $u_3$, $w_3$ for the two-sided case) are bounded. However, having a look at $G$, we see that there is no hope for (L2) to hold, since in both cases $G$ has components that are simply $x$. In [11] and [24], the authors construct Lyapunov functions for similar systems by using the Hamiltonian operator of the system and certain averaging techniques. Their approach is entirely different from Theorems 2.11 and 2.13.

## 2.3 Deterministic control

The basic idea of this chapter is that stability properties of the deterministic dynamical system (DDS) should translate into some kind of stability of the perturbed system (SDS). It turns out that this is indeed the case if our notion of stability for (SDS) is the concept of deterministic control which comes into play in assumption 5 of Theorem 2.3 and which we will now introduce and explain in some more detail.

Assume that

**(C1)** $C^1$**-volatility:** The mapping $\sigma\colon \mathbb{R}^N \to \mathbb{R}^{N\times M}$ is continuously differentiable.

and let

$$\tilde{b}\colon \mathbb{R}^N \to \mathbb{R}^N, \quad z \mapsto b(z) - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{M} \sigma_{i,j}(z) \begin{pmatrix} \partial_{z_i}\sigma_{1,j}(z) \\ \vdots \\ \partial_{z_i}\sigma_{N,j}(z) \end{pmatrix}, \tag{2.19}$$

denote the Stratonovich version of the drift $b$ with respect to $\sigma$, i.e. $\tilde{b}$ is the drift coefficient such that $Z$ satisfies

$$dZ_t = [S(t) + \tilde{b}(Z_t)]dt + \sigma(Z_t) \circ dW_t,$$

where "$\circ\, dW_t$" indicates that we want to interpret the stochastic integral in the sense of Stratonovich instead of Itō. For a short and clear introduction to the Stratonovich integral and its relation to the Itō integral, we refer to [2, Chapter I.9 and Chapter VIII.3]. For the moment, it is enough to know that the drift transformation (2.19) turns an Itō equation into a Stratonovich equation. Hence, if we define $\tilde{B}$ in the same way as $B$ in (1.3), only replacing $b$ with $\tilde{b}$, we obviously get the Stratonovich drift for the entire system (SDS), i.e.

$$d\mathbb{X}_t = \tilde{B}(t, \mathbb{X}_t)dt + \Sigma(\mathbb{X}_t) \circ dW_t.$$

Now fix some finite time horizon $t_0 \in (0,\infty)$ and look at solutions $\Psi^h \in C([0,t_0];\mathtt{E})$ to the corresponding deterministic integral equation

$$d\Psi^h(t) = \tilde{B}\big(t, \Psi^h(t)\big)dt + \Sigma\big(\Psi^h(t)\big)\dot{h}(t)dt \quad \text{for all } t \in [0,t_0], \quad \Psi^h(0) = \mathbb{X}_0,$$

where we replaced the Brownian Motion $W$ with a deterministic absolutely continuous function $h\colon [0,t_0] \to \mathbb{R}^M$ with weak derivative $\dot{h}$. The classical Support Theorem[1] states that in the case $V = \mathbb{R}^L$ (and hence $\mathtt{E} = \mathbb{R}^{N+L+N}$), the support of the probability law

$$\mathcal{L}([0,t_0] \ni t \mapsto \mathbb{X}_t \,|\, \mathbb{P}) \quad \text{on } \mathcal{B}\big(C([0,t_0];\mathtt{E})\big)$$

is given by the closure of the set

$$\left\{ \Psi^h \,\middle|\, h = \int_0^{\cdot} \dot{h}(s)ds \text{ with } \dot{h} \in \mathbb{L}^2\big([0,t_0];\mathbb{R}^M\big) \right\} \subset C([0,t_0];\mathtt{E}),$$

where both the Borel-$\sigma$-field and the closure are meant with respect to the topology of uniform convergence. The result is obtained by approximating $\mathbb{X}$ by $\Psi^h$ with functions $h$ that approximate the path of the driving Brownian Motion $W$ and then using Girsanov's Theorem (see [51]). The proof of Theorem 2.3 that is given in [37] uses a localised version of the Support Theorem which also works for other choices of $V$, provided that assumption 1 of Theorem 2.3 holds (see [37, Theorem 3.1]).

In order to continue the argument from Remark 2.8, we want to find a specific state that the process can attain from any starting point. In the spirit of the Support Theorem, we work with the following notion of attainability.

**Definition 2.19.** A point $(x^*, y^*, z^*) \in \operatorname{int}(\mathtt{E})$ is called *attainable in a sense of deterministic control* (or just *attainable* for short) if for any starting point $(x_0, y_0, z_0) \in \mathtt{E}$ there are $\dot{h} \in \mathbb{L}^2_{\text{loc}}\big([0,\infty);\mathbb{R}^M\big)$ and a solution $\Psi^h \in C([0,\infty);\mathtt{E})$ to

$$\begin{aligned} d\Psi^h(t) &= \tilde{B}\big(t, \Psi^h(t)\big)dt + \Sigma\big(\Psi^h(t)\big)\dot{h}(t)dt \quad \text{for all } t \in [0,\infty), \\ \Psi^h(0) &= (x_0, y_0, z_0) \end{aligned} \tag{2.20}$$

with

$$\Psi^h(t) \xrightarrow{t\to\infty} (x^*, y^*, z^*).$$

We call $h = \int_0^{\cdot} \dot{h}(t)dt$ the *control* and $\Psi^h$ the corresponding *control path.*

**Remark 2.20.** As in Remark 2.8, we want to discuss the relevance of this definition in the context of the proof of Theorem 2.3. Let $(x^*, y^*, z^*) \in \operatorname{int}(\mathtt{E})$ be an attainable point and $\mathcal{U} \subset \mathtt{E}$ any open environment of it. The Support Theorem implies (see [37, Corollary 3.6]) that for any $\mathbb{X}_0 \in \mathtt{E}$ we have

$$P^k_{0,T}(\mathbb{X}_0, \mathcal{U}) = \mathbb{P}(\mathbb{X}^{\mathbf{gr}}_k \in \mathcal{U}) > 0 \quad \text{for some } k = k(\mathbb{X}_0) \in \mathbb{N}. \tag{2.21}$$

At first glance, the dependence of $k$ on the initial value $\mathbb{X}_0$ appears to be problematic, but we can get rid of it by introducing the sampled chain $\mathbb{X}^{\mathbf{sa}}$ that is characterised by

[1] We use the variant that is presented in [51], for the original form see [63].

the transition kernel

$$R := (1-p)\sum_{k=1}^{\infty} p^{k-1} P_{0,kT} = (1-p)\sum_{k=1}^{\infty} p^{k-1} P_{0,T}^{k} \tag{2.22}$$

with some arbitrary $p \in (0,1)$. This kernel corresponds to sampling the grid chain at independent geometrically distributed times, and it follows directly from the definition of Harris recurrence that positive Harris recurrence of $\mathbb{X}^{\mathbf{sa}}$ is sufficient for positive Harris recurrence of $\mathbb{X}^{\mathbf{gr}}$. The property (2.21) implies

$$\mathbb{P}(\mathbb{X}_1^{\mathbf{sa}} \in \mathcal{U}) = R(\mathbb{X}_0, \mathcal{U}) > 0 \quad \text{for every open } \mathcal{U} \ni (x^*, y^*, z^*). \tag{2.23}$$

In Remark 2.36, we will explain how, together with the local weak Hörmander condition, this can be used to prove Theorem 2.3.

**Remark 2.21.** Note that if we write $(u, v, w) = \Psi^h$, the control system (2.20) in differential notation becomes

$$\begin{aligned} \dot{u} &= F(u,v) + \dot{w} = F(u,v) + S + \tilde{b}(w) + \sigma(w)\dot{h}, \\ \dot{v} &= G(u,v), \\ \dot{w} &= S + \tilde{b}(w) + \sigma(w)\dot{h}. \end{aligned} \tag{2.24}$$

We observe that the third equation is still autonomous and that the equations for $u$ and $v$ coincide with (DDS) where the signal is simply replaced by $\dot{w}$. In other words, if $\dot{w}$ is continuous, we can interpret $\dot{w}$ as a new signal, and then $(u,v)$ will equal $(x,y)[x_0, y_0, \dot{w}]$ with the notation we introduced in (1.1).

In order to establish the existence of an attainable point, we will have to impose suitable stability conditions on the system (DDS). Vaguely spoken, the basic idea is to proceed as follows:

1. Identify an equilibrium $(x^*, y^*) \in \mathbb{R}^N \times \mathrm{int}(V)$ of the zero-input system to which its solution $(x,y)[x_0, y_0, S \equiv 0_N]$ is attracted, provided that $(x_0, y_0)$ is close enough to $(x^*, y^*)$ in some sense.
2. Define $\dot{h}$ in such a way that the resulting function $w$ will eventually rest at a fixed value $z^* \in \mathbb{R}^N$ (that may **not** depend on the initial condition), while up until that time, $\dot{w}$ can serve as an input that steers $(u,v)$ into the domain of attraction of $(x^*, y^*)$.

This way we establish $(x^*, y^*, z^*)$ as an attainable point. The choice of $z^* \in \mathbb{R}^N$ turns out to be completely arbitrary for our technique and our setting (see Theorems 2.22 and 2.25 below).

Two different scenarios will be studied. We start with a very simple and restrictive setting in Theorem 2.22, the primary function of which is to illustrate the general idea in a context that allows to minimise the technical aspects. The main result of this section is then presented in Theorem 2.25 which treats a different, more nuanced and localised setting.

As a start, we introduce the following conditions.

**(C2) Zero-input equilibrium:** There is some $(x^*, y^*) \in \mathbb{R}^N \times \operatorname{int}(V)$ such that

$$\begin{pmatrix} F \\ G \end{pmatrix} (x^*, y^*) = 0_{N+L}.$$

**(C3) Global attractivity of $(x^*, y^*)$ under zero-input:** For any $(x_0, y_0) \in \mathbb{R}^N \times V$ there is a unique solution $(x, y)[x_0, y_0, S \equiv 0_N]\colon [0, \infty) \to \mathbb{R}^N \times V$ to (DDS) with zero-input, and it satisfies

$$(x, y)[x_0, y_0, S \equiv 0_N](t) \xrightarrow{t\to\infty} (x^*, y^*).$$

**(C4) Non-explosiveness for moderate input:** For any $(x_0, y_0) \in \mathbb{R}^N \times V$ there is a $\delta_0 = \delta_0(x_0, y_0) > 0$ such that for any $S \in C_b^\infty\big([0, \infty); \mathbb{R}^N\big)$ with $\|S\|_\infty \leq \delta_0$, the system (DDS) has a unique solution $(x, y)[x_0, y_0, S]\colon [0, \infty) \to \mathbb{R}^N \times V$.

Simply put, these conditions guarantee that for smooth and moderate input (DDS) has a unique non-explosive solution, and if the input is in fact constantly zero from some time on, this solution will ultimately converge to the unique equilibrium of the zero-input system, no matter what the starting point was. This is of course a very strong assumption that is rarely satisfied in applications (note Examples 2.23 and 2.30, though), but treating this setting helps to understand our general approach.

Apart from these conditions on the deterministic dynamical system (DDS), we require a non-degeneracy condition for the external noise that is applied to the signal.

**(C5) Non-degeneracy of $Z$:** At each point $z \in \mathbb{R}^N$, the linear mapping that is defined by the $N \times M$-matrix $\sigma(z)$ is surjective and thus $\sigma(z)$ has a right inverse which we simply denote by $\sigma^{-1}(z) \in \mathbb{R}^{M\times N}$.

Note that surjectivity of the linear mapping $\sigma(z)\colon \mathbb{R}^M \to \mathbb{R}^N$ implies $M \geq N$, so in this sense (C5) makes sure that the external equation for $Z$ is not degenerate itself (compare condition (H2) in Section 2.4 and condition (A1) in Section 3.1). Also note that $\sigma^{-1}(z)$ depends continuously on $z \in \mathbb{R}^N$, as $\sigma(\cdot)$ is continuous and so is the function that maps a surjective matrix to its right inverse.

**Theorem 2.22.** *Under the assumptions (C1) - (C5), the point $(x^*, y^*, z^*)$ is attainable in a sense of deterministic control for all $z^* \in \mathbb{R}^N$.*

*Proof.* Let $(x_0, y_0, z_0) \in \mathtt{E}$ and $z^* \in \mathbb{R}^N$. Using linear interpolation and a mollifier, we can easily construct a smooth function $\varrho \in C_b^\infty([0,\infty);\mathbb{R}^N)$ with $\varrho(0) = z_0$ and $\|\dot{\varrho}\|_\infty \leq \delta_0$ such that $\varrho \equiv z^*$ on $[t_1, \infty)$ for some $t_1 \in (0, \infty)$. Setting

$$\dot{h} := \sigma^{-1}(\varrho)\left(\dot{\varrho} - S - \tilde{b}(\varrho)\right) \in C\big([0,\infty);\mathbb{R}^M\big) \subset \mathbb{L}^2_{\mathrm{loc}}\big([0,\infty);\mathbb{R}^M\big), \tag{2.25}$$

the function $w := \varrho$ trivially satisfies the third equation

$$\dot{w} = S + \tilde{b}(w) + \sigma(w)\dot{h}$$

of the respective control system (2.24). The rest of it is then turned into

$$\begin{aligned} \dot{u} &= F(u,v) + \dot{\varrho}, \\ \dot{v} &= G(u,v). \end{aligned}$$

In other words, $(u, v)$ has to be the solution $(x, y)[x_0, y_0, \dot{\varrho}]$ to (DDS) which uniquely exists thanks to (C4). From time $t_1$ on, the pertaining signal $\dot{\varrho}$ rests at zero, so (C3) implies that

$$\lim_{t\to\infty}(u,v)(t) = \lim_{t\to\infty}(u,v)(t_1 + t) = \lim_{t\to\infty}(x,y)[u(t_1), v(t_1), S \equiv 0_N](t) = (x^*, y^*).$$

To sum things up, the differentiable function

$$(u, v, w) = \big((x,y)[x_0, y_0, \dot{\varrho}], \varrho\big) \colon [0,\infty) \to \mathtt{E}$$

is a solution to the control system (2.20) driven by the function $\dot{h}$ defined in (2.25), and it in fact converges to $(x^*, y^*, z^*)$ when time goes to infinity. □

Even if this method is quite simple and its requirements are fairly strong, we can present an example for which we argue heuristically that it allows the use of Theorem 2.22. Another one will be presented later in Example 2.30.

**Example 2.23.** In the rotor model from Example 1.3, we may think of a situation in which the interaction potentials have the rotors "attract" each other (for example $w_1 = w_3 = \sin(\cdot)$), while the pinning potentials pull the rotors to the centre with a force that grows with each rotation (for example $u_1 = u_2 = u_3 = -\varphi(|\cdot|)$ for some non-negative and non-decreasing function $\varphi$ with $\varphi(0) = 0$ and with a sufficiently steep slope). We do not give a rigorous proof, but it is intuitively clear that such a system will satisfy (C2) - (C4) with $(x^*, y^*) = 0_6$. If so, the rotor system with one-sided (or two-sided) input features $0_7$ (or $0_8$) as an attainable point (provided that $\sigma$ fulfills (C1) and (C5)).

In the proof of Theorem 2.22, the control is chosen in a way that forces $w$ to a specific trajectory, and (C3) takes care of $u$ and $v$. However, in other examples of (DDS), the function $(x, y)[x_0, y_0, S \equiv 0_N]$ may not always be attracted to the zero-input equilibrium $(x^*, y^*)$ as in (C3), but only if $(x_0, y_0)$ is already relatively close to it. Thus, apart from driving $w$ to $z^*$, we have to drive $(u, v)$ into a suitable neighbourhood of $(x^*, y^*)$. In general, these may obviously be two competing tasks, since $u$, $v$, and $w$ are interdependent by the dynamics of (2.24). One possible approach to this problem is to separate $x$ and $y$ and make assumptions on the behaviour of $y$ for $x(\cdot) \equiv x^*$. The method of steering the control paths in the right way then becomes more delicate, but our basic idea – exploiting asymptotic stability of the deterministic dynamical system with zero-input – remains unaltered. To put this into precise terms, we introduce the following assumptions which are refinements or variations of the ones used above.

**(C3') Local attractivity of $(x^*, y^*)$ under zero-input:** There is some $\varepsilon^* > 0$ such that for any initial value $(x_0, y_0) \in B_{\varepsilon^*}(x^*, y^*) \subset \mathbb{R}^N \times V$ there is a unique solution $(x, y)[x_0, y_0, S \equiv 0_N]\colon [0, \infty) \to \mathbb{R}^N \times V$ to (DDS) with zero-input, and it satisfies

$$(x, y)[x_0, y_0, S \equiv 0_N](t) \xrightarrow{t \to \infty} (x^*, y^*).$$

**(C4') Signal-dependent stability of $(x^*, y^*)$:** The mapping

$$\begin{aligned} \mathbb{R}^N \times V \times C_b^\infty\big([0, \infty); \mathbb{R}^N\big) &\to C\big([0, \infty); \mathbb{R}^N \times V\big), \\ (x_0, y_0, S) \qquad &\mapsto \quad (x, y)[x_0, y_0, S], \end{aligned}$$

is well-defined around and continuous in the point $(x^*, y^*, S \equiv 0_N)$ in the sense that for all $\varepsilon > 0$ there is some $\delta(\varepsilon) \in (0, \varepsilon]$ such that for all $S \in C_b^\infty\big([0, \infty); \mathbb{R}^N\big)$ with $\|S\|_\infty < \delta(\varepsilon)$ and all $(x_0, y_0) \in B_{\delta(\varepsilon)}(x^*, y^*)$ there is a unique solution $(x, y)[x_0, y_0, S]\colon [0, \infty) \to \mathbb{R}^N \times V$ to (DDS), and it satisfies

$$(x, y)[x_0, y_0, S](t) \in B_\varepsilon(x^*, y^*) \quad \text{for all } t \in [0, \infty).$$

**(C6) Global attractivity of $y^*$ for equilibrious $x$:** For all $y_0 \in V$ the differential equation

$$\dot{y}(t) = G(x^*, y(t)), \quad y(0) = y_0,$$

has a unique solution $y[y_0, x^*]\colon [0, \infty) \to V$, and it satisfies

$$y[y_0, x^*](t) \xrightarrow{t \to \infty} y^*.$$

**(C7) Relocation to the equilibrium in $x$:** For all $(x_0, y_0) \in \mathbb{R}^N \times V$ there are $t_1 \in (0, \infty)$ and a function $\gamma \in C^1\big([0, \infty); \mathbb{R}^N\big)$ with $\gamma(0) = x_0$ and $\gamma \equiv x^*$ on $[t_1, \infty)$ such that the differential equation

$$\dot{y}(t) = G(\gamma(t), y(t)), \quad y(0) = y_0, \tag{2.26}$$

has a unique local solution $\tilde{y}[y_0, \gamma]\colon [0, t_1] \to V$.

Assumption (C4') is a stability condition that takes into account not only the starting point but also the signal: If we want the solution of (DDS) to stay close to $(x^*, y^*)$, we can achieve this by starting not too far away from it, even if we slightly vary the input. If the input is in fact constantly zero, hypotheses (C3') and (C6) signify that the equilibrium is locally attractive for $x$ and $y$, and its attraction is even global for $y$ if $x$ is already in its equilibrium $x^*$. Condition (C7) makes sure that the $x$-variable can always be forced to a trajectory driving it to the equilibrium $x^*$, while the corresponding equation for $y$ still admits a local solution up to the time the $x$-variable hits $x^*$. Due to (C6), this solution can immediately be extended to a global solution. Thus, (C7) can equivalently be replaced by

**(C7') Relocation to the equilibrium in $x$:** For all $(x_0, y_0) \in \mathbb{R}^N \times V$ there are $t_1 \in (0, \infty)$ and a function $S \in C\big([0, \infty); \mathbb{R}^N\big)$ such that there is a unique solution $(x, y)[x_0, y_0, S]\colon [0, \infty) \to \mathbb{R}^N \times V$ to (DDS), and it satisfies $x(\cdot) \equiv x^*$ on $[t_1, \infty)$.

If (C7') is satisfied, we can simply take $\gamma$ as the corresponding trajectory of $x$ and acquire (C7). If on the other hand (C7) holds, we can deduce (C7') by extending $y$ to a global solution with (C6) and then taking $S = \dot{\gamma} - F(\gamma, y)$.

**Remark 2.24.** A relatively simple (but rather strong) sufficient set of conditions for (C7) is the following:

(i) There are constants $C_1, C_2 \in (0, \infty)$ such that

$$|G(x, y)| \leq C_1 + C_2 \, |(x, y)| \quad \text{for all } (x, y) \in \mathbb{R}^N \times V. \tag{2.27}$$

(ii) $V$ is closed and convex, and for any $y \in \partial V$ we have

$$G(x, y)^\top \nu \leq 0 \quad \text{for all } x \in \mathbb{R}^N \text{ and } \nu \in \mathcal{N}(y), \tag{2.28}$$

where $\mathcal{N}(y)$ denotes the set of all outer normals of $V$ in $y$, i.e. all non-zero vectors $\nu \in \mathbb{R}^L$ such that $B_{|\nu|}(y + \nu) \cap V = \emptyset$.

Under these conditions, any $t_1 \in (0, \infty)$ and any smooth function $\gamma\colon [0, \infty) \to \mathbb{R}^N$ with $\gamma(0) = x_0$ and $\gamma \equiv x^*$ on $[t_1, \infty)$ will work with regard to (C7). First, the Picard-Lindelöf Theorem yields unique local existence thanks to the local Lipschitz continuity of the mapping

$$\tilde{G}\colon [0, \infty) \times V \to \mathbb{R}^L, \quad (t, y) \mapsto G(\gamma(t), y).$$

With $G$ being continuous and $V$ being closed, it can be extended to a continuous function on $\mathbb{R}^N \times \mathbb{R}^L$ (by Tietze's Extension Theorem) for which (2.27) remains valid. This leads to a continuous extension of $\tilde{G}$ on $[0, \infty) \times \mathbb{R}^L$ with

$$\left|\tilde{G}(t, y)\right| \leq \left(C_1 + \sqrt{2}C_2 \left|\gamma(t)\right|\right) + \sqrt{2}C_2 \left|y\right| \quad \text{for all } (t, y) \in [0, \infty) \times \mathbb{R}^L.$$

By classical extensibility results (see [56, Korollar 2.5.1]), this guarantees global existence of a solution to (2.26), but so far this solution could take values anywhere. However, the boundary condition (2.28) ensures that for any $y \in \partial V$ we have

$$\tilde{G}(t, y)^\top \nu \leq 0 \quad \text{for all } t \in [0, \infty) \text{ and } \nu \in \mathcal{N}(y),$$

and thus the set $V$ is positively invariant for (2.26), i.e. solutions starting in $V$ will never leave it (see [56, Satz 7.3.4]).

The set of conditions is now complete, and we can state the main result of this section.

**Theorem 2.25.** *Under the assumptions (C1), (C2), (C3'), (C4'), and (C5) - (C7), the point $(x^*, y^*, z^*)$ is attainable in a sense of deterministic control for all $z^* \in \mathbb{R}^N$.*

*Proof.* Let $(x_0, y_0, z_0) \in \mathtt{E}$, $z^* \in \mathbb{R}^N$, and write $\Psi^h = (u, v, w)$ as in (2.24). As indicated above, the construction of a suitable control path is much more delicate than in Theorem 2.22. Therefore, we divide its construction into several steps, each of which corresponds to a distinct phase in approaching the desired limit (see Figure 2.1).

**Phase 1:** At first, we want to choose the control $h$ in such a way that $u$ is forced to the trajectory $\gamma\colon [0, \infty) \to \mathbb{R}^N$ from (C7). Therefore, we take a function $(u_\sharp, v_\sharp, w_\sharp)$ with

$$\begin{aligned} u_\sharp &= \gamma, \\ \dot{v}_\sharp &= G(\gamma, v_\sharp), \\ \dot{w}_\sharp &= \dot{\gamma} - F(\gamma, v_\sharp), \end{aligned}$$

and $v_\sharp(0) = y_0, w_\sharp(0) = z_0$. Here, the first equation indeed just prescribes $u_\sharp$ to coincide with $\gamma$, while (C7) secures that the second equation admits a solution in $[0, t_1]$, and the last equation is then simply solved by integrating, as the right hand side does not contain the variable $w_\sharp$. Rearranging the first equation of the control system (2.24) for $u$, we see that if we set

$$\begin{aligned}\dot h :=& \sigma^{-1}(w_\sharp)\left(\dot w_\sharp - S - \tilde b(w_\sharp)\right)\\ =& \sigma^{-1}(w_\sharp)\left(\dot u_\sharp - F(u_\sharp, v_\sharp) - S - \tilde b(w_\sharp)\right) \in C\big([0,\infty);\mathbb{R}^M\big),\end{aligned} \tag{2.29}$$

the function $(u,v,w) := (u_\sharp, v_\sharp, w_\sharp)$ satisfies (2.24) for all $t \in [0,t_1]$. Thus we have constructed a control path up until the time $t_1$ at which the first variable $u$ has reached the equilibrium $x^*$.

**Phase 2:** With $\gamma$ pinned to $x^*$ after time $t_1$, the stability condition (C6) implies that $v_\sharp$ can be extended to a global solution with

$$v_\sharp(t) \xrightarrow{t\to\infty} y^*.$$

Together with the continuity of $F$ and with (C2), this yields

$$\dot w_\sharp(t) = \dot\gamma(t) - F(\gamma(t), v_\sharp(t)) \xrightarrow{t\to\infty} -F(x^*,y^*) = 0_N \tag{2.30}$$

which allows us to set

$$t_2 := \inf\left\{t > t_1 \,\middle|\, v_\sharp(t) \in B_{\frac{1}{2}\delta(\varepsilon^*)}(y^*) \text{ and } \dot w_\sharp(t) \in B_{\frac{1}{2}\delta(\varepsilon^*)}(0_N)\right\} \in [t_1,\infty)$$

with $\varepsilon^*$ from (C3') and $\delta(\varepsilon^*)$ chosen according to (C4'). Thus, by the time $t_2$ we have driven $(u_\sharp, v_\sharp)$ into $B_{\delta(\varepsilon^*)}(x^*,y^*)$ and have slowed down the movement in $w_\sharp$ to less than $\delta(\varepsilon^*)$.

**Phase 3:** Up to time $t_2$, we have constructed a useful candidate for $h$ and the corresponding solution $(u,v,w) = (u_\sharp, v_\sharp, w_\sharp)$ to (2.24). Our next step will be to change the definition (2.29) of $\dot h$ beyond time $t_2$ and extend the solution $(u,v,w)$ accordingly. Our aim is to drive $w$ to $z^*$ while keeping $(u,v)$ close to $(x^*,y^*)$. Since $\dot w$ can be thought of as an alternate signal that is fed into the equation, assumption (C4') makes sure that we can achieve this, as long as we move $w$ sufficiently slowly. Linear interpolation and classical mollification techniques allow us to take some $\varrho \in C_b^\infty\big([t_2,\infty);\mathbb{R}^N\big)$ with $\varrho(t_2) = w_\sharp(t_2)$, $\dot\varrho(t_2) = \dot w_\sharp(t_2)$ and $\|\dot\varrho\|_\infty < \delta(\varepsilon^*)$ such that for some $t_3 \in (t_2,\infty)$ we have $\varrho \equiv z^*$ on $[t_3,\infty)$. On $[t_2,\infty)$, we change the definition of $\dot h$ to

$$\dot h := \sigma^{-1}(\varrho)\left(\dot\varrho - S - \tilde b(\varrho)\right) \in C\big([t_2,\infty);\mathbb{R}^M\big), \tag{2.31}$$

and now we can argue in the same way as in the proof of Theorem 2.22: We can obviously extend $w$ by taking it equal to $\varrho$ beyond time $t_2$, and thus $u$ and $v$ have to obey

$$\begin{aligned}\dot u &= F(u,v) + \dot\varrho,\\ \dot v &= G(u,v).\end{aligned}$$

Since $\dot{\varrho}$ now plays the role of a signal whose absolute value is bounded by $\delta(\varepsilon^*)$, assumption (C4') yields that we can indeed extend $(u, v)$ correspondingly and $(u, v)(t_3)$ will still be in $B_{\varepsilon^*}(x^*, y^*)$. Furthermore, $w \equiv z^*$ and $\dot{w} \equiv 0_N$ on $[t_3, \infty)$. Note that $\dot{h}$, as defined by (2.29) on $[0, t_2)$ and by (2.31) on $[t_2, \infty)$, is actually continuous.

**Phase 4:** With $w$ fixed at $z^*$, the control system after time $t_3$ is essentially reduced to the original dynamical system with constant input equal to zero, i.e.

$$\begin{aligned}\dot{u} &= F(u, v) + 0_N,\\ \dot{v} &= G(u, v),\\ \dot{w} &= 0_N.\end{aligned}$$

Since $(u, v)(t_3) \in B_{\varepsilon^*}(x^*, y^*)$, assumption (C3') ensures that $(u, v)(t)$ converges to $(x^*, y^*)$ when $t$ tends to infinity. In conclusion, we have obtained a continuous (and thus locally square-integrable) function $\dot{h}\colon [0, \infty) \to \mathbb{R}^M$ and a continuous control path $(u, v, w)\colon [0, \infty) \to \mathtt{E}$ solving the corresponding control system (2.24) with

$$(u, v, w)(t) \xrightarrow{t\to\infty} (x^*, y^*, z^*).$$

In other words, $(x^*, y^*, z^*)$ is attainable in a sense of deterministic control. $\square$

**Remark 2.26.** The route we take in this proof is inspired by the one used in the proof of Proposition 2.5 of [37], where the authors treat the stochastic Hodgkin-Huxley system (SHH). For the special case in which the autonomous equation for $Z$ is of Cox-Ingersoll-Ross type, they can in fact abandon the need for $\sigma$ to be well-defined on the entire space $\mathbb{R}$. This makes the early phases of the control more complicated, as one has to make sure that $w$ does not leave the state space by turning negative. The case that $Z$ is of Ornstein-Uhlenbeck type – of which our model is a generalisation – is treated mostly in the same way as here, albeit only for the dimension $M = N = 1$ and for constant volatility. However, there is one key difference in the third phase of the control (which corresponds to "Part (V)" of the proof in [37]): In the mentioned article, the control is defined such that $u$ is forced to a trajectory that moves it very slightly away from the equilibrium. If this is done slowly enough, $v$ is assumed to stay close to $y^*$, and $w$ approaches $z^*$ similarly as in our third phase. No rigorous proof is given that the deterministic Hodgkin-Huxley system actually shows this kind of behaviour. We think that in general it is a more intuitive approach to force not $u$ but $w$ to a specific trajectory, since we can interpret this as feeding a suitable signal into (DDS) that does not let $(x, y)$ escape from the domain of attraction of the equilibrium. Condition (C4') is the key to making this possible, and while for the Hodgkin-Huxley system we still

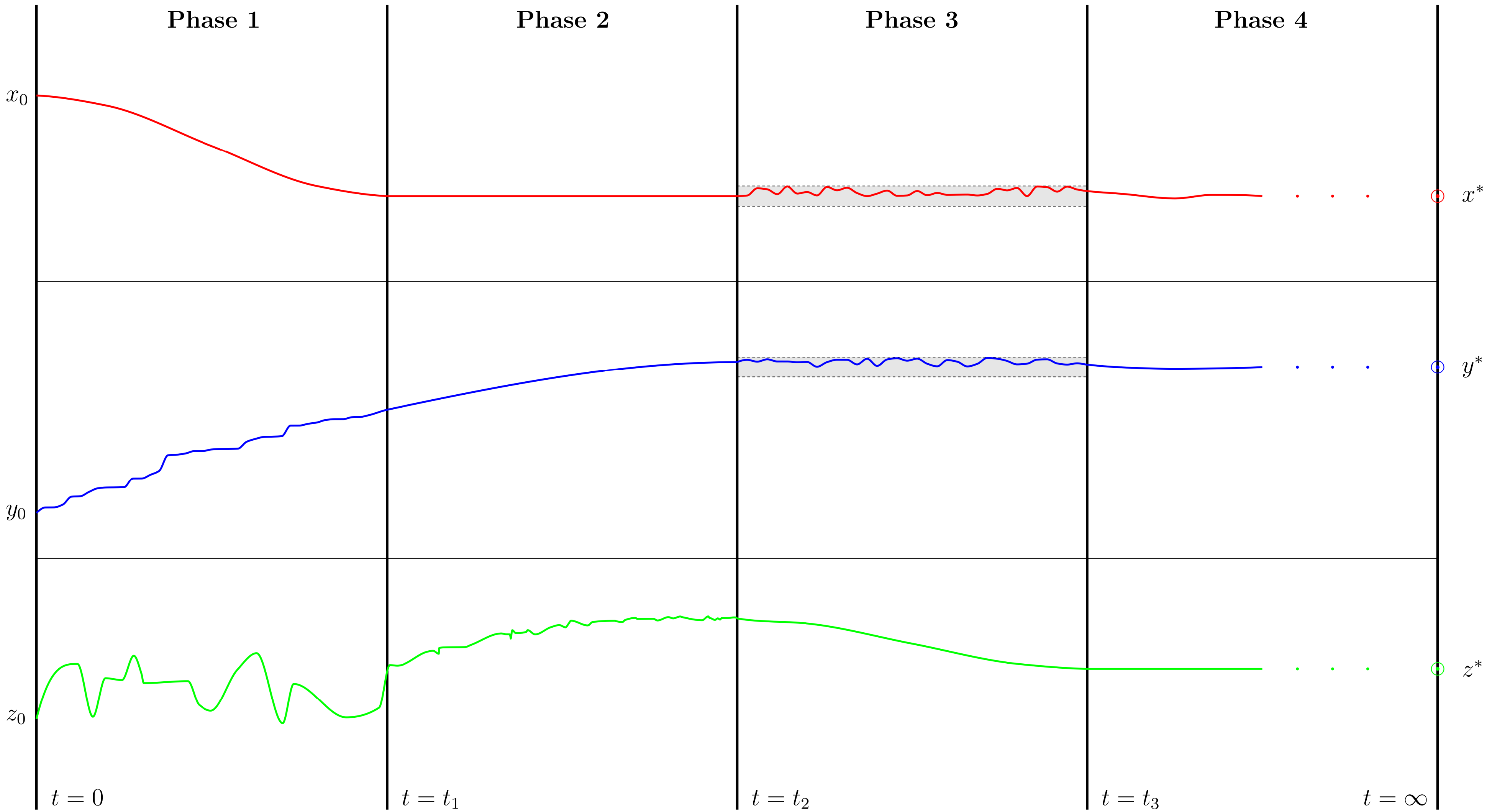


Figure 2.1: Phase 1 guides $u$ (red line) from $x_0$ to $x^*$, while no restrictions are imposed on $v$ (blue) or $w$ (green). In Phase 2, $u$ is kept at $x^*$, which makes $v$ tend to $y^*$, while $w$ loses momentum – but is not steered to a specific point. This happens in Phase 3, slowly enough that $(u, v)$ does not leave the domain of attraction of $(x^*, y^*)$. Once $w$ has reached $z^*$, it sticks there and Phase 4 begins, in which $(u, v)$ finally converges to $(x^*, y^*)$.

merely give a rough intuitive argument relying on numerical simulations (see Example 2.28), this property can be checked neatly for the system we treat in Example 2.29 below.

**Remark 2.27.** We decided to formulate the assumptions (C2) - (C4), (C6), (C3'), and (C4') in terms that are commonly used in stability theory for dynamical systems (see [56] or [65]). One may note that (C2) is mainly needed for the other conditions to make sense. Equation (2.30) is the only instance where we make explicit use of (C2), and all it does is secure that $\dot{h}$ is continuous in $t_2$. Since continuity of $\dot{h}$ is not required in Definition 2.19 anyway, this is of no major importance for the validity of the Theorem.

In the following examples, we want to discuss properties of certain systems like (DDS) that allow for the application of Theorem 2.25. The basic conditions (C1) and (C5) only concern $\sigma$, and we simply assume that they hold.

**Example 2.28.** We return to the stochastic Hodgkin-Huxley system (SHH), the stability of which has mainly been studied numerically. Simulations suggest that there is a number $C \in (0, \infty)$ such that every constant input $S \equiv c \in (-C, C)$ is injectively and continuously mapped to an equilibrium

$$(x_c^*, y_c^*) = \left( x_c^*, \frac{\alpha_1(x_c^*)}{(\alpha_1 + \beta_1)(x_c^*)}, \frac{\alpha_2(x_c^*)}{(\alpha_2 + \beta_2)(x_c^*)}, \frac{\alpha_3(x_c^*)}{(\alpha_3 + \beta_3)(x_c^*)} \right) \in \mathbb{R} \times (0,1)^3$$

which is stable and locally attractive (see [37, pages 533 and 548]), i.e. for all $\varepsilon > 0$ there is a $\delta = \delta(c, \varepsilon) > 0$ such that for all $(x_0, y_0) \in B_\delta(x_c^*, y_c^*)$ we have

$$(x, y)[x_0, y_0, S \equiv c](t) \in B_\varepsilon(x_c^*, y_c^*) \quad \text{for all } t \in [0, \infty), \tag{2.32}$$

and there is some $\varepsilon^* > 0$ such that for all $(x_0, y_0) \in B_{\varepsilon^*}(x_c^*, y_c^*)$ we actually even have

$$(x, y)[x_0, y_0, S \equiv c](t) \xrightarrow{t \to \infty} (x_c^*, y_c^*).$$

Therefore, (C2) and (C3') are taken for granted, and we write $(x^*, y^*) := (x_0^*, y_0^*)$. Simulations show that if started slightly beside the equilibrium, the solution is immediately pulled in its direction, suggesting that one can in fact take $\delta = \varepsilon$ in (2.32). This is also the reason why stability in the sense of (2.32) should remain valid when also slightly and slowly changing the input over time as we do with the function $\varrho$ in the third control phase in the proof of Theorem 2.25: A small change of the input $c$ simply corresponds to a small change of the equilibrium $(x_c^*, y_c^*)$ towards which the trajectory is headed. This is essentially (C4'). Concerning (C6), we first note that for a given trajectory of the membrane potential $x$, variation of constants yields that for every $i \in \{1, 2, 3\}$ and $t \in [0, \infty)$ we can write

$$y_i(t) = y_i(0) e^{-\int_0^t (\alpha_i + \beta_i)(x(s)) ds} + \int_0^t \alpha_i(x(s)) e^{-\int_s^t (\alpha_i + \beta_i)(x(r)) dr} ds. \tag{2.33}$$

For $x(\cdot) \equiv x^*$ this turns into

$$y_i(t) = y_i(0)e^{-(\alpha_i+\beta_i)(x^*)t} + \frac{\alpha_i(x^*)}{(\alpha_i+\beta_i)(x^*)}\left(1 - e^{-(\alpha_i+\beta_i)(x^*)t}\right),$$

which indeed converges to $y_i^*$ for $t \to \infty$, independently of the starting point $y_0 \in V$. Equation (2.33) also takes care of (C7), since these functions remain well-defined when $x$ is forced to a sufficiently nice trajectory $\gamma$.

**Example 2.29.** We revisit the toy example from Example 1.4, where at first we consider the system with $G$ chosen as in (1.19). The point

$$(x^*, y^*) = (0_N, g(0_N)) \in \mathbb{R}^N \times (0, \infty)^L$$

is obviously an equilibrium for the zero-input system in the sense of (C2). For the other assumptions, we will mostly discuss the variables $(x(t))_{t\in[0,\infty)} \subset \mathbb{R}^N$ and $(y(t))_{t\in[0,\infty)} \subset V$ separately, while one of them is considered to be fixed.

If $(x(t))_{t\in[0,\infty)}$ is any fixed continuously differentiable trajectory in $\mathbb{R}^N$, the function $y$ that is defined by

$$y(t) = e^{-t}y_0 + \int_0^t e^{-(t-s)}g(x(s))ds \in (0,\infty) \quad \text{for all } t \in [0,\infty) \tag{2.34}$$

is the solution to the initial value problem

$$\dot{y}(t) = G(x(t), y(t)) = -y(t) + g(x(t)), \quad y(0) = y_0 \in V, \tag{2.35}$$

which is why (C7) is fulfilled.

Next, we want to check the global attractivity property (C6), or actually a slightly stronger variant of it that will be useful to check (C3') below. If we suppose that $(x(t))_{t\in[0,\infty)} \subset \mathbb{R}^N$ is a trajectory that converges to $x^*$, then (2.34) and dominated convergence imply that the corresponding solution $y$ of (2.35) fulfills

$$\begin{aligned} y(t) &= e^{-t}y_0 + \int_0^\infty 1_{[0,t]}(s)e^{-s}g(x(t-s))ds \\ &\xrightarrow{t\to\infty} g(x^*) = g(0_N) = y^*. \end{aligned} \tag{2.36}$$

For the specific choice $x(\cdot) \equiv x^*$, this yields (C6).

In preparation of checking the properties (C3') and (C4'), at first we fix any continuously differentiable trajectory $(y(t))_{t\in[0,\infty)} \subset V$ and concentrate on the initial value problem

$$\begin{aligned} \dot{x}(t) &= F(x(t), y(t)) + S(t) \\ &= -f(y(t))\big(x(t) + h(x(t))\big) + j(x(t), y(t)) + S(t), \\ x(0) &= x_0 \in \mathbb{R}^N. \end{aligned} \tag{2.37}$$

Thanks to the assumptions (1.16), (1.17), and (1.18) on $f$, $h$, and $j$, we see that for all $x \in B_1(0_N)$ we have

$$\begin{aligned} x^\top (F(x, y(t)) + S(t)) &= -f(y(t)) \left(|x|^2 + x^\top h(x)\right) + x^\top S(t) \\ &\leq -(1 - h_0) |x|^2 + \|S\|_\infty |x| . \end{aligned} \tag{2.38}$$

The property (2.38) allows to determine the behaviour of the solution $x(\cdot)$ of (2.37) near the origin $0_N = x^*$, and it is thus the key to both (C3') and (C4'). We will explain this in detail in the next three paragraphs.

In order to check (C3'), we assume that $S \equiv 0_N$. This turns (2.38) into

$$x^\top F(x, y(t)) \leq -(1 - h_0) |x|^2 < 0 \quad \text{for all } x \in B_1(0_N) \setminus \{0_N\}, \tag{2.39}$$

where we used that $h_0 \in [0, 1)$ by (1.17). This means that near the origin the solution $x(\cdot)$ to (2.37) for zero-input is driven back towards $0_N = x^*$ with a force that is independent of $t$. Put in rigorous terms from the theory of dynamical systems, $V(t, x) := \frac{1}{2} |x|^2$ defines a Lyapunov function for (2.37) with $S \equiv 0_N$ in the sense of [56, Definition 8.1.1] and the left hand side of (2.39) is its orbital derivative in the sense of [56, (8.3)]. Consequently, [56, Satz 8.3.3.3] yields that $x^*$ is stable and locally attractive for the system (2.37) with zero-input, i.e. for all $\varepsilon > 0$ there is a $\delta > 0$ such that for all $x_0 \in B_\delta(x^*)$ we have $x(t) \in B_\varepsilon(x^*)$ for all $t \in [0, \infty)$, and if $\delta$ is sufficiently small we also have $x(t) \to x^*$ for $t \to \infty$. We stress the crucial point that the estimate in (2.39) works independently of $(y(t))_{t \in [0,\infty)}$, which is why we can combine this result with the attractivity property (2.36) and conclude that (C3') holds.

In the context of (C4'), we have to consider non-vanishing but very small smooth signals. We note that in this case the upper bound in (2.38) is still strictly negative for $|x| > (1 - h_0)^{-1} \|S\|_\infty$ (recall that $h_0 \in [0, 1)$ by (1.17)). This suggests that, given a neighbourhood of $x^* = 0_N$, we should be able to keep $x(\cdot)$ from escaping from it by choosing $S$ sufficiently small. The formula (2.34) and the continuity of $g$ should then also take care of taming the $y$-variables at the same time. The properties we have just described are exactly what (C4') demands.

We will now turn this intuitive idea into a rigorous proof. Let $\varepsilon > 0$ and, using the continuity of $g$, choose

$$\zeta \in \left(0, \min\left\{1, \varepsilon/(3\sqrt{L})\right\}\right) \tag{2.40}$$

such that

$$g(B_\zeta(0_N)) \subset B_{\varepsilon/(3\sqrt{L})}(g(0_N)). \tag{2.41}$$

If we choose

$$\delta \in (0, (1 - h_0)\zeta), \tag{2.42}$$

and the signal $S \in C_b^\infty\big([0,\infty);\mathbb{R}^N\big)$ satisfies $\|S\|_\infty < \delta$, the property (2.38) yields that in the non-empty half-open ring

$$B_\zeta(0_N) \setminus B_{(1-h_0)^{-1}\|S\|_\infty}(0_N),$$

the solution $x(\cdot)$ of (2.37) is driven back towards the origin – again, independently of $t$ and $(y(t))_{t\in[0,\infty)}$. Using the same arguments as in [56, Lemma 8.3.1 and Satz 8.3.3.1],[2] one can use this to prove that

$$\begin{gathered}\text{for any trajectory } (y(t))_{t\in[0,\infty)} \text{ and any } x_0 \in B_\delta(0_N)\\ \text{the solution } x(\cdot) \text{ of (2.37) will stay in } B_\zeta(0_N) \text{ for all time.}\end{gathered} \tag{2.43}$$

We will now show that for all $(x_0, y_0) \in B_\delta(x^*) \times B_\delta(y^*)$ we have

$$(x,y)[x_0,y_0,S](t) \in B_\varepsilon(x^*) \times B_\varepsilon(y^*) \quad \text{for all } t \in [0,\infty) \tag{2.44}$$

which implies that (C4') holds. Indeed, we have just shown that, no matter what the $y$-variables do, the $x$-variables cannot leave $B_\zeta(x^*)$ which is contained in $B_\varepsilon(x^*)$ by (2.40). In order to prove the converse, let $y_0 \in B_\delta(y^*)$. Then for all $k \in \{1,\dots,L\}$ and $t \in [0,\infty)$ we can use (2.34), (2.43), and (2.41) to obtain that

$$\begin{aligned} y_k(t) &= e^{-t}y_k(0) + \int_0^t e^{-(t-s)}g_k(x(s))ds\\ &\geq e^{-t}y_k(0) + \big(g_k(0_N) - \varepsilon/(3\sqrt{L})\big)(1-e^{-t})\\ &= y_k^* - \varepsilon/(3\sqrt{L}) - e^{-t}\big(y_k^* - y_k(0) - \varepsilon/(3\sqrt{L})\big)\\ &> y_k^* - \varepsilon/(3\sqrt{L}) - \big(\delta + \varepsilon/(3\sqrt{L})\big)\\ &\geq y_k^* - \varepsilon/\sqrt{L}, \end{aligned}$$

where the last step made use of (2.42) and (2.40). In the same way, one can show that

$$y_k(t) < y_k^* + \varepsilon/\sqrt{L}.$$

This means that $y(t) \in B_\varepsilon(y^*)$ which implies (2.44) and hence (C4').

In conclusion, every assumption of Theorem 2.25 is fulfilled, and thus $(x^*, y^*, z^*)$ is attainable in a sense of deterministic control for any choice of $z^* \in \mathbb{R}^N$.

[2] The sole difference in our case is that, if the bound for the signal is fixed, we are merely interested in not leaving a specific ball, while we do not care about arbitrarily small balls centred around the origin. It is also important to note that the function $\psi$ in the proof of [56, Satz 8.3.3.1] can be chosen uniformly for all trajectories of $y$, which is why the same is true for $\delta$. The reason why this works is that (locally around $x^*$) the function $f$ is the only object involved that is influenced by the $y$-variables, and it is bounded from below by (1.16).

Using basically the same arguments, one can show that for the system with $G$ chosen as in (1.20), the point

$$(x^*, y^*) = \big(0, g_1(0), g_2(g_1(0)), \dots, g_L(\dots(g_2(g_1(0))))\big)^\top \in \mathbb{R} \times (0, \infty)^L$$

is attainable.

**Example 2.30.** If we take $j \equiv 0_N$ in Example 2.29, the same reasoning as there can be used to show that $(x^*, y^*)$ is in fact globally attractive in the sense of (C3), and hence the conclusion also follows from the simpler Theorem 2.22.

## 2.4 Hörmander's condition

In this section, we will present conditions on the external equation and on the way $X$ and $Y$ interact through the functions $F$ and $G$ that imply the so-called *local weak Hörmander condition* (see Definition 2.35 below), which is needed in assumption 5 of Theorem 2.3.

**Notation 2.31.** Whenever it is convenient, we will denote elements of $\mathbb{R}^{1+N+L+N}$ by

$$\xi = (\xi_0, \dots, \xi_{N+L+N})^\top := (t, x_1, \dots, x_N, y_1, \dots, y_L, z_1, \dots, z_N)^\top = (t, x, y, z)$$

and use the abbreviation

$$U(x, y) := \begin{pmatrix} F(x, y) \\ G(x, y) \end{pmatrix} \quad \text{for all } (x, y) \in \mathbb{R}^N \times V. \tag{2.45}$$

Furthermore, we set

$$\hat{b}(t, z) := S(t) + \tilde{b}(z) \quad \text{for all } (t, z) \in [0, \infty) \times \mathbb{R}^N,$$

where $\tilde{b}$ is as it was defined in (2.19). We will also use the notation

$$A_{\cdot,k} := \begin{pmatrix} A_{1,k} \\ \vdots \\ A_{n,k} \end{pmatrix} \in \mathbb{R}^n \quad \text{for all } k \in \{1, \dots, m\}$$

for the $k$-th column of a matrix $A \in \mathbb{R}^{n \times m}$ with $m, n \in \mathbb{N}$.

**Definition 2.32.** Let $n \in \mathbb{N}$ and for $i \in \{1, 2\}$ let $V_i$ be a continuously differentiable vector field on some $D \subset \mathbb{R}^n$, i.e.

$$V_i = \big(V_i^{(1)}, \dots, V_i^{(n)}\big)^\top \in C^1(D; \mathbb{R}^n).$$

The *Lie bracket of* $V_1$ *and* $V_2$ is defined as the vector field

$$[V_1, V_2] := J_{V_2}V_1 - J_{V_1}V_2 \in C(D;\mathbb{R}^n),$$

where $J_{V_i}$ denotes the respective Jacobian matrix, i.e.

$$J_{V_i}(x) = (\partial_{x_l} V_i^{(k)}(x))_{k,l\in\{1,\dots,n\}} \quad \text{for all } x \in \mathbb{R}^n.$$

**Remark 2.33.** Abstractly speaking, the Lie bracket is the commutator of two vector fields with respect to the non-commutative operation

$$C^1(D;\mathbb{R}^n) \times C^1(D;\mathbb{R}^n) \ni (V_1, V_2) \mapsto J_{V_2}V_1 \in C(D;\mathbb{R}^n),$$

so in particular the Lie bracket of a vector field with itself is zero. Moreover, the mapping

$$C^1(D;\mathbb{R}^n) \times C^1(D;\mathbb{R}^n) \ni (V_1, V_2) \mapsto [V_1, V_2] \in C(D;\mathbb{R}^n).$$

is bilinear and antisymmetric. If $V_1$ and $V_2$ are vector fields that are differentiable of any order, so is their Lie bracket. Hence, the Lie bracket can be viewed as a binary operation on $C^\infty(D;\mathbb{R}^n)$.

**Definition 2.34.** For any $\mathcal{C} \subset C^\infty(D;\mathbb{R}^n)$ the *Lie algebra generated by* $\mathcal{C}$ is defined as the smallest linear subspace of $C^\infty(D;\mathbb{R}^n)$ that contains $\mathcal{C}$ and that is closed with respect to the binary operation of taking Lie brackets. We denote it by $\mathcal{C}^*$.

The idea behind the Lie bracket in the sense of Definition 2.32 is that by combining motions in the directions $V_1$ and $V_2$ one can effectively approximate motion in the direction $[V_1, V_2]$ and ultimately in any direction in $\{V_1, V_2\}^*$. If $V_1$ and $V_2$ occur in the drift or volatility of a Stratonovich stochastic differential equation, this can be exploited in order to determine along which directions (i.e. along which subspaces of $\mathbb{R}^n$) its solution can evolve locally. Two (not quite identical) detailed heuristic explanations of this idea can be found in [3, pages 73-75] and [23, Section 2].

If we want to apply this reasoning to our time-inhomogeneous setting, we have to consider the homogeneous $(1+N+L+N)$-dimensional time-space process $(t, \mathbb{X}_t)_{t\in[0,\infty)}$. It solves the Stratonovich stochastic differential equation

$$d(t, \mathbb{X}_t) = V_0(t, \mathbb{X}_t)dt + \sum_{k=1}^{M} V_k(t, \mathbb{X}_t) \circ dW_t^{(k)}$$

with the vector fields

$$V_k \colon [0,\infty) \times \mathtt{E} \to \mathbb{R}^{1+N+L+N},$$

defined by

$$V_0(\xi) := \begin{pmatrix} 1 \\ \tilde{B}(\xi) \end{pmatrix} = \begin{pmatrix} 1 \\ F(x,y) + S(t) + \tilde{b}(z) \\ G(x,y) \\ S(t) + \tilde{b}(z) \end{pmatrix} = \begin{pmatrix} 0 \\ F(x,y) \\ G(x,y) \\ 0_N \end{pmatrix} + \begin{pmatrix} 1 \\ \hat{b}(t,z) \\ 0_L \\ \hat{b}(t,z) \end{pmatrix} \tag{2.46}$$

and

$$V_k(\xi) := \begin{pmatrix} 0 \\ \Sigma_{1,k}(x,y,z) \\ \vdots \\ \Sigma_{N+L+N,k}(x,y,z) \end{pmatrix} = \begin{pmatrix} 0 \\ \sigma_{\cdot,k}(z) \\ 0_L \\ \sigma_{\cdot,k}(z) \end{pmatrix} \quad \text{for all } k \in \{1,\dots,M\}. \tag{2.47}$$

The idea outlined above suggests that we should study the collections of vector fields that are recursively constructed by setting

$$\begin{aligned} \mathbb{L}_0 &:= \{V_0,\dots,V_M\}, \\ \mathbb{L}_n &:= \mathbb{L}_{n-1} \cup \{\pm[V_k, V] \,|\, V \in \mathbb{L}_{n-1}, k \in \{0,\dots,M\}\}, \quad n \in \mathbb{N}, \\ \mathbb{L}(\xi) &:= \operatorname{span} \bigcup_{n\in\mathbb{N}} \{V(\xi) \,|\, V \in \mathbb{L}_n^*\}, \quad \xi \in [0,\infty) \times \mathtt{E}. \end{aligned} \tag{2.48}$$

In order for the definition in (2.48) to make sense, we will of course have to require that all of these vector fields are smooth. More precisely, we suppose:

**(H1) Smooth coefficients:** The coefficient functions $F$, $G$, $\sigma$, $S$, and $b$ are differentiable of any order.

This will be a standing assumption for this entire section and it will not be mentioned explicitly again.

According to the idea explained above, $\mathbb{L}(\xi)$ should describe where $(t, \mathbb{X}_t)$ can move from $\xi$ within an infinitesimal time step. The following Definition 2.35 formalises the idea that it can move in any direction. Its roots go back to [38] and it has since become classical to work with context-specific variants of it in order to prove the existence of transition densities, usually incorporating tools from the Malliavin calculus (see for example [23] and the references therein). The variant presented below is customised for our time-inhomogeneous setting, in which we think this formulation is the most intuitive. It is essentially the same as [36, (LWH)] and its connection to the common terminology in other literature is discussed in Remark 2.38 below.

**Definition 2.35.** Let $(x,y,z) \in \operatorname{int}(\mathtt{E})$. If

$$\mathbb{L}\big((t,x,y,z)\big) = \mathbb{R}^{1+N+L+N} \quad \text{for all } t \in [0,\infty),$$

we say that the *local weak Hörmander condition* holds at $(x,y,z)$.

**Remark 2.36.** We would like to continue the thread of Remarks 2.8 and 2.20 and thus complete our sketch of the proof of Theorem 2.3. Suppose that $(x^*, y^*, z^*)$ is an attainable point in the interior of $\mathtt{E}$ at which the local weak Hörmander condition holds. The continuity of all coefficient functions yields that it automatically holds in an open environment of $(x^*, y^*, z^*)$. The crucial consequence of this is that one can find an open neighbourhood $\mathcal{U} \subset \mathtt{E}$ of $(x^*, y^*, z^*)$ such that for all $t \in (0, \infty)$ and all $\mathbb{X}_0 \in \mathtt{E}$ the measure $P_{0,t}(\mathbb{X}_0, \cdot)$ locally admits a Lebesgue density $p_{0,t}(\mathbb{X}_0, \cdot) \in C^\infty(\mathcal{U})$ which, for fixed argument, is lower semi-continuous with respect to $\mathbb{X}_0$ (see [37, Lemma 4.1]). Lower semi-continuity extends to the mapping $P_{0,t}(\cdot, A)\colon \mathtt{E} \to [0, \infty)$ for any measurable subset $A \subset \mathcal{U}$. Mutatis mutandis, these properties are passed on to the transition measure $R(\mathbb{X}_0, \cdot)$ of the sampled chain $\mathbb{X}^{\mathbf{sa}}$ that we introduced in (2.22).

All of the essential pieces of the puzzle that is Theorem 2.3 are now at hand, and it remains to put them together. The main point is that, in combination with (2.23), the transition densities and their properties enable us to find an open ball $C \subset \mathcal{U}$ around $x^*$ and another open ball $D \subset \mathcal{U}$ such that the minorisation condition

$$R(\mathbb{X}_0, A) \geq \alpha 1_C(\mathbb{X}_0)\nu(A) \quad \text{for all } \mathbb{X}_0 \in \mathtt{E} \text{ and } A \in \mathcal{B}(\mathtt{E}), \tag{2.49}$$

holds, where $\alpha$ is some positive real constant and $\nu$ is the uniform law on $D$. This is the content of Lemma 3.7 of [37]. The property (2.49) is known as Nummelin's splitting condition (compare [53]) and it entails that $C$ is a small set in the sense of [54, Definition 2.3]. If there is at least one starting point from which $\mathbb{P}$-almost surely $C$ is visited infinitely often by $\mathbb{X}^{\mathbf{sa}}$, part (v) of Theorem 3.7 in [54] implies that $\mathbb{X}^{\mathbf{sa}}$ is Harris recurrent. In order to check this condition, we first note that our Lyapunov function for $\mathbb{X}^{\mathbf{gr}}$ is also a Lyapunov function for $\mathbb{X}^{\mathbf{sa}}$. This provides a compact set $K \subset \mathtt{E}$ to which $\mathbb{X}^{\mathbf{sa}}$ returns infinitely often from any state. The property (2.23) and lower semi-continuity imply

$$\inf_{\mathbb{X}_0 \in K} R(\mathbb{X}_0, C) > 0, \tag{2.50}$$

and the Borel-Cantelli Lemma allows us to conclude that $\mathbb{X}^{\mathbf{sa}}$ indeed visits $C$ infinitely often, no matter where it is started. Hence, $\mathbb{X}^{\mathbf{sa}}$ is Harris recurrent, and it remains to show that recurrence is even positive.[3]

By combining (2.49) and (2.50), it is easy to prove the minorisation condition

$$R^2(\mathbb{X}_0, A) \geq \tilde{\alpha} 1_K(\mathbb{X}_0)\nu(A) \quad \text{for all } \mathbb{X}_0 \in \mathtt{E} \text{ and } A \in \mathcal{B}(\mathtt{E}),$$

where

$$\tilde{\alpha} := \alpha \inf_{\mathbb{X}_0 \in K} R(\mathbb{X}_0, C) > 0.$$

[3]Note that this final part differs slightly from the argument given in [37], where the authors conclude the proof with similar tools from [50].

Consequently, $K$ is a small set as well. Thanks to our Lyapunov function, Proposition 5.10 of [54] then yields that Harris recurrence of $\mathbb{X}^{\mathbf{sa}}$ is indeed positive.[4] As indicated in Remark 2.20, this implies positive Harris recurrence of the grid chain $\mathbb{X}^{\mathbf{gr}}$ and thus also of $\mathbb{X}^{\mathbf{ps}}$ and $\mathbb{X}^{\mathbf{ts}}$ (see [37, page 531, lines 19-21]).

We will now collect some basic observations about the objects and properties we have defined so far in this section. We note that for any two continuously differentiable vector fields $V$ and $W$ on $[0,\infty) \times \mathtt{E}$ we can write their Lie bracket as

$$[V,W] = \sum_{i=0}^{N+L+N} \big(V^{(i)}\partial_{\xi_i} W - W^{(i)}\partial_{\xi_i} V\big) \tag{2.51}$$

which will be convenient for the following remarks.

**Remark 2.37.** Let us make some comments on the role of time, i.e. the (0)-components of the occurring vector fields on the one hand and the derivatives with respect to $t = \xi_0$ on the other hand.

1.) First, note that for all $k \in \{0,\dots,M\}$ the (0)-component of $V_k$ is constant. Hence, (2.51) yields that for any vector field $W \in C^1\big([0,\infty)\times \mathtt{E}; \mathbb{R}^{1+N+L+N}\big)$ we have

$$[V_k, W]^{(0)} = \sum_{i=0}^{N+L+N} V_k^{(i)} \partial_{\xi_i} W^{(0)}$$

which vanishes everywhere whenever $W^{(0)}$ is constant as well. In particular, this is the case for $W \in \{V_0,\dots,V_M\}$. Consequently, any vector field we could possibly create by consecutively taking Lie brackets of $V_0,\dots,V_M$ will have a vanishing (0)-component.

2.) Let $V, W \in C^1\big([0,\infty)\times \mathtt{E}; \mathbb{R}^{1+N+L+N}\big)$. Separating the sum in (2.51) yields

$$[V,W] = V^{(0)}\partial_t W - W^{(0)}\partial_t V + \sum_{i=1}^{N+L+N} \big(V^{(i)}\partial_{\xi_i} W - W^{(i)}\partial_{\xi_i} V\big),$$

so any possible influence of time derivatives of $W$ is killed by a vanishing $V^{(0)}$ and vice versa. The first part of this remark yields that $V_0$ is the only vector field in $\mathbb{L}_n$, $n \in \mathbb{N}$, whose (0)-component $V_0^{(0)} \equiv 1$ does not vanish. Thus, no time derivatives occur in

$$[V,W] = \sum_{i=1}^{N+L+N} \big(V^{(i)}\partial_{\xi_i} W - W^{(i)}\partial_{\xi_i} V\big) \quad \text{for all } V, W \in \mathbb{L}_n \setminus \{V_0\},\ n \in \mathbb{N}, \tag{2.52}$$

while

$$[V_0,W] = \partial_t W + \sum_{i=1}^{N+L+N} \big(V_0^{(i)}\partial_{\xi_i} W - W^{(i)}\partial_{\xi_i} V_0\big) \quad \text{for all } W \in \mathbb{L}_n \setminus \{V_0\},\ n \in \mathbb{N}. \tag{2.53}$$

[4] In fact, Theorem 2.1 of [18] shows that if $R$ is aperiodic (see [54, page 21]), the returning time of $K$ (and any other compact set containing $K$) not only has finite expectation but even admits exponential moments.

As $[V_0, V_0]$ vanishes and so does $\partial_t V_k$ for all $k \in \{1, \dots, M\}$, time derivatives can occur in $\mathbb{L}_2$ at the earliest, but never in $\mathbb{L}_1$.

**Remark 2.38.** The first part of Remark 2.37 allows us to shed some more light on the special role of $V_0$. In close analogy to (2.48), we define

$$\begin{aligned} \mathcal{L}_0 :=& \{V_1, \dots, V_M\}, \\ \mathcal{L}_n :=& \mathcal{L}_{n-1} \cup \{\pm[V_k, V] \,|\, V \in \mathcal{L}_{n-1}, k \in \{0, \dots, M\}\}, \quad n \in \mathbb{N}, \\ \mathcal{L}(\xi) :=& \operatorname{span} \bigcup_{n \in \mathbb{N}} \{V(\xi) \,|\, V \in \mathcal{L}_n^*\}, \quad \xi \in [0, \infty) \times \mathtt{E}. \end{aligned} \tag{2.54}$$

The only difference to (2.48) is in the "initialisation" $\mathcal{L}_0$ or $\mathbb{L}_0$, where the latter contains $V_0$ while the former does not. After that, the construction steps for $n \geq 1$ follow the exact same rules. The first part of Remark 2.37 implies that for every $n \in \mathbb{N}$ the (0)-component of any vector field in $\mathcal{L}_n$ is zero, and in turn the dimension of $\mathcal{L}(\cdot)$ can never exceed $N + L + N$. The set $\mathbb{L}_n$ on the other hand does contain $V_0$ which has the non-vanishing (0)-component $V_0^{(0)} \equiv 1$. In fact, thanks to the antisymmetry of the Lie-bracket,

$$\begin{aligned} \mathbb{L}_1 &= \{V_0, \dots, V_M\} \cup \{\pm[V_k, V_l] \,|\, k, l \in \{0, \dots, M\}\} \\ &= \{V_0\} \cup \mathcal{L}_0 \cup \{\pm[V_k, V_l] \,|\, k \in \{0, \dots, M\}, l \in \{1, \dots, M\}\} \\ &= \{V_0\} \cup \mathcal{L}_1, \end{aligned}$$

and by induction it follows from (2.54) that

$$\mathbb{L}_n = \{V_0\} \cup \mathcal{L}_n \quad \text{for all } n \in \mathbb{N}.$$

In particular, this means that

$$\dim \mathbb{L}(\xi) = \dim \mathcal{L}(\xi) + 1 \tag{2.55}$$

at any point $\xi \in [0, \infty) \times \mathtt{E}$ (cf. [36, Proposition 1]), and hence the local weak Hörmander condition holds at $\xi$ if and only if

$$\dim \mathcal{L}(\xi) = N + L + N. \tag{2.56}$$

This variant of the local weak Hörmander condition is the time-inhomogeneous analogue to the common version from [23, Definition 1.2]. In the sequel, we will check the local weak Hörmander condition by checking (2.56).

For general diffusions, there is very little one can say about sufficient conditions for the local weak Hörmander condition. One usually has to use ad hoc arguments that depend on the particular shape of the drift and the volatility matrix. For systems of the special type (SDS) however, there are two fairly general scenarios which we can treat without confining us to entirely specific examples. Both are based on the idea that external noise (in the form of the $Z$-variable) is explicitly imported only into the $X$-variable, while the coefficient functions $F$ and $G$ have to transport and distribute its influence suitably among all of the $X$- and $Y$-variables.

**(I) Star shape:** Every component depends on the $X$-variables in such a way that sufficient amounts of noise are able to spread from $Z$ via $X$ to the rest of the system.

**(II) Cascade structure:** There is a chain of components such that with each step exactly one more of them is directly influenced by the previous one, and this chain ultimately runs through the entire system.

Of course, in both of these scenarios we also have to make sure that $X$ and $Z$ are not coupled in a degenerate way. However, since the random terms in their respective equations coincide, their interaction is basically coded in the difference of their drift coefficients which is again simply given by the function $F$. It therefore seems natural that, as long as $b$, $\sigma$, and $S$ are nice enough for $Z$ not to be degenerate itself, it should be possible for us to find conditions solely on $F$ and $G$ that are sufficient for the local weak Hörmander condition in the sense of Definition 2.35.

We will treat each of these situations separately in the Subsections 2.4.1 and 2.4.2 below (the main results being Corollaries 2.48, 2.49, and 2.50 for the star shape and Theorem 2.56 for the cascade structure), but first we will collect some preparatory observations and calculations.

Given the particular shape of $V_0, \ldots, V_M$, the following basic Lemma will be helpful in the sequel. It should be stressed once more at this occasion that in accordance with Definition 2.32 we use the same Lie bracket notation – consistently – not only for vector fields on $[0, \infty) \times \mathtt{E}$ but also for vector fields of lower dimensions.

**Lemma 2.39.** *For $i \in \{1, 2\}$ consider continuously differentiable functions*

$$
\begin{aligned}
[0, \infty) \times \mathbb{R}^N \ni (t, z) &\mapsto A_i(t, z) \in \mathbb{R}^N, \\
\mathbb{R}^N \times V \ni (x, y) &\mapsto W_i(x, y) \in \mathbb{R}^{N+L}, \\
[0, \infty) \times \mathbb{R}^N \ni (t, z) &\mapsto \varphi_i(t, z) \in \mathbb{R}.
\end{aligned}
$$

*Then for all* $(t, x, y, z) \in [0, \infty) \times \mathtt{E}$ *we have*

$$\text{1. } \left[\begin{pmatrix} A_1(t,\cdot) \\ 0_L \\ A_1(t,\cdot) \end{pmatrix}, \begin{pmatrix} A_2(t,\cdot) \\ 0_L \\ A_2(t,\cdot) \end{pmatrix}\right](x,y,z) = \begin{pmatrix} [A_1(t,\cdot), A_2(t,\cdot)](z) \\ 0_L \\ [A_1(t,\cdot), A_2(t,\cdot)](z) \end{pmatrix},$$

$$\text{2. } \left[\varphi_1(t,\cdot)\begin{pmatrix} W_1 \\ 0_N \end{pmatrix}, \varphi_2(t,\cdot)\begin{pmatrix} W_2 \\ 0_N \end{pmatrix}\right](x,y,z) = \varphi_1(t,z)\varphi_2(t,z)\begin{pmatrix} [W_1, W_2](x,y) \\ 0_N \end{pmatrix},$$

$$\text{3. } \left[\varphi_1(t,\cdot)\begin{pmatrix} W_1 \\ 0_N \end{pmatrix}, \begin{pmatrix} A_1(t,\cdot) \\ 0_L \\ A_1(t,\cdot) \end{pmatrix}\right](x,y,z) = -\sum_{j=1}^{N} A_1^{(j)}(t,z)\Bigg(\varphi_1(t,z)\begin{pmatrix} \partial_{x_j} W_1(x,y) \\ 0_N \end{pmatrix}$$
$$+ \partial_{z_j}\varphi_1(t,z)\begin{pmatrix} W_1(x,y) \\ 0_N \end{pmatrix}\Bigg).$$

*Proof.* These formulas follow immediately from the definition of the Lie bracket by straight forward calculations. □

The following notational convention is of utmost importance for understanding the rest of this section.

**Notation 2.40.** 1.) As noted under 1.) in Remark 2.37, the only relevant vector field to feature a non-vanishing (0)-component is $V_0$ – which is not contained in $\mathcal{L}_n$ for any $n \in \mathbb{N}$. In order to simplify our notation, in the sequel **we will therefore systematically omit this component**. More precisely, we identify every vector field

$$W\colon [0,\infty) \times \mathtt{E} \to \mathbb{R}^{1+N+L+N}, \quad (t,x,y,z) \mapsto \begin{pmatrix} 0 \\ W^{(1)}(t,x,y,z) \\ \vdots \\ W^{(N+L+N)}(t,x,y,z) \end{pmatrix}, \tag{2.57}$$

with the collection of vector fields on $\mathtt{E}$ given by

$$W(t,\cdot)\colon \mathtt{E} \to \mathbb{R}^{N+L+N}, \quad (x,y,z) \mapsto \begin{pmatrix} W^{(1)}(t,x,y,z) \\ \vdots \\ W^{(N+L+N)}(t,x,y,z) \end{pmatrix}, \quad \text{for all } t \in [0,\infty). \tag{2.58}$$

Either of these objects will simply be denoted by $W$. Let us stress that both directions of this identification are actively used: on the one hand, we tacitly omit vanishing zero-components when a vector field is given as in (2.57), but on the other hand, we also think of each vector field that is given in terms of (2.58) as one on $[0, \infty) \times \mathtt{E}$ with a vanishing zero-component.

2.) If $V$ and $W$ are of this type, so is their Lie bracket $[V, W]$ (confer part 1.) of Remark 2.37). Thus, it is natural (and consistent) to identify

$$[V, W](t, x, y, z) = [V(t, \cdot), W(t, \cdot)](x, y, z) \in \mathbb{R}^{N+L+N} \tag{2.59}$$

for all $(t, x, y, z) \in [0, \infty) \times \mathtt{E}$.

3.) As indicated above, $V_0$ is the only relevant vector field not of the type in (2.57) and (2.58), but we only need it when taking Lie brackets as in (2.53). Using the notation we just introduced (and the abbreviation (2.45)), (2.53) can be rewritten as

$$[V_0, W] = \partial_t W + \left[ \begin{pmatrix} \hat{b} \\ 0_L \\ \hat{b} \end{pmatrix} + \begin{pmatrix} U \\ 0_N \end{pmatrix}, W \right]. \tag{2.60}$$

Note that this formula is completely consistent: $[V_0, W]^{(0)}$ and $\partial_t W^{(0)}$ vanish again and are therefore omitted, while the Lie bracket on the right hand side is to be understood as the Lie bracket of vector fields on $\mathtt{E}$ (which carry an additional dependence on time) and thus has only $N + L + N$ components to begin with.

4.) Note also that with this notational convention there is no difference between $V_k$ and $\Sigma_{\cdot,k}$ for $k \in \{1, \ldots, M\}$.

5.) Let us further illustrate this notation by rewriting the formulas from Lemma 2.39 accordingly. With the identification of (2.57) and (2.58) and with the convention (2.59), they read

1. $$\left[ \begin{pmatrix} A_1 \\ 0_L \\ A_1 \end{pmatrix}, \begin{pmatrix} A_2 \\ 0_L \\ A_2 \end{pmatrix} \right](\xi) = \begin{pmatrix} [A_1(t, \cdot), A_2(t, \cdot)](z) \\ 0_L \\ [A_1(t, \cdot), A_2(t, \cdot)](z) \end{pmatrix},$$

2. $$\left[ \varphi_1 \begin{pmatrix} W_1 \\ 0_N \end{pmatrix}, \varphi_2 \begin{pmatrix} W_2 \\ 0_N \end{pmatrix} \right](\xi) = \varphi_1(t, z)\varphi_2(t, z) \begin{pmatrix} [W_1, W_2](x, y) \\ 0_N \end{pmatrix},$$

3. $$\left[ \varphi_1 \begin{pmatrix} W_1 \\ 0_N \end{pmatrix}, \begin{pmatrix} A_1 \\ 0_L \\ A_1 \end{pmatrix} \right](\xi) = -\sum_{j=1}^{N} A_1^{(j)}(t, z) \Bigg( \varphi_1(t, z) \begin{pmatrix} \partial_{x_j} W_1(x, y) \\ 0_N \end{pmatrix} + \partial_{z_j} \varphi_1(t, z) \begin{pmatrix} W_1(x, y) \\ 0_N \end{pmatrix} \Bigg)$$

for all $\xi = (t, x, y, z) \in [0, \infty) \times \mathtt{E}$.

Using Notation 2.40 and Lemma 2.39, we will now see which vector fields we can construct by taking Lie brackets with $V_0, \ldots, V_M$. With the first formula in Lemma

2.39, it is easy to see that

$$\{V_1, \ldots, V_M\}^* = \left\{ \begin{pmatrix} w \\ 0_L \\ w \end{pmatrix} \,\middle|\, w \in \{\sigma_{\cdot,1}, \ldots, \sigma_{\cdot,M}\}^* \right\}. \tag{2.61}$$

This means that by taking Lie brackets using only the vector fields corresponding to the volatility, we can never construct more than $N$ linearly independent vectors. However, for some fixed $z \in \mathbb{R}^N$ the assumption

**(H2) Non-degeneracy of $Z$:** $\{w(z) \,|\, w \in \{\sigma_{\cdot,1}, \ldots, \sigma_{\cdot,M}\}^*\} = \mathbb{R}^N$

is a reasonable starting point.[5] If, for example, $\sigma(z)$ is surjective (as is needed by condition (C5) in Section 2.3 anyway), it has to have $N$ linearly independent columns, and hence we see without even taking Lie brackets that (H2) holds in $z$.

In order to construct suitable sequences of vector fields that can span the remaining $N + L$ dimensions, we will always start by taking the Lie bracket of the drift $V_0$ with some column $V_k$ of the volatility matrix, $k \in \{1, \ldots, M\}$. Using (2.60) and Lemma 2.39 and exploiting bilinearity as well as antisymmetry of the Lie bracket, we see that

$$\begin{aligned}
[V_0, V_k](\xi) &= \partial_t \begin{pmatrix} \sigma_{\cdot,k}(z) \\ 0_L \\ \sigma_{\cdot,k}(z) \end{pmatrix} + \left[ \begin{pmatrix} \hat{b} \\ 0_L \\ \hat{b} \end{pmatrix} + \begin{pmatrix} U \\ 0_N \end{pmatrix}, \begin{pmatrix} \sigma_{\cdot,k} \\ 0_L \\ \sigma_{\cdot,k} \end{pmatrix} \right](\xi) \\
&= 0_{N+L+N} + \left[ \begin{pmatrix} \hat{b} \\ 0_L \\ \hat{b} \end{pmatrix}, \begin{pmatrix} \sigma_{\cdot,k} \\ 0_L \\ \sigma_{\cdot,k} \end{pmatrix} \right](\xi) + \left[ \begin{pmatrix} U \\ 0_N \end{pmatrix}, \begin{pmatrix} \sigma_{\cdot,k} \\ 0_L \\ \sigma_{\cdot,k} \end{pmatrix} \right](\xi) \\
&= \begin{pmatrix} [\hat{b}(t,\cdot), \sigma_{\cdot,k}](z) \\ 0_L \\ [\hat{b}(t,\cdot), \sigma_{\cdot,k}](z) \end{pmatrix} - \sum_{i=1}^{N} \sigma_{i,k}(z) \begin{pmatrix} \partial_{x_i} U(x,y) \\ 0_N \end{pmatrix}
\end{aligned} \tag{2.62}$$

and hence

$$[V_k, V_0](\xi) = -[V_0, V_k](\xi) = \begin{pmatrix} [\sigma_{\cdot,k}, \hat{b}(t,\cdot)](z) \\ 0_L \\ [\sigma_{\cdot,k}, \hat{b}(t,\cdot)](z) \end{pmatrix} + \sum_{i=1}^{N} \sigma_{i,k}(z) \begin{pmatrix} \partial_{x_i} U(x,y) \\ 0_N \end{pmatrix}. \tag{2.63}$$

In order to get a better idea of what we are dealing with here, let us see what happens when we take an extra Lie bracket with $V_0$ or some column $V_l$ of the volatility matrix,

[5] This is basically the local strong Hörmander condition for the external equation, where "strong" refers to the fact that we do not include its drift at all.

$l \in \{1, \dots, M\}$. Using (2.60), Lemma 2.39, and basic properties of the Lie bracket again, we see that $[V_0, [V_0, V_k]](\xi)$ is equal to

$$
\begin{aligned}
&\partial_t [V_0, V_k](\xi) + \left[ \begin{pmatrix} \hat{b} \\ 0_L \\ \hat{b} \end{pmatrix} + \begin{pmatrix} U \\ 0_N \end{pmatrix}, [V_0, V_k] \right](\xi) \\
&= \begin{pmatrix} \partial_t [\hat{b}(t,\cdot), \sigma_{\cdot,k}](z) \\ 0_L \\ \partial_t [\hat{b}(t,\cdot), \sigma_{\cdot,k}](z) \end{pmatrix} + \left[ \begin{pmatrix} \hat{b} \\ 0_L \\ \hat{b} \end{pmatrix} + \begin{pmatrix} U \\ 0_N \end{pmatrix}, \begin{pmatrix} [\hat{b}(t,\cdot), \sigma_{\cdot,k}] \\ 0_L \\ [\hat{b}(t,\cdot), \sigma_{\cdot,k}] \end{pmatrix} \right](\xi) \\
&\quad - \sum_{i=1}^{N} \left[ \begin{pmatrix} \hat{b} \\ 0_L \\ \hat{b} \end{pmatrix} + \begin{pmatrix} U \\ 0_N \end{pmatrix}, \sigma_{i,k} \begin{pmatrix} \partial_{x_i} U \\ 0_N \end{pmatrix} \right](\xi) \\
&= \begin{pmatrix} \partial_t [\hat{b}(t,\cdot), \sigma_{\cdot,k}](z) \\ 0_L \\ \partial_t [\hat{b}(t,\cdot), \sigma_{\cdot,k}](z) \end{pmatrix} + \left[ \begin{pmatrix} \hat{b} \\ 0_L \\ \hat{b} \end{pmatrix}, \begin{pmatrix} [\hat{b}(t,\cdot), \sigma_{\cdot,k}] \\ 0_L \\ [\hat{b}(t,\cdot), \sigma_{\cdot,k}] \end{pmatrix} \right](\xi) + \left[ \begin{pmatrix} U \\ 0_N \end{pmatrix}, \begin{pmatrix} [\hat{b}(t,\cdot), \sigma_{\cdot,k}] \\ 0_L \\ [\hat{b}(t,\cdot), \sigma_{\cdot,k}] \end{pmatrix} \right](\xi) \\
&\quad - \sum_{i=1}^{N} \left( \left[ \begin{pmatrix} \hat{b} \\ 0_L \\ \hat{b} \end{pmatrix}, \sigma_{i,k} \begin{pmatrix} \partial_{x_i} U \\ 0_N \end{pmatrix} \right](\xi) + \left[ \begin{pmatrix} U \\ 0_N \end{pmatrix}, \sigma_{i,k} \begin{pmatrix} \partial_{x_i} U \\ 0_N \end{pmatrix} \right](\xi) \right) \\
&= \begin{pmatrix} \partial_t [\hat{b}(t,\cdot), \sigma_{\cdot,k}](z) \\ 0_L \\ \partial_t [\hat{b}(t,\cdot), \sigma_{\cdot,k}](z) \end{pmatrix} + \begin{pmatrix} [\hat{b}(t,\cdot), [\hat{b}(t,\cdot), \sigma_{\cdot,k}]](z) \\ 0_L \\ [\hat{b}(t,\cdot), [\hat{b}(t,\cdot), \sigma_{\cdot,k}]](z) \end{pmatrix} - \sum_{i=1}^{N} [\hat{b}(t,\cdot), \sigma_{\cdot,k}]^{(i)}(z) \begin{pmatrix} \partial_{x_i} U(x,y) \\ 0_N \end{pmatrix} \\
&\quad - \sum_{i=1}^{N} \left( \sum_{j=1}^{N} \hat{b}^{(j)}(t,z) \left( \sigma_{i,k}(z) \begin{pmatrix} \partial_{x_j} \partial_{x_i} U(x,y) \\ 0_N \end{pmatrix} + \partial_{z_j} \sigma_{i,k}(z) \begin{pmatrix} \partial_{x_i} U(x,y) \\ 0_N \end{pmatrix} \right) \right. \\
&\qquad\qquad \left. + \sigma_{i,k}(z) \begin{pmatrix} [U, \partial_{x_i} U](x,y) \\ 0_N \end{pmatrix} \right) \\
&= \begin{pmatrix} [\hat{b}(t,\cdot), [\hat{b}(t,\cdot), \sigma_{\cdot,k}]](z) + \partial_t [\hat{b}(t,\cdot), \sigma_{\cdot,k}](z) \\ 0_L \\ [\hat{b}(t,\cdot), [\hat{b}(t,\cdot), \sigma_{\cdot,k}]](z) + \partial_t [\hat{b}(t,\cdot), \sigma_{\cdot,k}](z) \end{pmatrix} \\
&\quad - \sum_{i=1}^{N} \left( [\hat{b}(t,\cdot), \sigma_{\cdot,k}]^{(i)}(z) \begin{pmatrix} \partial_{x_i} U(x,y) \\ 0_N \end{pmatrix} + \sigma_{i,k}(z) \begin{pmatrix} [U, \partial_{x_i} U](x,y) \\ 0_N \end{pmatrix} \right. \\
&\qquad\qquad \left. + \sum_{j=1}^{N} \hat{b}^{(j)}(t,z) \left( \sigma_{i,k}(z) \begin{pmatrix} \partial_{x_j} \partial_{x_i} U(x,y) \\ 0_N \end{pmatrix} + \partial_{z_j} \sigma_{i,k}(z) \begin{pmatrix} \partial_{x_i} U(x,y) \\ 0_N \end{pmatrix} \right) \right).
\end{aligned} \tag{2.64}
$$

Similarly, we can calculate

$$
\begin{aligned}
[V_l, [V_k, V_0]](\xi) &= \left[ \begin{pmatrix} \sigma_{\cdot,l} \\ 0_L \\ \sigma_{\cdot,l} \end{pmatrix}, \begin{pmatrix} [\sigma_{\cdot,k}, \hat{b}(t,\cdot)] \\ 0_L \\ [\sigma_{\cdot,k}, \hat{b}(t,\cdot)] \end{pmatrix} + \sum_{i=1}^{N} \sigma_{i,k} \begin{pmatrix} \partial_{x_i} U \\ 0_N \end{pmatrix} \right] (\xi) \\
&= \left[ \begin{pmatrix} \sigma_{\cdot,l} \\ 0_L \\ \sigma_{\cdot,l} \end{pmatrix}, \begin{pmatrix} [\sigma_{\cdot,k}, \hat{b}(t,\cdot)] \\ 0_L \\ [\sigma_{\cdot,k}, \hat{b}(t,\cdot)] \end{pmatrix} \right] (\xi) + \sum_{i=1}^{N} \left[ \begin{pmatrix} \sigma_{\cdot,l} \\ 0_L \\ \sigma_{\cdot,l} \end{pmatrix}, \sigma_{i,k} \begin{pmatrix} \partial_{x_i} U \\ 0_N \end{pmatrix} \right] (\xi) \\
&= \begin{pmatrix} \zeta_{k,l}(t,z) \\ 0_L \\ \zeta_{k,l}(t,z) \end{pmatrix} + \sum_{i,j=1}^{N} \left( \sigma_{j,l}(z) \sigma_{i,k}(z) \begin{pmatrix} \partial_{x_j} \partial_{x_i} U(x,y) \\ 0_N \end{pmatrix} \right. \\
&\qquad \left. + \sigma_{j,l}(z) \partial_{z_j} \sigma_{i,k}(z) \begin{pmatrix} \partial_{x_i} U(x,y) \\ 0_N \end{pmatrix} \right),
\end{aligned} \tag{2.65}
$$

where

$$
\begin{aligned}
\zeta_{k_1}(t,z) &:= [\sigma_{\cdot,k_1}, \hat{b}(t,\cdot)](z), \\
\zeta_{k_1,\dots,k_n}(t,z) &:= [\sigma_{\cdot,k_n}, \zeta_{k_1,\dots,k_{n-1}}(t,\cdot)](z) \quad \text{for all } n \geq 2,
\end{aligned} \tag{2.66}
$$

with any $k_1, k_2, \ldots \in \{1, \ldots, M\}$.

Having thus acquired a basic idea of the typical structure of higher iterations $[V_0, [\ldots, [V_0, V_k]]]$ or $[V_{k_n}, [\ldots, [V_{k_1}, V_0]]]$, we formulate the following two Lemmas which will help us treat these in detail later. Their proofs are straight forward, following the same line of arguments as in the above calculations of $[V_0, [V_0, V_k]]$ and $[V_l, [V_k, V_0]]$.

**Lemma 2.41.** *Let $n \in \mathbb{N}$. For all $i \in \{1, \ldots, n\}$ consider continuously differentiable functions*

$$
\begin{aligned}
[0,\infty) \times \mathbb{R}^N \ni (t,z) &\mapsto \varphi_i(t,z) \in \mathbb{R}, \\
\mathbb{R}^N \times V \ni (x,y) &\mapsto W_i(x,y) \in \mathbb{R}^{N+L}, \\
[0,\infty) \times \mathbb{R}^N \ni (t,z) &\mapsto A(t,z) \in \mathbb{R}^N
\end{aligned}
$$

*and set*

$$
W(\xi) := \begin{pmatrix} A(t,z) \\ 0_L \\ A(t,z) \end{pmatrix} - \sum_{i=1}^{n} \varphi_i(t,z) \begin{pmatrix} W_i(x,y) \\ 0_N \end{pmatrix} \quad \textit{for all } \xi \in [0,\infty) \times \mathtt{E}.
$$

*Then*

$$[V_0, W](\xi) = \begin{pmatrix} [\hat{b}(t,\cdot), A(t,\cdot)](z) + \partial_t A(t,z) \\ 0_L \\ [\hat{b}(t,\cdot), A(t,\cdot)](z) + \partial_t A(t,z) \end{pmatrix} - \sum_{i=1}^{n} \partial_t \varphi_i(t,z) \begin{pmatrix} W_i(x,y) \\ 0_N \end{pmatrix}$$
$$- \sum_{j=1}^{N} A^{(j)}(t,z) \begin{pmatrix} \partial_{x_j} U(x,y) \\ 0_N \end{pmatrix} - \sum_{i=1}^{n} \varphi_i(t,z) \begin{pmatrix} [U, W_i](x,y) \\ 0_N \end{pmatrix}$$
$$- \sum_{i=1}^{n} \sum_{j=1}^{N} \hat{b}^{(j)}(t,z) \left( \varphi_i(t,z) \begin{pmatrix} \partial_{x_j} W_i(x,y) \\ 0_N \end{pmatrix} + \partial_{z_j} \varphi_i(t,z) \begin{pmatrix} W_i(x,y) \\ 0_N \end{pmatrix} \right)$$

*for all* $\xi \in [0,\infty) \times \mathtt{E}$.

**Lemma 2.42.** *Let* $n \in \mathbb{N}$. *For all* $i \in \{1,\dots,n\}$ *consider continuously differentiable functions*

$$\mathbb{R}^N \ni z \mapsto \varphi_i(z) \in \mathbb{R},$$
$$\mathbb{R}^N \times V \ni (x,y) \mapsto W_i(x,y) \in \mathbb{R}^{N+L},$$
$$[0,\infty) \times \mathbb{R}^N \ni (t,z) \mapsto A(t,z) \in \mathbb{R}^N$$

*and set*

$$W(\xi) := \begin{pmatrix} A(t,z) \\ 0_L \\ A(t,z) \end{pmatrix} + \sum_{i=1}^{n} \varphi_i(z) \begin{pmatrix} W_i(x,y) \\ 0_N \end{pmatrix} \quad \text{for all } \xi \in [0,\infty) \times \mathtt{E}.$$

*Then*

$$[V_k, W](\xi) = \begin{pmatrix} [\sigma_{\cdot,k}, A(t,\cdot)](z) \\ 0_L \\ [\sigma_{\cdot,k}, A(t,\cdot)](z) \end{pmatrix} + \sum_{i=1}^{n} \sum_{j=1}^{N} \sigma_{j,k}(z) \varphi_i(z) \begin{pmatrix} \partial_{x_j} W_i(x,y) \\ 0_N \end{pmatrix}$$
$$+ \sum_{i=1}^{n} \sum_{j=1}^{N} \sigma_{j,k}(z) \partial_{z_j} \varphi_i(z) \begin{pmatrix} W_i(x,y) \\ 0_N \end{pmatrix}$$

*for all* $\xi \in [0,\infty) \times \mathtt{E}$ *and any* $k \in \{1,\dots,M\}$.

**Remark 2.43.** Of course, if the pertaining vector fields are of the respective form only in some open subset of $[0,\infty) \times \mathtt{E}$, all of the formulas in Lemmas 2.41 and 2.42 still hold locally in this subset.

Having collected these calculations for reference, we can now proceed to treat the star shape situation (Subsection 2.4.1) and the cascade structure (Subsection 2.4.2).

### 2.4.1 Star shape

For the star shape, our strategy is to span the entire space with vector fields that are constructed in the following fashion: Let $l \in \mathbb{N}$ and $\kappa = (k_1, \dots, k_l) \in \{1, \dots, M\}^l$ and define

$$L_{\kappa,1} := [V_{k_1}, V_0] \quad \text{and} \quad L_{\kappa,n} := [V_{k_n}, L_{\kappa,n-1}] \quad \text{for all } n \in \{2, \dots, l\}.$$

When this recursive procedure is finished, we have acquired the vector field $L_{\kappa,l}$. As this is the result of successively taking Lie brackets with vector fields that follow the "path" $\kappa = (k_1, \dots, k_l)$ through the columns of the volatility matrix, we will use the notation $L_\kappa := L_{\kappa,l}$. Our goal is to find verifiable criteria under which for some sequence $\kappa_1, \dots, \kappa_{N+L}$ of such paths (of possibly different lengths) and for some

$$W_1, \dots, W_N \in \{V_1, \dots, V_M\}^*,$$

the vector fields

$$W_1, \dots, W_N, L_{\kappa_1}, \dots, L_{\kappa_{N+L}},$$

evaluated at a suitable point $\xi \in [0, \infty) \times \mathtt{E}$, span the entire $N + L + N$-dimensional euclidean space.[6] To this end, we will have to calculate $L_\kappa$ for a general path

$$\kappa = (k_1, \dots, k_{l(\kappa)}) \in \bigcup_{l \in \mathbb{N}} \{1, \dots, M\}^l,$$

where we write $l(\kappa)$ for the path length of $\kappa$. Thanks to Lemma 2.42, these calculations are not too hard. Recall the definition of the functions $\zeta_{k_1,\dots,k_n}$, $n \in \{1, \dots, l(\kappa)\}$, that was given in (2.66).

**Lemma 2.44.** *For any $\kappa = (k_1, \dots, k_l) \in \{1, \dots, M\}^l$, $n \in \{1, \dots, l\}$, and $\xi \in [0, \infty) \times \mathtt{E}$ we have*

$$\begin{aligned} L_{\kappa,n}(\xi) = \begin{pmatrix} \zeta_{k_1,\dots,k_n}(t,z) \\ 0_L \\ \zeta_{k_1,\dots,k_n}(t,z) \end{pmatrix} &+ \sum_{i_1,\dots,i_n=1}^{N} \sigma_{i_1,k_1}(z) \cdots \sigma_{i_n,k_n}(z) \begin{pmatrix} \partial_{x_{i_1}} \cdots \partial_{x_{i_n}} U(x,y) \\ 0_N \end{pmatrix} \\ &+ \sum_{\alpha \in \mathbb{N}_0^N, 1 \le |\alpha|_1 \le n-1} p_{\kappa,n,\alpha}(z) \begin{pmatrix} \partial_x^\alpha U(x,y) \\ 0_N \end{pmatrix}, \end{aligned} \tag{2.67}$$

*where each coefficient function $p_{\kappa,n,\alpha}$ is a polynomial expression of the terms*

$$\partial_z^\beta \sigma_{i,k_j} \quad \textit{with } \beta \in \mathbb{N}_0^N,\ |\beta|_1 \le |\alpha|_1,\ i \in \{1, \dots, N\},\ j \in \{1, \dots, n\}.$$

[6] Recall that due to the Notation 2.40 we interpret these vector fields in such a way that they take values in $\mathbb{R}^{N+L+N}$.

*Proof.* Fix some $\kappa = (k_1, \dots, k_l) \in \{1, \dots, M\}^l$. We want to prove this Lemma by induction, so at first we notice that its claim is true for $n = 1$ and $n = 2$ – this was shown in (2.63) and (2.65). Assume now that this Lemma's claim holds for some $n \in \{1, \dots, l-1\}$. Then Lemma 2.42 yields

$$\begin{aligned}
L_{\kappa,n+1}(\xi) = &\begin{pmatrix} \zeta_{k_1,\dots,k_{n+1}}(t,z) \\ 0_L \\ \zeta_{k_1,\dots,k_{n+1}}(t,z) \end{pmatrix} \\
&+ \sum_{i_1,\dots,i_{n+1}=1}^{N} \sigma_{i_1,k_1}(z) \cdots \sigma_{i_{n+1},k_{n+1}}(z) \begin{pmatrix} \partial_{x_{i_1}} \cdots \partial_{x_{i_{n+1}}} U(x,y) \\ 0_N \end{pmatrix} \\
&+ \sum_{i_1,\dots,i_{n+1}=1}^{N} \sigma_{i_{n+1},k_{n+1}}(z) \partial_{z_{i_{n+1}}} [\sigma_{i_1,k_1}(z) \cdots \sigma_{i_n,k_n}(z)] \begin{pmatrix} \partial_{x_{i_1}} \cdots \partial_{x_{i_n}} U(x,y) \\ 0_N \end{pmatrix} \\
&+ \sum_{\alpha \in \mathbb{N}_0^N, 1 \le |\alpha|_1 \le n-1} \sum_{i=1}^{N} \sigma_{i,k_{n+1}}(z) p_{\kappa,n,\alpha}(z) \begin{pmatrix} \partial_{x_i} \partial_x^\alpha U(x,y) \\ 0_N \end{pmatrix} \\
&+ \sum_{\alpha \in \mathbb{N}_0^N, 1 \le |\alpha|_1 \le n-1} \sum_{i=1}^{N} \sigma_{i,k_{n+1}}(z) \partial_{z_i} p_{\kappa,n,\alpha}(z) \begin{pmatrix} \partial_x^\alpha U(x,y) \\ 0_N \end{pmatrix},
\end{aligned}$$

where the summands in the first two lines are already exactly of the shape we are aiming for. All of the other terms correspond to derivatives of $U$ that are of the order $n$ at most. By the induction hypothesis, each $p_{\kappa,n,\alpha}$ is a polynomial expression of the terms $\partial_z^\beta \sigma_{i,k_j}$ with $\beta \in \mathbb{N}_0^N$, $|\beta|_1 \le |\alpha|_1$, $i \in \{1, \dots, N\}$, $j \in \{1, \dots, n\}$. Trivially, the same is then true for any partial derivative $\partial_{z_i} p_{\kappa,n,\alpha}$ and hence all of the respective coefficients in the above formula are polynomial expressions of $\partial_z^\beta \sigma_{i,k_j}$ with $i \in \{1, \dots, N\}$, $j \in \{1, \dots, n+1\}$ and $\beta \in \mathbb{N}_0^N$ with $|\beta|_1$ bounded by the order of the respective derivative of $U$. Thus, $L_{\kappa,n+1}(\xi)$ is indeed of the desired form, which completes the proof. □

**Remark 2.45.** Having a closer look at the formula in the proof of Lemma 2.44, we note that the coefficients $p_{\kappa,n,\alpha}$ can be calculated by the following scheme:

1. For all $n \in \mathbb{N}$ define
$$p_{\kappa,n,0_N} := 0$$
and
$$p_{\kappa,n,\alpha} := \prod_{j=1}^{n} \sigma_{i_j,k_j}$$
for all $\alpha \in \mathbb{N}_0^N$ with $|\alpha|_1 = n$ and $\partial_x^\alpha = \partial_{x_{i_1}} \cdots \partial_{x_{i_n}}$.

2. Now we can recursively calculate
$$p_{\kappa,n,\alpha} = \sum_{i=1}^{N} \sigma_{i,k_n} \left( \partial_{z_i} p_{\kappa,n-1,\alpha} + p_{\kappa,n-1,\alpha-e_i} \cdot 1_{\alpha_i \neq 0} \right)$$

for all $n \geq 2$ and $\alpha \in \mathbb{N}_0^N$ with $|\alpha|_1 \in \{1, \dots, n-1\}$, where $e_i$ denotes the $i$-th canonical unit vector in $\mathbb{R}^N$.

Let us now consider a finite sequence of paths

$$\kappa_1, \dots, \kappa_{N+L} \in \bigcup_{l \in \mathbb{N}} \{1, \dots, M\}^l$$

and the corresponding sequence of vector fields

$$L_{\kappa_1}, \dots, L_{\kappa_{N+L}} \in \mathcal{L}_{\max\{l(\kappa_1), \dots, l(\kappa_{N+L})\}},$$

as defined in (2.67). Let $\xi = (t, x, y, z) \in [0, \infty) \times \mathtt{E}$ and assume that (H2) holds in $z$, i.e. there are

$$w_1, \dots, w_N \in \{\sigma_{\cdot,1}, \dots, \sigma_{\cdot,M}\}^*$$

such that $w_1(z), \dots, w_N(z) \in \mathbb{R}^N$ are linearly independent. Setting

$$W_1 := \begin{pmatrix} w_1 \\ 0_L \\ w_1 \end{pmatrix}, \dots, W_N := \begin{pmatrix} w_N \\ 0_L \\ w_N \end{pmatrix} \in \{V_1, \dots, V_M\}^*,$$

we want to find a sufficient condition for

$$W_1(z), \dots, W_N(z), L_{\kappa_1}(\xi), \dots, L_{\kappa_{N+L}}(\xi) \in \mathcal{L}(\xi)$$

to be linearly independent. We first note that for each $\kappa_n = (k_{n,1}, \dots, k_{n,l(\kappa_n)})$ the term

$$\zeta_{k_{n,1}, \dots, k_{n,l(\kappa_n)}}(t, z) \in \mathbb{R}^N$$

can of course be expressed as a linear combination of the linearly independent vectors $w_1(z), \dots, w_N(z) \in \mathbb{R}^N$, and consequently the first summand

$$\begin{pmatrix} \zeta_{k_{n,1}, \dots, k_{n,l(\kappa_n)}}(t, z) \\ 0_L \\ \zeta_{k_{n,1}, \dots, k_{n,l(\kappa_n)}}(t, z) \end{pmatrix} \in \mathbb{R}^{N+L+N}$$

of $L_{\kappa_n}(\xi)$ is a linear combination of $W_1(z), \dots, W_N(z) \in \mathbb{R}^{N+L+N}$. All of its other summands are zero in their last $N$ components. Therefore, the vectors

$$W_1(z), \dots, W_N(z), L_{\kappa_1}(\xi), \dots, L_{\kappa_{N+L}}(\xi) \in \mathbb{R}^{N+L+N}$$

are linearly independent if and only if

$$\bar{L}_{\kappa_1}(\xi), \dots, \bar{L}_{\kappa_{N+L}}(\xi) \in \mathbb{R}^{N+L}$$

are, where

$$\begin{aligned}\bar{L}_{\kappa_n}(\xi) = &\sum_{i_1,\dots,i_{l(\kappa_n)}=1}^{N} \sigma_{i_1,k_{n,1}}(z)\cdots\sigma_{i_{l(\kappa_n)},k_{n,l(\kappa_n)}}(z)\partial_{x_{i_1}}\cdots\partial_{x_{i_{l(\kappa_n)}}}U(x,y)\\ &+ \sum_{\alpha\in\mathbb{N}_0^N,1\le|\alpha|\le l(\kappa_n)-1} p_{\kappa_n,l(\kappa_n),\alpha}(z)\partial_x^\alpha U(x,y)\end{aligned} \tag{2.68}$$

for every $n \in \{1,\dots,N+L\}$. Note that these vector fields in fact no longer depend on time, which is why we can simply write

$$\bar{L}_{\kappa_n}(x,y,z) = \bar{L}_{\kappa_n}(\xi).$$

The following Theorem collects the insight we have gained so far.

**Theorem 2.46.** *Let* $(x,y,z) \in \mathrm{int}(\mathtt{E})$ *and assume that the following two conditions hold.*

*(i) Condition (H2) holds in* $z$.

*(ii) There are*

$$\kappa_1,\dots,\kappa_{N+L} \in \bigcup_{l\in\mathbb{N}}\{1,\dots,M\}^l$$

*such that*

$$\bar{L}_{\kappa_1}(x,y,z),\dots,\bar{L}_{\kappa_{N+L}}(x,y,z) \quad \textit{are linearly independent.}$$

*Then the local weak Hörmander condition holds at* $(x,y,z)$.

*Proof.* As we have reasoned above, the assumptions of this Theorem imply that there are $W_1,\dots,W_N \in \{V_1,\dots,V_M\}^*$ such that the span of

$$W_1(z),\dots,W_N(z),L_{\kappa_1}(t,x,y,z),\dots,L_{\kappa_{N+L}}(t,x,y,z) \in \mathcal{L}\big(t,x,y,z\big)$$

has the maximum dimension $N+L+N$ for all $t \in [0,\infty)$, and this is sufficient for (2.56). $\square$

**Remark 2.47.** We note that the signal (the only time-dependent object present) is not part of any of the vector fields featured in Theorem 2.46, which is why uniformity in time (as required for the local weak Hörmander condition) is entirely unproblematic. Since time-derivatives do not occur at all in the construction in this section, we could in principal exclude the signal $S$ from our standing smoothness assumption (H1).

In general, the assumptions of Theorem 2.46 can turn out to be still very hard to verify, since the vector fields in (2.68) have a rather complicated structure. However, we can present certain interesting situations in which they can be radically simplified.

**Corollary 2.48.** *Assume that $\sigma$ is constant and that its only value is a surjective matrix. Then the local weak Hörmander condition holds at any $(x, y, z) \in \operatorname{int}(\mathtt{E})$ where for some $\kappa_1, \dots, \kappa_{N+L} \in \bigcup_{l\in\mathbb{N}}\{1, \dots, M\}^l$ the vectors*

$$\sum_{i_1,\dots,i_{l(\kappa_n)}=1}^{N} \sigma_{i_1,k_{n,1}} \cdots \sigma_{i_{l(\kappa_n)},k_{n,l(\kappa_n)}} \partial_{x_{i_1}} \cdots \partial_{x_{i_{l(\kappa_n)}}} \begin{pmatrix} F(x,y) \\ G(x,y) \end{pmatrix}, \quad n \in \{1, \dots, N+L\},$$

*are linearly independent.*

*Proof.* Firstly, surjectivity of $\sigma(z) \equiv \sigma$ implies condition (i) of Theorem 2.46 for all $z \in \mathbb{R}^N$, as was already explained right after the introduction of (H2). Secondly, it follows from Remark 2.45 that for constant $\sigma$ all of the coefficients $p_{\kappa_n,l(\kappa_n),\alpha}$ with $1 \leq |\alpha|_1 \leq l(\kappa_n)-1$ vanish. Therefore, $\bar{L}_{\kappa_1}, \dots, \bar{L}_{\kappa_{N+L}}$ can be simplified to the respective expressions given in this Corollary. Their linear independence yields condition (ii) of Theorem 2.46 which can then be applied to complete the proof. □

Next, we will focus on the situation in which $M = N$, and the sequence of paths $\kappa_1, \dots, \kappa_{N+L}$ is such that for some $k_1, \dots, k_{N+L} \in \{1, \dots, N\}$ we have

$$\kappa_n = (k_1, \dots, k_n) \quad \text{for every } n \in \{1, \dots, N+L\}, \tag{2.69}$$

i.e. each path is a one-step extension of its predecessor.

**Corollary 2.49.** *Let $M = N$ and assume that $\sigma\colon \mathbb{R}^N \to \mathbb{R}^{N\times N}$ takes values only in the set of invertible diagonal matrices. Then the local weak Hörmander condition holds at any $(x, y, z) \in \operatorname{int}(\mathtt{E})$ where for some $k_1, \dots, k_{N+L} \in \{1, \dots, N\}$ the vectors*

$$\partial_{x_{k_1}} \cdots \partial_{x_{k_n}} \begin{pmatrix} F(x,y) \\ G(x,y) \end{pmatrix}, \quad n \in \{1, \dots, N+L\}, \tag{2.70}$$

*are linearly independent.*

*Proof.* Once again, we want to apply Theorem 2.46. By invertibility of $\sigma(z)$, all of its columns are linearly independent, and hence condition (i) is trivially fulfilled for all $z \in \mathbb{R}^N$. Only the validity of condition (ii) remains to be shown. We let $n \in \{1, \dots, N+L\}$ and we use the same notation as in (2.69). The vector field $\bar{L}_{\kappa_n}$ can be calculated by the formula (2.68), and since $\sigma(z)$ is a diagonal matrix, the first sum is reduced to just the one summand

$$\sigma_{k_1,k_1}(z) \cdots \sigma_{k_n,k_n}(z) \partial_{x_{k_1}} \cdots \partial_{x_{k_n}} U(x,y),$$

while the second sum is reduced to

$$\sum_{l=1}^{n-1} p_{\kappa_n,n,\alpha(l)}(z) \partial_{x_{k_1}} \cdots \partial_{x_{k_l}} U(x,y)$$

with $\alpha(l) := e_{k_1} + \ldots + e_{k_l}$, as follows immediately from Remark 2.45. Invertibility of $\sigma(z)$ yields that $\sigma_{k_1,k_1}(z) \cdots \sigma_{k_n,k_n}(z)$ can never be zero, so none of the leading coefficients of $\bar{L}_{\kappa_n}(x, y, z)$ vanish. Hence, starting with $n = 1$, we can successively add multiples of $\bar{L}_{\kappa_n}(x, y, z)$ to $\bar{L}_{\kappa_{n+1}}(x, y, z)$ in order to eliminate any summand in $\bar{L}_{\kappa_{n+1}}(x, y, z)$ with derivatives of the order less than $n + 1$. The resulting vectors are clearly linearly independent if and only if $\bar{L}_{\kappa_1}(x, y, z), \ldots, \bar{L}_{\kappa_{N+L}}(x, y, z)$ are. This leads to the conclusion that linear independence of

$$\sigma_{k_1,k_1}(z) \cdots \sigma_{k_n,k_n}(z) \partial_{x_{k_1}} \cdots \partial_{x_{k_n}} U(x, y), \quad n \in \{1, \ldots, N + L\},$$

is sufficient for condition (ii) of Theorem 2.46. As leaving out scaling factors has no effect on linear independence, the proof is completed. □

**Corollary 2.50.** *If $M = N = 1$ and $\sigma \colon \mathbb{R} \to \mathbb{R}$ is strictly positive, the local weak Hörmander condition holds at any $(x, y, z) \in \operatorname{int}(\mathtt{E})$ where the vectors*

$$\partial_x^n \begin{pmatrix} F(x, y) \\ G(x, y) \end{pmatrix}, \quad n \in \{1, \ldots, 1 + L\},$$

*are linearly independent.*

*Proof.* For $N = 1$ we have no other choice than $k_n = 1$ for all $n \in \{1, \ldots, 1 + L\}$, and the assertion follows immediately from Corollary 2.49. □

**Remark 2.51.** Since all of the calculations in this section are done locally, the representations of $L_\kappa$ (and $\bar{L}_\kappa$) for different coefficient functions coincide in some open subset of $[0, \infty) \times \mathtt{E}$, when the respective coefficient functions coincide in that set (compare Remark 2.43). In particular, Corollaries 2.49 and 2.50 remain valid if $\sigma$ is an invertible diagonal matrix (or strictly positive) only in a small neighbourhood of the particular point $z$.

Corollary 2.50 contains Theorem 3 of [36] as a special case, and thus the following example can be treated in the same way as in Section 5.3 of [36].

**Example 2.52.** Let us consider the stochastic Hodgkin-Huxley model (SHH) with Ornstein-Uhlenbeck type input with strictly positive volatility $\sigma \in C^\infty(\mathbb{R})$. In order to apply Corollary 2.50, we have to calculate the derivatives of $F$ and $G$ from (1.5) and (1.6) with respect to $x$ in $\mathbb{R} \times (0, 1)^3$. We get

$$\partial_x F(x, y) = -36 y_1^4 - 120 y_2^3 y_3 < 0 \quad \text{for all } (x, y) \in \mathbb{R} \times (0, 1)^3$$

and then

$$\partial_x^n F(x, y) = 0 \quad \text{for all } (x, y) \in \mathbb{R} \times (0, 1)^3$$

for any $n \geq 2$. Hence, the vectors

$$\partial_x^n \begin{pmatrix} F(x,y) \\ G(x,y) \end{pmatrix}, \quad n \in \{1, \dots, 4\},$$

are linearly independent if and only if

$$\partial_x^n G(x,y) = \begin{pmatrix} \partial_x^n \alpha_1(x)(1-y_1) - \partial_x^n \beta_1(x) y_1 \\ \partial_x^n \alpha_2(x)(1-y_2) - \partial_x^n \beta_2(x) y_2 \\ \partial_x^n \alpha_3(x)(1-y_3) - \partial_x^n \beta_3(x) y_3 \end{pmatrix}, \quad n \in \{2, 3, 4\},$$

are. This is of course equivalent to the condition

$$D(x,y) := \det \begin{pmatrix} \partial_x^2 G(x,y) & \partial_x^3 G(x,y) & \partial_x^4 G(x,y) \end{pmatrix} \neq 0.$$

This determinant is discussed numerically in Section 5.4 of [36]. In particular, this numerical study yields that in the equilibrium $(x^*, y^*)$ of the zero-input deterministic Hodgkin-Huxley system we have $D(x^*, y^*) \neq 0$.[7] Continuing our line of thought from Example 2.28, this is the other half of assumption 4 of Theorem 2.3. We have taken care of assumptions 1 and 5 in Examples 2.4 and 2.16, respectively. Since smoothness of the coefficients is evident and the signal $S$ is smooth and periodic by assumption, this means that every requirement of Theorem 2.3 is fulfilled, and hence the stochastic Hodgkin-Huxley system is positive Harris recurrent (compare [37, Theorem 2.7]).

**Example 2.53.** In order to provide an elementary application of Corollary 2.49 with $N > 1$, we take a look at the toy example (see Example 1.4) and show that, for a suitable choice of the coefficient functions $g$ and $h$, the local weak Hörmander condition holds at $(x, y, z) \in \operatorname{int}(\mathtt{E})$ whenever $x$ is taken from some small open environment $\mathcal{U} \subset \mathbb{R}^N$ of $x^* = 0_N$. For $x$-values close to this point, the function $j$ from (1.15) vanishes and is of no relevance for our considerations. Let $k_1, \dots, k_{N+L} \in \{1, \dots, N\}$. For $F$ and $G$ as in (1.15) and (1.19), the vectors from (2.70) are clearly linearly independent if and only if

$$e_{k_1} + \partial_{x_{k_1}} \begin{pmatrix} h(x) \\ g(x) \end{pmatrix}, \partial_{x_{k_n}} \cdots \partial_{x_{k_1}} \begin{pmatrix} h(x) \\ g(x) \end{pmatrix}, \quad n \in \{2, \dots, N+L\}, \tag{2.71}$$

are linearly independent. For a simple example let us choose $L = M = N = 2$ and suppose that for $x \in \mathcal{U}$ we have

$$h(x) = \begin{pmatrix} x_1 \\ x_1 x_2 \end{pmatrix}, \quad g(x) = \begin{pmatrix} x_1^2 x_2 + 1 \\ x_1^2 x_2^2 + 1 \end{pmatrix}. \tag{2.72}$$

[7] For the sake of completeness, we should mention Proposition 7 from [36] which states that the set of all $(x,y) \in \mathbb{R} \times (0,1)^3$ where $D(x,y)$ vanishes has in fact Lebesgue measure zero.

Note that, for sufficiently small $\mathcal{U}$, it is obviously possible to extend this choice for $h$ beyond $\mathcal{U}$ while still obeying (1.17). Calculating the terms from (2.71) with $k_1 = k_3 = 1$, $k_2 = k_4 = 2$, the resulting vectors

$$\begin{pmatrix} 2 \\ x_2 \\ 2x_1x_2 \\ 2x_1x_2^2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2x_1 \\ 4x_1x_2 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 2 \\ 4x_2 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 4 \end{pmatrix}$$

are trivially linearly independent. Thus, the local weak Hörmander condition is satisfied in any $(x, y, z) \in \operatorname{int}(\mathtt{E})$ where $x \in \mathcal{U}$.

Together with Examples 2.5, 2.17, and 2.29, this means that for any smooth periodic signal we can apply Theorem 2.3. The respective degenerate diffusion with internal variables and randomly perturbed time-inhomogeneous deterministic input that is derived from (1.14) is therefore positive Harris recurrent.

### 2.4.2 Cascade structure

For this section, we will stick to the case

$$M = N = 1.$$

In other words, the $x$- and $z$-variables have only one component each, and a typical element of the state space is denoted as

$$\xi = (t, x, y, z) = (t, x, y_1, \dots, y_L, z)^\top \in [0, \infty) \times \mathtt{E} = [0, \infty) \times \mathbb{R} \times V \times \mathbb{R}.$$

Hence, the function $F\colon \mathbb{R} \times V \to \mathbb{R}$ also has only one component and

$$U(x, y) = \begin{pmatrix} F(x, y) \\ G(x, y) \end{pmatrix} = \begin{pmatrix} F(x, y) \\ G_1(x, y) \\ \vdots \\ G_L(x, y) \end{pmatrix} \in \mathbb{R}^{1+L} \quad \text{for all } (x, y) \in \mathbb{R} \times V.$$

We set

$$\mathtt{B} := \{z \in \mathbb{R} \mid \sigma(z) \neq 0\} \subset \mathbb{R} \tag{2.73}$$

which is open, since $\sigma$ is continuous. We tacitly assume that $\mathtt{B}$ is non-empty, since otherwise any randomness would be removed from our model. Note that since in dimension $M = N = 1$ the condition (H2) becomes much simpler, $\mathtt{B}$ is in fact the set of all $z \in \mathbb{R}$ for which (H2) holds. Moreover, we fix an open set $\mathtt{A} \subset \mathbb{R} \times V$ and write

$$\mathtt{E}^* := \mathtt{A} \times \mathtt{B} \subset \mathtt{E}.$$

Let us first give a formal definition of what we mean by a cascade structure.

**(H3) Cascade condition for $G$:** In $\mathtt{A}$ we have

1. $\partial_x G_1, \partial_{y_1} G_2, \partial_{y_2} G_3, \dots, \partial_{y_{L-1}} G_L \neq 0$,
2. $\partial_{y_i} G_j = 0$ for all $i \in \{1, \dots, L\}$ and $j \in \{i+2, \dots, L+1\}$.

A more compact formulation of (H3) would be that in $\mathtt{A}$ for all $i \in \{1, \dots, L\}$ we have

$$\partial_{\xi_i} U_{i+1} \neq 0 \quad \text{and} \quad \partial_{\xi_i} U_j = 0 \text{ for all } j \in \{i+2, \dots, L+1\},$$

or in less strict but maybe more intuitive notation

$$\partial_{\xi_i} U = \begin{pmatrix} \square_i \\ \neq 0 \\ 0_{L-i} \end{pmatrix} \quad \text{for all } i \in \{1, \dots, L\},$$

where $\square_i \in \mathbb{R}^i$ indicates that we do not care about the values of these first $i$ entries. Intuitively speaking, while the variable $\xi_i$ may or may not influence the first $i$ components of $U$, it definitely influences $U_{i+1}$ and definitely does not immediately influence any components afterwards – hence the term cascade structure.

This setting is inspired by Section 5.3 of [17], where the authors discuss an approximating diffusion for a model of interacting neurons which are divided into two groups. In each of these groups, a current is passed on from one neuron to the next, while only at the ends of these chains neurons are subject to noise and interact with neurons from the other group. The diffusion given by (5.26) in [17] features two cascades that are similar in nature to the scenario we study in this section.

If we start with the vector field

$$V_1 = \begin{pmatrix} \sigma \\ 0_L \\ \sigma \end{pmatrix}$$

corresponding to the volatility and then keep taking Lie brackets with the vector field $V_0$ corresponding to the drift, intuition suggests that the structure from (H3) will turn up there again in some sense, providing more and more linearly independent vectors. So, this time we set

$$L_1 := [V_0, V_1], \quad \text{and} \quad L_n := [V_0, L_{n-1}] \quad \text{for all } n \geq 2. \tag{2.74}$$

We have already calculated $L_1$ and $L_2$ in (2.62) and (2.64). Using Lemma 2.41, we are able to get a better grasp of what $L_n$ looks like for larger $n$. This is the content of the following Lemma.

**Lemma 2.54.** *Let $n \in \{1, \dots, L\}$ and assume that the cascade condition (H3) holds. Then for all $\xi = (t, x, y, z) \in [0, \infty) \times \mathtt{E}^*$ we can write*

$$L_n(\xi) = \begin{pmatrix} A_n(t,z) \\ 0_L \\ A_n(t,z) \end{pmatrix} - \begin{pmatrix} \bar{L}_n(\xi) \\ 0 \end{pmatrix}, \tag{2.75}$$

*where*

$$\bar{L}_n(\xi) := \sum_{i=1}^{m_n - 1} \varphi_{n,i}(t,z) W_{n,i}(x,y) + (-1)^{n+1} \sigma(z) W_{n,m_n}(x,y) \tag{2.76}$$

*for some $m_n \in \mathbb{N}$ and smooth functions*

$$\begin{aligned} [0,\infty) \times \mathtt{B} \ni (t,z) &\mapsto \varphi_{n,i}(t,z) \in \mathbb{R}, && i \in \{1, \dots, m_n - 1\}, \\ \mathtt{A} \ni (x,y) &\mapsto W_{n,i}(x,y) \in \mathbb{R}^{1+L}, && i \in \{1, \dots, m_n\}, \\ [0,\infty) \times \mathtt{B} \ni (t,z) &\mapsto A_n(t,z) \in \mathbb{R}, \end{aligned}$$

*with*

$$W_{n,i} = \begin{pmatrix} \square_n \\ 0_{1+L-n} \end{pmatrix} \quad \text{for all } i \in \{1, \dots, m_n - 1\}, \quad W_{n,m_n} = \begin{pmatrix} \square_n \\ \neq 0 \\ 0_{1+L-(n+1)} \end{pmatrix}$$

*everywhere in* $\mathtt{A}$.

*Proof.* In analogy to Lemma 2.44, we prove this Lemma by induction. As seen in (2.62), a representation of $L_n$ as in (2.75) holds for $n = 1$. We assume now that such a representation holds for some $n \in \{1, \dots, L-1\}$. Set $\varphi_{n,m_n}(t,z) := (-1)^{n+1}\sigma(z)$ for all $(t,z) \in [0,\infty) \times \mathtt{B}$ and let $\xi = (t,x,y,z) \in [0,\infty) \times \mathtt{E}^*$. Then, according to Lemma 2.41, we obtain

$$\begin{aligned} L_{n+1}(\xi) = {} & \begin{pmatrix} [\hat{b}(t,\cdot), A_n(t,\cdot)](z) + \partial_t A_n(t,z) \\ 0_L \\ [\hat{b}(t,\cdot), A_n(t,\cdot)](z) + \partial_t A_n(t,z) \end{pmatrix} - \sum_{i=1}^{m_n} \partial_t \varphi_{n,i}(t,z) \begin{pmatrix} W_{n,i}(x,y) \\ 0 \end{pmatrix} \\ & - A_n(t,z) \begin{pmatrix} \partial_x U(x,y) \\ 0 \end{pmatrix} - \sum_{i=1}^{m_n} \varphi_{n,i}(t,z) \begin{pmatrix} [U, W_{n,i}](x,y) \\ 0 \end{pmatrix} \\ & - \sum_{i=1}^{m_n} \hat{b}(t,z) \left( \varphi_{n,i}(t,z) \begin{pmatrix} \partial_x W_{n,i}(x,y) \\ 0 \end{pmatrix} + \partial_z \varphi_{n,i}(t,z) \begin{pmatrix} W_{n,i}(x,y) \\ 0 \end{pmatrix} \right). \end{aligned}$$

The first summand is already of the desired type, and every other summand has a vanishing last component. Therefore, we are left with the task to study the term

$$\begin{aligned}\bar{L}_{n+1}(\xi) :=& \sum_{i=1}^{m_n} \partial_t\varphi_{n,i}(t,z)W_{n,i}(x,y) + A_n(t,z)\partial_x U(x,y) + \sum_{i=1}^{m_n} \varphi_{n,i}(t,z)[U,W_{n,i}](x,y) \\ &+ \sum_{i=1}^{m_n} \hat{b}(t,z)\Big(\varphi_{n,i}(t,z)\partial_x W_{n,i}(x,y) + \partial_z\varphi_{n,i}(t,z)W_{n,i}(x,y)\Big).\end{aligned} \tag{2.77}$$

We have to prove that we can extract one summand from this sum that is of the type $(-1)^{n+2}\sigma(z)W(x,y)$ with some

$$W(x,y) = \begin{pmatrix} \square_{n+1} \\ \neq 0 \\ 0_{1+L-(n+2)} \end{pmatrix},$$

while every other summand is of the type $\varphi(t,z)V(x,y)$ with some

$$V(x,y) = \begin{pmatrix} \square_{n+1} \\ 0_{1+L-(n+1)} \end{pmatrix}.$$

In order to do so, we will expand the expressions in (2.77) step for step and discuss the occurring terms one by one.

Due to our induction hypothesis, we get that

$$W_{n,i}(x,y),\ \partial_x W_{n,i}(x,y) \text{ are of the type } \begin{pmatrix} \square_{n+1} \\ 0_{1+L-(n+1)} \end{pmatrix} \text{ for all } i \in \{1,\dots,m_n\},$$

and due to (H3), the same is true for $\partial_x U(x,y)$. This is why it remains to look at

$$\sum_{i=1}^{m_n} \varphi_{n,i}(t,z)[U,W_{n,i}](x,y) = \sum_{i=1}^{m_n} \varphi_{n,i}(t,z) \sum_{j=1}^{1+L} \Big(U^{(j)}\partial_{\xi_j}W_{n,i} - W_{n,i}^{(j)}\partial_{\xi_j}U\Big)(x,y).$$

Thanks again to the induction hypothesis,

$$\partial_{\xi_j}W_{n,i}(x,y) = \begin{pmatrix} \square_{n+1} \\ 0_{1+L-(n+1)} \end{pmatrix} \text{ for all } i \in \{1,\dots,m_n\} \text{ and } j \in \{1,\dots,1+L\},$$

which is of course still valid for the same terms multiplied by $U^{(j)}(x,y)$. Therefore, it remains to discuss

$$-\sum_{i=1}^{m_n} \varphi_{n,i}(t,z) \sum_{j=1}^{1+L} W_{n,i}^{(j)}(x,y)\partial_{\xi_j}U(x,y).$$

Using the induction hypothesis once again, we know in particular that

$$W_{n,i}^{(j)}(x,y) = 0 \quad \text{for all } i \in \{1,\dots,m_n-1\} \text{ and } j \in \{n+1,\dots,1+L\}$$

and

$$W^{(j)}_{n,m_n}(x,y) = 0 \quad \text{for all } j \in \{n+2, \dots, 1+L\}.$$

Thus, it remains to look at

$$-\sum_{i=1}^{m_n-1} \varphi_{n,i}(t,z) \sum_{j=1}^{n} W^{(j)}_{n,i}(x,y)\partial_{\xi_j} U(x,y) - (-1)^{n+1}\sigma(z) \sum_{j=1}^{n+1} W^{(j)}_{n,m_n}(x,y)\partial_{\xi_j} U(x,y).$$

Using (H3) again, we see that

$$\partial_{\xi_j} U(x,y) = \begin{pmatrix} \square_{n+1} \\ 0_{1+L-(n+1)} \end{pmatrix} \quad \text{for all } j \in \{1, \dots, n\},$$

which implies that it now remains to comment on the summand

$$(-1)^{n+2}\sigma(z) W^{(n+1)}_{n,m_n}(x,y)\partial_{\xi_{n+1}} U(x,y).$$

The two leading scalar factors do not vanish (by (2.73) and by the induction hypothesis), and

$$\partial_{\xi_{n+1}} U(x,y) = \begin{pmatrix} \square_{n+1} \\ \neq 0 \\ 0_{1+L-(n+2)} \end{pmatrix}$$

thanks to (H3). Thus, $\bar{L}_{n+1}(\xi)$ is in fact of the desired form with

$$W_{n+1,m_{n+1}}(x,y) := W^{(n+1)}_{n,m_n}(x,y)\partial_{\xi_{n+1}} U(x,y) = \begin{pmatrix} \square_{n+1} \\ \neq 0 \\ 0_{1+L-(n+2)} \end{pmatrix},$$

and the proof by induction is complete. $\square$

Lemma 2.54 states that the structure of $U$ that is given by (H3) is indeed passed on to $L_1, \dots, L_L$ in some sense. In particular, this entails the following Corollary.

**Corollary 2.55.** *Assume that (H3) holds. Then for all $\xi \in [0,\infty) \times \mathtt{E}^*$ the vectors*

$$\bar{L}_1(\xi), \dots, \bar{L}_L(\xi) \in \mathbb{R}^{1+L}$$

*from* (2.76) *are linearly independent.*

*Proof.* An immediate consequence of Lemma 2.54 is that for arguments from $[0,\infty) \times \mathtt{E}^*$ the vector fields $\bar{L}_1, \dots, \bar{L}_L$ are of the type

$$\bar{L}_n = \begin{pmatrix} \square_n \\ \neq 0 \\ 0_{1+L-(n+1)} \end{pmatrix} \quad \text{for all } n \in \{1, \dots, L\},$$

which in turn yields their linear independence. $\square$

Let us now assume that (H3) holds and let $\xi = (t, x, y, z) \in [0, \infty) \times \mathtt{E}^*$. Lemma 2.54 implies that $L_n$ differs from

$$-\begin{pmatrix} \bar{L}_n(\xi) \\ 0 \end{pmatrix}$$

only by a multiple of the volatility $V_1(z)$ which is non-zero since $z \in \mathtt{B}$. Corollary 2.55 above lets us conclude that

$$V_1(z), L_1(\xi), \ldots, L_L(\xi)$$

are linearly independent. However, $1 + L$ vectors are obviously not enough to span $\mathbb{R}^{1+L+1}$ which is our goal in view of (2.56). What we need is one more vector that decouples the first and the last variable (which coincide in $V_1(z)$). Defining an additional vector field $L_{L+1}$ by taking the Lie bracket of $V_0$ and $L_L$ again does not seem particularly promising, as the cascade has already run through all of the variables.[8] Instead, we work with the volatility as we did in the star shape situation, i.e. we set

$$\begin{aligned} L_{L+1}(\xi) :=& [V_1, [V_1, V_0]](\xi) = -[V_1, L_1](\xi) \\ =& \begin{pmatrix} \zeta_{1,1}(t, z) \\ 0_L \\ \zeta_{1,1}(t, z) \end{pmatrix} + \sigma^2(z) \begin{pmatrix} \partial_x^2 U(x, y) \\ 0 \end{pmatrix} + \sigma(z)\sigma'(z) \begin{pmatrix} \partial_x U(x, y) \\ 0 \end{pmatrix}, \end{aligned}$$

where the last equality follows from (2.63). If (H3) holds, this expression takes the slightly simpler shape

$$L_{L+1}(\xi) = \begin{pmatrix} \zeta_{1,1}(t, z) \\ 0_L \\ \zeta_{1,1}(t, z) \end{pmatrix} + \sigma^2(z) \begin{pmatrix} \partial_x^2 F(x, y) \\ \partial_x^2 G_1(x, y) \\ 0_L \end{pmatrix} + \sigma(z)\sigma'(z) \begin{pmatrix} \partial_x F(x, y) \\ \partial_x G_1(x, y) \\ 0_L \end{pmatrix}.$$

Using similar arguments as in the proof of Corollary 2.49, we see after subtracting a suitable linear combination of $V_0(z)$ and $L_1(\xi)$ that $L_{L+1}(\xi)$ is linearly independent from $V_1(z), L_1(\xi), \ldots, L_L(\xi)$ whenever the following condition is fulfilled:

**(H4) Decoupling of $X$ and $Z$:** For all $(x, y) \in \mathtt{A}$ we have

$$\partial_x F(x, y)\partial_x^2 G_1(x, y) \neq \partial_x^2 F(x, y)\partial_x G_1(x, y).$$

Note that (if (H3) already holds) this is clearly equivalent to the following more handy variant.

**(H4') Decoupling of $X$ and $Z$:** For all $(x, y) \in \mathtt{A}$ the vectors $\partial_x U(x, y)$ and $\partial_x^2 U(x, y)$ are linearly independent.

[8] Of course, this construction might work in certain cases, but in general we simply cannot tell.

This section's essence is summed up in the following Theorem which follows immediately from the considerations above.

**Theorem 2.56.** *If (H3) and (H4) are valid, the local weak Hörmander condition holds at any* $(x, y, z) \in \mathtt{E}^*$.

*Proof.* As follows from Corollary 2.55 and the reasoning outlined above, the assumptions (H3) and (H4) imply that for all $\xi = (t, x, y, z) \in [0, \infty) \times \mathtt{E}^*$ the vectors

$$V_1(z), L_1(\xi), \ldots, L_{L+1}(\xi) \in \mathcal{L}(\xi)$$

are linearly independent. This is sufficient for the local weak Hörmander condition to hold at $(x, y, z)$. $\square$

**Example 2.57.** Let us return to the toy example, but this time with the choice for $G$ that was taken in (1.20). Let $\mathtt{A} \subset (-1, 1) \times [0, \infty)^L$ be any non-empty open set, assume that the set $\mathtt{B}$ from (2.73) is non-empty, and let $(t, x, y, z) \in [0, \infty) \times \mathtt{E}^*$. Since

$$\partial_x G(x, y) = \begin{pmatrix} g_1'(x) \\ 0_{L-1} \end{pmatrix} \quad \text{and} \quad \partial_{y_k} G(x, y) = \begin{pmatrix} 0_{k-1} \\ -1 \\ g_{k+1}'(y_k) \\ 0_{L-(k+1)} \end{pmatrix} \quad \text{for all } k \in 1, \ldots, L-1,$$

the condition (H3) is obviously fulfilled if and only if

$$g_1'(x) \neq 0 \quad \text{and} \quad g_{k+1}'(y_k) \neq 0 \quad \text{for all } k \in 1, \ldots, L-1, \tag{2.78}$$

for all $(x, y) \in \mathtt{A}$. Note that by (1.18) the function $j$ vanishes in $\mathtt{A}$ and hence

$$\partial_x F(x, y) = -f(y)(1 + h'(x)), \quad \partial_x^2 F(x, y) = -f(y) h''(x), \quad \partial_x^2 G_1(x, y) = g_1''(x).$$

Since $f$ is strictly positive, this implies that the condition (H4) is equivalent to

$$(1 + h'(x)) g_1''(x) \neq h''(x) g_1'(x) \tag{2.79}$$

in this context. Note that this condition depends solely on $x$.

Now let us assume that $\mathtt{A}$ is some small open environment of the equilibrium $(x^*, y^*)$ that we established in Example 2.29. Since for the sake of this toy example we have complete freedom in choosing $g_1, \ldots, g_L$ and $h$, the conditions (2.78) and (2.79) are quite easy to satisfy – just suppose, for example, that in $\mathtt{E}^*$ the functions $g_1, \ldots, g_L$ are locally linear but non-constant and that the curvature $h''$ is locally non-zero.

Together with Examples 2.5, 2.17, and 2.29, this means that the assumptions of Theorem 2.3 are fulfilled for any smooth periodic signal. As in Example 2.53, we have thus established positive Harris recurrence for the stochastic version of this variant of (1.14).

**Example 2.58.** Even though it might seem like the rotor model from Example 1.3 should involve some kind of cascade structure (at least in the setting with one-sided input), calculating the derivatives of the respective function $U$ shows that it actually does not (even after renumbering the variables). If we apply the strategy from (2.74) anyway[9] and do calculations in analogy to Lemma 2.54, we can span $N + L$ space directions in the one-sided case at points where $\partial_{y_3} G_1$ and $\partial_{y_4} G_2$ do not vanish and in the two-sided case at points where $\partial_{y_2} G_1$ and $\partial_{y_4} G_1$ do not vanish. For the time-homogeneous system from [11] which includes no explicit external equations, one can allow those derivatives to vanish as long as some higher order of them does not, as the authors prove in [11, Lemma 5.3]. In our case however, time-inhomogeneity and the extra dependence on $z$ in the drift make this more difficult. Decoupling of $X$ and $Z$ is even more problematic, since any dependence of $U$ on the $x$-variable is linear, which rules out working with a condition like (H4). At the moment, we do not know how to provide a method that will work under assumptions that are neither too restrictive nor physically irrelevant.

[9] For the case of two-sided input, we have $N = 2$ and therefore we have to do the same construction twice: Once just as in (2.74) and once with $V_1$ replaced with $V_2$. For the respective calculations, recall that Lemma 2.41 works for any $N \in \mathbb{N}$.

# Chapter 3

# Estimation of the periodicity and the shape of the deterministic signal

In this chapter, we want to study a statistical model in which the deterministic signal $S$ depends on a set of parameters. More precisely, we assume that there is an open set $\Theta \subset \mathbb{R}^d$ such that

$$S = S_{(\vartheta,T)} \quad \text{with } (\vartheta, T) \in \Theta \times (0, \infty),$$

where $T$ is the signal's periodicity and $\vartheta$ is a $d$-dimensional shape parameter. A natural goal is to estimate $\vartheta$ and $T$ simultaneously from continuous observation of $\mathbb{X}$. However, observing $\mathbb{X}$ entirely may not make sense in many models: The external variable $Z$ can be of a rather abstract nature and, for example, in the Hodgkin-Huxley model the only variable that is arguably observable is the membrane potential $X$.

In spite of that, Section 3.1 shows that from a statistical point of view it does not matter whether we can observe $X$, $Z$ or the entire process $\mathbb{X}$ (Remark 3.1, Proposition 3.3). Since $Z$ is the most convenient process to handle among all of these, our considerations in the sequel are confined to this external variable. Being able to relate statistical problems entirely to $Z$ means that as long as this variable fits our setting, we can treat any example of (SDS) (including in particular those that were introduced in Section 1.2).

In Section 3.2, we prove Local Asymptotic Normality for the sequence of statistical experiments corresponding to continuous observation of $Z$ over growing time intervals $[0, n]$ for $n \to \infty$. The local scales are identified as $n^{-1/2}$ for the shape and $n^{-3/2}$ for the periodicity (Theorem 3.11). Section 3.2 is essentially a generalised and extended version of the article [28] in which we only treated the case $M = N = 1$.

## 3.1 Observing $\mathbb{X}$, $X$, or $Z$

We start this section with a fundamental observation: If the starting point is known, observing only $X$ is actually no restriction, since we can successively reconstruct the remaining variables. Let us explain this step for step in the following remark.

**Remark 3.1.** Assume that the starting point $\mathbb{X}_0 = (x_0, y_0, z_0) \in \mathtt{E}$ is known. Fix a finite time horizon $t_0 \in (0, \infty)$ and assume that the trajectory $(X_t)_{t\in[0,t_0]}$ has been observed and is thus also known. Then the function

$$[0, t_0] \times V \ni (t, y) \mapsto G(X_t, y)$$

is completely known, and given the structure of the internal equation in (SDS), the trajectory $(Y_t)_{t\in[0,t_0]}$ is now given as the solution to the deterministic initial value problem

$$\begin{aligned} dY_t &= G(X_t, Y_t)dt \quad \text{for all } t \in [0, t_0], \\ Y_0 &= y_0. \end{aligned}$$

Now we know both $(X_t)_{t\in[0,t_0]}$ and $(Y_t)_{t\in[0,t_0]}$, and by rearranging the first line of (SDS), this information allows us to calculate

$$Z_t = z_0 + X_t - x_0 - \int_0^t F(X_s, Y_s)ds \quad \text{for all } t \in [0, t_0].$$

All in all, we have reconstructed every component of $(\mathbb{X}_t)_{t\in[0,t_0]}$ just from $(X_t)_{t\in[0,t_0]}$ and the starting point $\mathbb{X}_0$.

Remark 3.1 is the legitimation for us to work with the idealised assumption that we can in fact observe $\mathbb{X}$. Next, we will precisely describe the corresponding statistical experiment.

Incorporating the parameters into our general notation, we rewrite the equation (SDS) for $\mathbb{X} = (X, Y, Z)$ as

$$d\mathbb{X}_t = B_{(\vartheta,T)}(t, \mathbb{X}_t)dt + \Sigma(\mathbb{X}_t)dW_t, \tag{SDS'}$$

where

$$B_{(\vartheta,T)} \colon [0, \infty) \times \mathtt{E} \to \mathbb{R}^{N+L+N}, \quad (t, x, y, z) \mapsto \begin{pmatrix} F(x, y) + S_{(\vartheta,T)}(t) + b(z) \\ G(x, y) \\ S_{(\vartheta,T)}(t) + b(z) \end{pmatrix},$$

for each $(\vartheta, T) \in \Theta \times (0, \infty)$, while the volatility $\Sigma$ remains the same as in (1.4).

We will always assume that for all sets of parameters $(\vartheta, T) \in \Theta \times (0, \infty)$ and all deterministic starting points $\mathbb{X}_0 \in \mathtt{E}$ the equation (SDS') has a unique strong solution $\mathbb{X}^{(\vartheta,T)}$ on $(\Omega, \mathcal{A})$ under $\mathbb{P}$.

In order to conveniently work within the theoretical framework of parameter estimation, we do not want to have a family of processes but rather a family of probability measures. To that end, we introduce the canonical path space

$$\big(C([0,\infty);\mathtt{E}),\mathcal{B}\big(C([0,\infty);\mathtt{E})\big)\big),$$

where $\mathcal{B}\big(C([0,\infty);\mathtt{E})\big)$ denotes the Borel-$\sigma$-field with respect to the topology of locally uniform convergence. On this space, we can define the canonical process $\pi=(\pi_t)_{t\in[0,\infty)}$ by setting

$$\pi_t\colon C([0,\infty);\mathtt{E})\to\mathtt{E},\quad \omega\mapsto\omega(t)\quad \text{for all } t\in[0,\infty).$$

We write

$$\mathbb{P}^{(\vartheta,T)}:=\mathcal{L}\left([0,\infty)\ni t\mapsto\mathbb{X}_t^{(\vartheta,T)}\,\middle|\,\mathbb{P}\right)$$

for the law on $\mathcal{B}\big(C([0,\infty);\mathtt{E})\big)$ of (the trajectory of) the unique strong solution of (SDS') when issued from $\mathbb{X}_0\in\mathtt{E}$ with the parameter $(\vartheta,T)\in\Theta\times(0,\infty)$. Then

$$\mathcal{L}\left(\pi\,\middle|\,\mathbb{P}^{(\vartheta,T)}\right)=\mathcal{L}\left(\mathbb{X}^{(\vartheta,T)}\,\middle|\,\mathbb{P}\right),$$

and hence we have shifted the parameter dependence from the process to the measure.

Observing the process continuously then means working with the filtration given by

$$\mathcal{F}_t:=\bigcap_{r\in(t,\infty)}\sigma(\pi_s\,|\,s\in[0,r])\subset\mathcal{B}\big(C([0,\infty);\mathtt{E})\big)\quad\text{for all } t\in[0,\infty)$$

and gives rise to the sequence of statistical experiments defined by

$$\mathcal{E}_{\mathbb{X}}:=\big(C([0,\infty);\mathtt{E}),\mathcal{F}_n,\big\{\mathbb{P}^{(\vartheta,T)}|_{\mathcal{F}_n}\,\big|\,(\vartheta,T)\in\Theta\times(0,\infty)\big\}\,\big)_{n\in\mathbb{N}}.$$

Our next step will be to calculate the log-likelihood ratios

$$\log\frac{d\mathbb{P}^{(\tilde{\vartheta},\tilde{T})}|_{\mathcal{F}_t}}{d\mathbb{P}^{(\vartheta,T)}|_{\mathcal{F}_t}}\quad\text{with } (\vartheta,T),(\tilde{\vartheta},\tilde{T})\in\Theta\times(0,\infty),\ t\in[0,\infty),$$

for this experiment, as our ultimate goal is to show that it is Locally Asymptotically Normal. Before we proceed, let us briefly recall what that means.

**Definition 3.2.** Let $\Xi\subset\mathbb{R}^D$ be an open set and assume that for each $n\in\mathbb{N}$ the set $\{\mathbb{P}_n^\xi\,|\,\xi\in\Xi\}$ is a family of probability measures on the measurable space $(\Omega_n,\mathcal{A}_n)$. The sequence

$$\big(\Omega_n,\mathcal{A}_n,\{\mathbb{P}_n^\xi\,|\,\xi\in\Xi\}\big)_{n\in\mathbb{N}}$$

of statistical experiments is termed *Locally Asymptotically Normal in* $\xi$, if there is a sequence $\mathcal{S}^\xi=(\mathcal{S}_n^\xi)_{n\in\mathbb{N}}$ of random variables

$$\mathcal{S}_n^\xi\colon(\Omega_n,\mathcal{A}_n)\to\big(\mathbb{R}^D,\mathcal{B}(\mathbb{R}^D)\big),$$

a symmetric and positive definite matrix $\mathcal{I}_\xi \in \mathbb{R}^{D\times D}$, and a sequence of symmetric and positive definite matrices $(\delta_n^\xi)_{n\in\mathbb{N}} \subset \mathbb{R}^{D\times D}$ with $\delta_n^\xi \downarrow 0_{D\times D}$,[1] such that the following two properties hold.

1. For every bounded sequence $(h_n)_{n\in\mathbb{N}} \subset \mathbb{R}^D$ the log-likelihood ratio admits a quadratic expansion

$$\log \frac{d\mathbb{P}_n^{\xi+\delta_n^\xi h_n}}{d\mathbb{P}_n^\xi} = h_n^\top \mathcal{S}_n^\xi - \frac{1}{2} h_n^\top \mathcal{I}_\xi h_n + o_{\mathbb{P}_n^\xi}(1) \tag{3.1}$$

under $\mathbb{P}_n^\xi$. Here, $o_{\mathbb{P}_n^\xi}(1)$ denotes an (arbitrary) sequence of random variables

$$R_n \colon (\Omega_n, \mathcal{A}_n) \to \big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big) \quad \text{for all } n \in \mathbb{N}$$

that vanishes in probability under $(\mathbb{P}_n^\xi)_{n\in\mathbb{N}}$, i.e.

$$\mathbb{P}_n^\xi(|R_n| > \varepsilon) \xrightarrow{n\to\infty} 0 \quad \text{for all } \varepsilon > 0.$$

2. We have weak convergence

$$\mathcal{L}\big(\mathcal{S}_n^\xi \,\big|\, \mathbb{P}_n^\xi\big) \xrightarrow{n\to\infty} \mathcal{N}\big(0_D, \mathcal{I}_\xi\big). \tag{3.2}$$

In this situation, we call $(\delta_n^\xi)_{n\in\mathbb{N}}$ the *local scale*, $\mathcal{S}^\xi$ the *Score*, and $\mathcal{I}_\xi$ the *Fisher Information* in $\xi$.

The concept behind Local Asymptotic Normality is to fix $\xi \in \Xi$ and then reparameterise the experiment at stage $n$ around $\xi$ by looking at $\xi + \delta_n^\xi h$ and letting

$$h \in \Xi_n^\xi := \big\{v \in \mathbb{R}^D \,\big|\, \xi + \delta_n^\xi v \in \Xi\big\}$$

take the role of a local parameter. Since $\delta_n^\xi$ is decreasing, for growing $n \in \mathbb{N}$ the range $\Xi_n^\xi$ of possible values of $h$ will tend to the full space $\mathbb{R}^D$, and thanks to (3.1) and (3.2), we can associate a limit experiment

$$\big(\Omega_\infty^\xi, \mathcal{A}_\infty^\xi, \{\mathbb{P}_\infty^{\xi,h} \,|\, h \in \mathbb{R}^D\}\big)$$

which has

$$\log \frac{d\mathbb{P}_\infty^{\xi,h}}{d\mathbb{P}_\infty^{\xi,0}} = h^\top \mathcal{S}_\infty^\xi - \frac{1}{2} h^\top \mathcal{I}_\xi h \quad \text{for all } h \in \mathbb{R}^D,$$

where

$$\mathcal{L}\big(\mathcal{S}_\infty^\xi \,\big|\, \mathbb{P}_\infty^{\xi,0}\big) = \mathcal{N}\big(0_D, \mathcal{I}_\xi\big).$$

[1] Monotonicity is to be understood in the sense of the partial order induced by setting $A \geq B$ if and only if $A - B$ is non-negative definite.

In other words, the limit experiment has the structure of a Gaussian shift experiment – a well-studied model in which much is known about efficient parameter estimation. The importance of Local Asymptotic Normality lies in the fact that certain statistical properties of a Gaussian shift can be carried over to the pre-asymptotic level in some sense. This concept was introduced by LeCam in 1960 (see [47]) and has since turned out to be very fruitful. The most important results that can be proved following this idea are Hájek's Convolution Theorem and the Local Asymptotic Minimax Theorem which are strong tools in establishing optimality for sequences $(\hat{\xi}_n)_{n\in\mathbb{N}}$ of estimators for the unknown parameter $\xi$, when the rescaled estimation errors are stochastically asymptotically equivalent to the central statistic $\mathcal{Z}_n := \mathcal{I}_\xi^{-1}\mathcal{S}_n^\xi$ of the experiment, i.e.

$$(\delta_n^\xi)^{-1}(\hat{\xi}_n - \xi) = \mathcal{Z}_n + o_{\mathbb{P}_n^\xi}(1) \quad \text{for } n \to \infty.$$

See [48], [12], [45], or [29] for a detailed presentation of the relevant theory.

Having revised our knowledge about Local Asymptotic Normality, we can now return to our train of thought. A closer look at (SDS') reveals that the drift coefficient depends on the parameter $(\vartheta, T) \in \Theta \times (0,\infty)$, while the volatility does not. Hence, we can use [29, Theorem 6.10] in order to determine the log-likelihood ratios. For technical reasons (which will become apparent shortly), we will make the following assumption.

**(A1) Uniform ellipticity of $\sigma\sigma^\top$:** The mapping $\sigma\sigma^\top\colon \mathbb{R}^N \to \mathbb{R}^{N\times N}$ is uniformly elliptic, i.e. there is a $\sigma_0 \in (0,\infty)$ such that

$$x^\top \left(\sigma\sigma^\top(z)\right) x \geq \sigma_0 \, |x|^2 \quad \text{for all } x, z \in \mathbb{R}^N.$$

Uniform ellipticity is of course equivalent to the assertion that at each $z \in \mathbb{R}^N$ the symmetric $N \times N$-matrix $\sigma\sigma^\top(z)$ is positive definite and its smallest eigenvalue is no smaller than $\sigma_0$ which does not depend on $z$. This is in turn equivalent to $\sigma\sigma^\top(z)$ being invertible with $\sigma_0^{-1}$ as an upper bound for the eigenvalues of its symmetric inverse, so in particular

$$x^\top \left(\sigma\sigma^\top(z)\right)^{-1} x \leq \sigma_0^{-1} \, |x|^2 \quad \text{for all } x, z \in \mathbb{R}^N. \tag{3.3}$$

Since $\left(\sigma\sigma^\top(z)\right)^{-1}$ is symmetric and positive definite, it possesses a square root, i.e. there is some symmetric and positive definite $\left(\sigma\sigma^\top(z)\right)^{-1/2} \in \mathbb{R}^{N\times N}$ such that

$$\left(\left(\sigma\sigma^\top(z)\right)^{-1/2}\right)^2 = \left(\sigma\sigma^\top(z)\right)^{-1}.$$

Note that invertibility of $\sigma\sigma^\top(z)$ also implies that $\sigma^\top \left(\sigma\sigma^\top\right)^{-1}(z) \in \mathbb{R}^{M\times N}$ is a right inverse of $\sigma(z)$. Thus, (A1) is an even stronger variant of the non-degeneracy condition (C5) that we used in Section 2.3.

We are now prepared to apply [29, Theorem 6.10]. Let $\xi = (t, x, y, z) \in [0, \infty) \times \mathtt{E}$. Comparing the drift coefficients of (SDS') with different parameters $(\tilde{\vartheta}, \tilde{T}), (\vartheta, T) \in \Theta \times (0, \infty)$, we see that

$$\begin{aligned}
\big(B_{(\tilde{\vartheta},\tilde{T})} - B_{(\vartheta,T)}\big)(\xi) &= \begin{pmatrix} S_{(\tilde{\vartheta},\tilde{T})} - S_{(\vartheta,T)} \\ 0_L \\ S_{(\tilde{\vartheta},\tilde{T})} - S_{(\vartheta,T)} \end{pmatrix}(t) \\
&= \begin{pmatrix} \sigma\sigma^\top(z) & 0_{N\times L} & \sigma\sigma^\top(z) \\ 0_{L\times N} & 0_{L\times L} & 0_{L\times N} \\ \sigma\sigma^\top(z) & 0_{N\times L} & \sigma\sigma^\top(z) \end{pmatrix} \begin{pmatrix} 0_{N+L} \\ \big(\sigma\sigma^\top\big)^{-1}(z)\big(S_{(\tilde{\vartheta},\tilde{T})} - S_{(\vartheta,T)}\big)(t) \end{pmatrix} \\
&= \Sigma\Sigma^\top(x,y,z) \begin{pmatrix} 0_{N+L} \\ \big(\sigma\sigma^\top\big)^{-1}(z)\big(S_{(\tilde{\vartheta},\tilde{T})} - S_{(\vartheta,T)}\big)(t) \end{pmatrix} \\
&=: \Sigma\Sigma^\top(x,y,z)\Gamma(t,x,y,z).
\end{aligned}$$

Defining $\Gamma$ in this way yields

$$\begin{aligned}
\int_0^t \big(\Gamma^\top\Sigma\Sigma^\top\Gamma\big)(s,\pi_s)ds &= \int_0^t \begin{pmatrix} 0_{N+L} \\ \big(\sigma\sigma^\top\big)^{-1}\big(S_{(\tilde{\vartheta},\tilde{T})} - S_{(\vartheta,T)}\big) \end{pmatrix}^\top \begin{pmatrix} S_{(\tilde{\vartheta},\tilde{T})} - S_{(\vartheta,T)} \\ 0_L \\ S_{(\tilde{\vartheta},\tilde{T})} - S_{(\vartheta,T)} \end{pmatrix}(s,\pi_s)ds \\
&= \int_0^t \big(S_{(\tilde{\vartheta},\tilde{T})} - S_{(\vartheta,T)}\big)^\top \big(\sigma\sigma^\top\big)^{-1}\big(S_{(\tilde{\vartheta},\tilde{T})} - S_{(\vartheta,T)}\big)(s,\pi_s)ds \\
&\leq \sigma_0^{-1} \int_0^t \Big|S_{(\tilde{\vartheta},\tilde{T})}(s) - S_{(\vartheta,T)}(s)\Big|^2 ds \\
&< \infty,
\end{aligned}$$

because the signals are continuous. Thence, both conditions (+) and (++) of [29, Theorem 6.10] are fulfilled. Writing $m^{\mathbb{X},(\vartheta,T)}$ for the local martingale part of $\pi$ under $\mathbb{P}^{(\vartheta,T)}$, we can conclude that

$$\log \frac{d\mathbb{P}^{(\tilde{\vartheta},\tilde{T})}|_{\mathcal{F}_t}}{d\mathbb{P}^{(\vartheta,T)}|_{\mathcal{F}_t}} = \int_0^t \Gamma(s,\pi_s)^\top dm_s^{\mathbb{X},(\vartheta,T)} - \frac{1}{2}\int_0^t \left(\Gamma^\top\Sigma\Sigma^\top\Gamma\right)(s,\pi_s)ds.$$

Setting $\pi^Z := \big(\pi^{(N+L+1)}, \dots, \pi^{(N+L+N)}\big)^\top$ and writing $\tilde{m}^{Z,(\vartheta,T)}$ for its local martingale part under $\mathbb{P}^{(\vartheta,T)}$, the expression for the log-likelihood ratio can be rewritten as

$$\begin{aligned}
&\int_0^t \Big(\big(\sigma\sigma^\top(\pi_s^Z)\big)^{-1}\big(S_{(\tilde{\vartheta},\tilde{T})} - S_{(\vartheta,T)}\big)(s)\Big)^\top d\tilde{m}_s^{Z,(\vartheta,T)} \\
&\qquad - \frac{1}{2}\int_0^t \big(S_{(\tilde{\vartheta},\tilde{T})} - S_{(\vartheta,T)}\big)^\top(s)\left(\sigma\sigma^\top(\pi_s^Z)\right)^{-1}\left(S_{(\tilde{\vartheta},\tilde{T})} - S_{(\vartheta,T)}\right)(s)ds.
\end{aligned}$$

In order to eliminate the rather unintuitive integral with respect to $\tilde{m}^{Z,(\vartheta,T)}$, we introduce the local $\big(\mathbb{P}^{(\vartheta,T)}, (\mathcal{F}_t)_{t\in[0,\infty)}\big)$-martingale $\tilde{B}^{(\vartheta,T)} := \big(\tilde{B}_t^{(\vartheta,T)}\big)_{t\in[0,\infty)}$ that is given by

$$\tilde{B}_t^{(\vartheta,T)} = \int_0^t (\sigma\sigma^\top)^{-1/2}(\pi_s^Z)d\tilde{m}_s^{Z,(\vartheta,T)} \quad \text{for all } t \in [0,\infty). \tag{3.4}$$

Its quadratic variation process is

$$\begin{aligned}
\left\langle \int_0^\cdot (\sigma\sigma^\top)^{-1/2}(\pi_s^Z)d\tilde{m}_s^{Z,(\vartheta,T)} \right\rangle_t &= \int_0^t (\sigma\sigma^\top)^{-1/2}(\pi_s^Z)\big((\sigma\sigma^\top)^{-1/2}\big)^\top(\pi_s^Z)d\langle \tilde{m}^{Z,(\vartheta,T)}\rangle_s \\
&= \int_0^t (\sigma\sigma^\top)^{-1}(\pi_s^Z)d\left(\int_0^s \sigma\sigma^\top(\pi_r^Z)dr\right) \\
&= t \cdot 1_{N\times N}
\end{aligned}$$

for all $t \in [0,\infty)$, so Lévy's Characterisation Theorem [40, Theorem II.6.1] yields that $\tilde{B}^{(\vartheta,T)}$ is an $N$-dimensional $\big(\mathbb{P}^{(\vartheta,T)}, (\mathcal{F}_t)_{t\in[0,\infty)}\big)$-Brownian Motion. Incorporating this process, we can write

$$\begin{aligned}
\log \frac{d\mathbb{P}^{(\tilde{\vartheta},\tilde{T})}|_{\mathcal{F}_t}}{d\mathbb{P}^{(\vartheta,T)}|_{\mathcal{F}_t}} &= \int_0^t \Big((\sigma\sigma^\top)^{-1/2}\big(S_{(\tilde{\vartheta},\tilde{T})} - S_{(\vartheta,T)}\big)(s)\Big)^\top d\tilde{B}_s^{(\vartheta,T)} \\
&\quad - \frac{1}{2}\int_0^t \big(S_{(\tilde{\vartheta},\tilde{T})} - S_{(\vartheta,T)}\big)^\top(s)\,\big(\sigma\sigma^\top(\pi_s^Z)\big)^{-1}\,\big(S_{(\tilde{\vartheta},\tilde{T})} - S_{(\vartheta,T)}\big)(s)ds.
\end{aligned} \tag{3.5}$$

We note immediately that the only component of $\pi$ that is featured explicitly in this expression is the $\pi^Z$-component. This is hardly surprising, since by (SDS') it is the only one that directly receives any random influence. It seems plausible that we should get the same expression for the log-likelihood ratio in an experiment that does not even know that any variables other than $Z$ exist. Let us make this formally rigorous.

In analogy to $\pi$, for all $t \in [0,\infty)$ we define the mapping

$$\eta_t\colon C\big([0,\infty);\mathbb{R}^N\big) \to \mathbb{R}^N, \quad \omega \mapsto \omega(t), \tag{3.6}$$

such that $\eta = (\eta_t)_{t\in[0,\infty)}$ is the canonical process on $C\big([0,\infty);\mathbb{R}^N\big)$. We write

$$\mathbb{Q}^{(\vartheta,T)} := \mathcal{L}\left([0,\infty) \ni t \mapsto Z_t^{(\vartheta,T)} \,\Big|\, \mathbb{P}\right)$$

for the law on $\mathcal{B}\big(C\big([0,\infty);\mathbb{R}^N\big)\big)$ of the unique strong solution $Z^{(\vartheta,T)}$ on $(\Omega,\mathcal{A})$ under $\mathbb{P}$ of

$$dZ_t = [S_{(\vartheta,T)}(t) + b(Z_t)]dt + \sigma(Z_t)dW_t, \tag{3.7}$$

when issued from $z_0 \in \mathbb{R}^N$ with the parameter $(\vartheta,T) \in \Theta \times (0,\infty)$. For any $t \in [0,\infty)$ let

$$\mathcal{G}_t := \bigcap_{r\in(t,\infty)} \sigma(\eta_s \,|\, s \in [0,r])$$

and consider the sequence of experiments given by

$$\mathcal{E}_Z := \Big(C\big([0,\infty);\mathbb{R}^N\big), \mathcal{G}_n, \big\{\mathbb{Q}^{(\vartheta,T)}|_{\mathcal{G}_n} \,\big|\, (\vartheta,T) \in \Theta \times (0,\infty)\big\}\Big)_{n\in\mathbb{N}}. \tag{3.8}$$

As above, we want to calculate the log-likelihood ratios corresponding to this sequence. Comparing the drift coefficients of (3.7) with different parameters $(\tilde{\vartheta}, \tilde{T}), (\vartheta, T) \in \Theta \times (0, \infty)$, we see that for all $(t, z) \in [0, \infty) \times \mathbb{R}^N$ we have

$$\begin{aligned}\big(b(z) + S_{(\tilde{\vartheta},\tilde{T})}(t)\big) - \big(b(z) + S_{(\vartheta,T)}(t)\big) &= \big(S_{(\tilde{\vartheta},\tilde{T})} - S_{(\vartheta,T)}\big)(t) \\ &= \sigma\sigma^\top(z)\big(\sigma\sigma^\top\big)^{-1}(z)\big(S_{(\tilde{\vartheta},\tilde{T})} - S_{(\vartheta,T)}\big)(t) \\ &=: \sigma\sigma^\top(z)\gamma(t, z).\end{aligned}$$

Just as above, we obtain

$$\begin{aligned}\int_0^t \left(\gamma^\top\sigma\sigma^\top\gamma\right)(s, \eta_s)ds &= \int_0^t \big(S_{(\tilde{\vartheta},\tilde{T})} - S_{(\vartheta,T)}\big)^\top \big(\sigma\sigma^\top\big)^{-1}\big(S_{(\tilde{\vartheta},\tilde{T})} - S_{(\vartheta,T)}\big)(s, \eta_s)ds \\ &\leq \sigma_0^{-1} \int_0^t \left|S_{(\tilde{\vartheta},\tilde{T})}(s) - S_{(\vartheta,T)}(s)\right|^2 ds \\ &< \infty\end{aligned}$$

for all $t \in (0, \infty)$, again thanks to the continuity of the signals.

Writing $m^{Z,(\vartheta,T)}$ for the local martingale part of $\eta$ under $\mathbb{Q}^{(\vartheta,T)}$, we can again use [29, Theorem 6.10] which yields

$$\begin{aligned}\log \frac{d\mathbb{Q}^{(\tilde{\vartheta},\tilde{T})}|_{\mathcal{G}_t}}{d\mathbb{Q}^{(\vartheta,T)}|_{\mathcal{G}_t}} &= \int_0^t \gamma(s, \eta_s)^\top dm_s^{Z,(\vartheta,T)} - \frac{1}{2}\int_0^t \left(\gamma^\top\sigma\sigma^\top\gamma\right)(s, \eta_s)ds \\ &= \int_0^t \Big((\sigma\sigma^\top)^{-1/2}(\eta_s)\big(S_{(\tilde{\vartheta},\tilde{T})} - S_{(\vartheta,T)}\big)(s)\Big)^\top dB_s^{(\vartheta,T)} \\ &\quad - \frac{1}{2}\int_0^t \big(S_{(\tilde{\vartheta},\tilde{T})} - S_{(\vartheta,T)}\big)^\top(s) \left(\sigma\sigma^\top(\eta_s)\right)^{-1} \big(S_{(\tilde{\vartheta},\tilde{T})} - S_{(\vartheta,T)}\big)(s)ds,\end{aligned} \tag{3.9}$$

where the process $B^{(\vartheta,T)} := \big(B_t^{(\vartheta,T)}\big)_{t\in[0,\infty)}$ that is given by

$$B_t^{(\vartheta,T)} = \int_0^t (\sigma\sigma^\top)^{-1/2}(\eta_s)dm_s^{Z,(\vartheta,T)} \quad \text{for all } t \in [0, \infty) \tag{3.10}$$

is again an $N$-dimensional $\big(\mathbb{Q}^{(\vartheta,T)}, (\mathcal{G}_t)_{t\in[0,\infty)}\big)$-Brownian Motion, which is seen in the same way as for $\tilde{B}^{(\vartheta,T)}$ above.

We now have calculated the log-likelihood ratios for both $\mathcal{E}_{\mathbb{X}}$ and $\mathcal{E}_Z$. Comparing them leads to the following result.

**Proposition 3.3.** *The sequences $\mathcal{E}_{\mathbb{X}}$ and $\mathcal{E}_Z$ corresponding to continuous observation of $\mathbb{X}$ or $Z$ respectively, with the same deterministic starting point $\mathbb{X}_0 = (X_0, Y_0, Z_0) \in \mathtt{E}$, are statistically equivalent in the sense that*

$$\mathcal{L}\left(\left(\log \frac{d\mathbb{Q}^{(\tilde{\vartheta},\tilde{T})}|_{\mathcal{G}_t}}{d\mathbb{Q}^{(\vartheta,T)}|_{\mathcal{G}_t}}\right)_{t\in[0,\infty)} \middle| \, \mathbb{Q}^{(\vartheta,T)}\right) = \mathcal{L}\left(\left(\log \frac{d\mathbb{P}^{(\tilde{\vartheta},\tilde{T})}|_{\mathcal{F}_t}}{d\mathbb{P}^{(\vartheta,T)}|_{\mathcal{F}_t}}\right)_{t\in[0,\infty)} \middle| \, \mathbb{P}^{(\vartheta,T)}\right). \tag{3.11}$$

*In particular, we have Local Asymptotic Normality for $\mathcal{E}_{\mathbb{X}}$ if and only if we have it for $\mathcal{E}_Z$ with the same local scale, Fisher Information and Score (in distribution).*

*Proof.* Due to the definition of $\mathbb{P}^{(\vartheta,T)}$ and $\mathbb{Q}^{(\vartheta,T)}$, we have

$$\mathcal{L}\big(\eta \,\big|\, \mathbb{Q}^{(\vartheta,T)}\big) = \mathcal{L}\big(\pi^Z \,\big|\, \mathbb{P}^{(\vartheta,T)}\big),$$

and in view of (3.4), (3.5), (3.9), and (3.10), this implies (3.11) from which the second statement of this Proposition follows immediately. □

**Remark 3.4.** Proposition 3.3 is the justification for us to restrict ourselves to studying the simpler process $Z$ instead of the more complex $\mathbb{X}$. Let us stress that Proposition 3.3 basically lets us apply Theorem 3.11 (which we state and prove in the following section) to $\mathbb{X}$ just the same as to $Z$. In particular, this Theorem is of practical relevance for all of the examples that were introduced in Section 1.2, including the stochastic Hodgkin-Huxley system (Example 1.1) and also the rotor model from Example 1.3. If we can treat the external noise, we can treat the entire system (provided we know its starting configuration).

## 3.2 Local Asymptotic Normality for $Z$

This section centres around the sequence of statistical experiments defined by $\mathcal{E}_Z$ in (3.8) which corresponds to continuous observation over growing time intervals of the $N$-dimensional diffusion $Z$ following the stochastic differential equation

$$dZ_t = [S_{(\vartheta,T)}(t) + b(Z_t)]dt + \sigma(Z_t)dW_t, \tag{3.12}$$

where $\vartheta \in \Theta$ is an unknown $d$-dimensional shape parameter of the signal, and $T \in (0, \infty)$ is its unknown periodicity. Just as we previously did for the entire process $\mathbb{X}$, we always make the following basic assumption.

**(A2) Unique solvability:** For each $(\vartheta, T) \in \Theta \times (0, \infty)$, the equation (3.12) has a unique strong solution.

As mentioned in Section 1.1, taking $M = N = 1$, $b \equiv 0$, and $\sigma \equiv 1$ leads to the classical "signal in white noise" model. For this special case, Ibragimov and Khasminskii proved Local Asymptotic Normality with rate $n^{-3/2}$ for a smooth signal with known $\vartheta$ and discussed asymptotic efficiency for certain estimators (see [39, Sections II.7 and III.5]). In [21], Golubev extended their approach with $\mathbb{L}^2$-methods in order to estimate $T$ at the same rate for unknown shape, which in turn was the basis for Castillo, Lévy-Leduc and Matias for non-parametric estimation of the shape under unknown $T$ (see [7]). For our more general diffusion (3.12), we will stay within the confines of parametric estimation.

Our main result is Local Asymptotic Normality for the sequence of experiments $\mathcal{E}_Z$ with unknown $\vartheta$ and unknown $T$ (Theorem 3.11). For $M = N = 1$ Höpfner and

Kutoyants had already solved this problem both for known $T$ with unknown $\vartheta$ (see [31]) and for known $\vartheta$ with unknown $T$ (see [33]). Our result extends both of these and allows for application to simultaneous estimation of the shape and the periodicity.

Recall from (3.9) that for all $(\vartheta, T), (\tilde{\vartheta}, \tilde{T}) \in \Theta \times (0, \infty)$ and $t \in [0, \infty)$ the log-likelihood ratio for this sequence of experiments is given by

$$\begin{aligned}\Lambda_t^{(\tilde{\vartheta},\tilde{T})/(\vartheta,T)} &:= \log\left(\frac{d\mathbb{Q}^{(\tilde{\vartheta},\tilde{T})}|_{\mathcal{G}_t}}{d\mathbb{Q}^{(\vartheta,T)}|_{\mathcal{G}_t}}\right)\\ &= \int_0^t \Big((\sigma\sigma^\top)^{-1/2}(\eta_s)\big(S_{(\tilde{\vartheta},\tilde{T})} - S_{(\vartheta,T)}\big)(s)\Big)^\top dB_s^{(\vartheta,T)}\\ &\quad - \frac{1}{2}\int_0^t \big(S_{(\tilde{\vartheta},\tilde{T})} - S_{(\vartheta,T)}\big)^\top(s)\,\big(\sigma\sigma^\top(z)\big)^{-1}\,\big(S_{(\tilde{\vartheta},\tilde{T})} - S_{(\vartheta,T)}\big)(s)ds,\end{aligned}$$

where $B^{(\vartheta,T)}$ is the Brownian Motion defined in (3.10). Striving for Local Asymptotic Normality, we need to find a suitable quadratic expansion of $\Lambda_t^{(\tilde{\vartheta},\tilde{T})/(\vartheta,T)}$ in the sense of (3.1). Examining its structure suggests that we have to impose appropriate smoothness conditions on the signal with respect to the parameters. The following set of conditions (S1) - (S4) turns out to be sufficient:

**(S1) Periodicity and basic regularity:** For each $\vartheta \in \Theta$ we have a 1-periodic function

$$S_\vartheta = \begin{pmatrix} S_\vartheta^{(1)} \\ \vdots \\ S_\vartheta^{(N)} \end{pmatrix} \in C^2\big([0,\infty);\mathbb{R}^N\big)$$

such that

$$S_\cdot(s) \in C^1\big(\Theta;\mathbb{R}^N\big) \quad \text{for every } s \in [0,\infty)$$

and

$$\partial_{\vartheta_i} S_\vartheta(\cdot) \in \mathbb{L}^2_{\text{loc}}\big([0,\infty);\mathbb{R}^N\big) \quad \text{for every } \vartheta \in \Theta \text{ and } i \in \{1,\dots,d\}.$$

**(S2) $\mathbb{L}^2_{\text{loc}}$-differentiability with respect to $(\vartheta, T)$:** The mapping

$$\begin{aligned} S\colon \Theta \times (0,\infty) &\to \mathbb{L}^2_{\text{loc}}\big([0,\infty);\mathbb{R}^N\big),\\ (\vartheta,T) \quad &\mapsto S_{(\vartheta,T)} := S_\vartheta\left(\frac{\cdot}{T}\right),\end{aligned}$$

is $\mathbb{L}^2_{\text{loc}}$-differentiable with the derivative

$$\begin{aligned} \dot{S}\colon \Theta \times (0,\infty) &\to \mathbb{L}^2_{\text{loc}}\big([0,\infty);\mathbb{R}^{N\times(d+1)}\big),\\ (\vartheta,T) \quad &\mapsto \dot{S}_{(\vartheta,T)} := \begin{pmatrix} \partial_{\vartheta_1} S^{(1)}_{(\vartheta,T)} & \cdots & \partial_{\vartheta_d} S^{(1)}_{(\vartheta,T)} & \partial_T S^{(1)}_{(\vartheta,T)} \\ \vdots & & \vdots & \vdots \\ \partial_{\vartheta_1} S^{(N)}_{(\vartheta,T)} & \cdots & \partial_{\vartheta_d} S^{(N)}_{(\vartheta,T)} & \partial_T S^{(N)}_{(\vartheta,T)} \end{pmatrix},\end{aligned}$$

in the sense that for every $t \in (0,\infty)$ and $(\vartheta,T) \in \Theta \times (0,\infty)$ we have

$$\int_0^t \left| \frac{S_{(\tilde\vartheta,\tilde T)}(s) - S_{(\vartheta,T)}(s) - \dot S_{(\vartheta,T)}(s)\big((\tilde\vartheta,\tilde T) - (\vartheta,T)\big)}{\left|(\tilde\vartheta,\tilde T) - (\vartheta,T)\right|} \right|^2 ds \to 0, \text{ as } (\tilde\vartheta,\tilde T) \to (\vartheta,T).$$

**(S3)** $\mathbb{L}^2_{\mathrm{loc}}$**-continuity of the** $(\vartheta,T)$**-derivative:** The mapping $\dot S$ is $\mathbb{L}^2_{\mathrm{loc}}$-continuous in the sense that for all $t \in (0,\infty)$ and $(\vartheta,T) \in \Theta \times (0,\infty)$ we have

$$\int_0^t \left| \dot S_{(\tilde\vartheta,\tilde T)}(s) - \dot S_{(\vartheta,T)}(s) \right|^2 ds \to 0, \text{ as } (\tilde\vartheta,\tilde T) \to (\vartheta,T),$$

where the notation $|\cdot|$ is used for the Frobenius norm of a matrix.

**(S4)** $\mathbb{L}^2_{\mathrm{loc}}$**-Hölder condition with respect to** $T$ **for the** $\vartheta$**-derivative:** For any fixed $\vartheta \in \Theta$ the mapping

$$(0,\infty) \ni T \mapsto D_\vartheta S_{(\vartheta,T)} := \begin{pmatrix} \partial_{\vartheta_1} S^{(1)}_{(\vartheta,T)} & \cdots & \partial_{\vartheta_d} S^{(1)}_{(\vartheta,T)} \\ \vdots & & \vdots \\ \partial_{\vartheta_1} S^{(N)}_{(\vartheta,T)} & \cdots & \partial_{\vartheta_d} S^{(N)}_{(\vartheta,T)} \end{pmatrix} \in \mathbb{L}^2_{\mathrm{loc}}\big([0,\infty);\mathbb{R}^{N\times d}\big)$$

satisfies the following local Hölder condition: For each $T \in (0,\infty)$ there are

$$\alpha \in (0,2] \quad \text{and} \quad \beta \in [0, 1 + 3\alpha/2)$$

such that for suitable $\varepsilon > 0$ and $t_0 \in [0,\infty)$ we have

$$\int_{t_0}^t \left| D_\vartheta S_{(\vartheta,\tilde T)}(s) - D_\vartheta S_{(\vartheta,T)}(s) \right|^2 ds \le C t^\beta \left| \tilde T - T \right|^\alpha$$

for all $t > t_0$, $\tilde T \in (T-\varepsilon, T+\varepsilon)$, and for some constant $C \in (0,\infty)$ that does not depend on $\tilde T$ or $t$.

**Remark 3.5.** 1.) We observe that if (S1) holds and $\dot S_{(\vartheta,T)}(s)$ is continuous (and thus also locally bounded) with respect to $\vartheta$, $T$, and $s$, (S2) and (S3) are immediate by dominated convergence. Note that in general, (S1) does not require that for example $\partial_{\vartheta_1} S_{(\vartheta,T)}(s)$ is continuous (or even locally bounded) in $T$ or $s$.

2.) Suppose that (S1) holds. If for every $(\vartheta,T) \in \Theta \times (0,\infty)$ and $t \in (0,\infty)$ there are $\delta = \delta(\vartheta,T) \in (0,1]$ and $C(\vartheta,t) \le \mathrm{cst}\, t^\zeta$ with $\zeta \in [0,\delta/2)$ such that the mapping

$$[0,\infty) \ni s \mapsto D_\vartheta S_\vartheta(s) := \begin{pmatrix} \partial_{\vartheta_1} S^{(1)}_\vartheta(s) & \cdots & \partial_{\vartheta_d} S^{(1)}_\vartheta(s) \\ \vdots & & \vdots \\ \partial_{\vartheta_1} S^{(N)}_\vartheta(s) & \cdots & \partial_{\vartheta_d} S^{(N)}_\vartheta(s) \end{pmatrix} \in \mathbb{R}^{N\times d}$$

is Hölder-$\delta$-continuous on $[0,t]$ with Hölder-constant $C(\vartheta,t)$, we get that for sufficiently small $\varepsilon > 0$ and for all $\tilde{T} \in (T-\varepsilon, T+\varepsilon)$

$$\begin{aligned}\int_0^t \left| D_\vartheta S_{(\vartheta,\tilde{T})}(s) - D_\vartheta S_{(\vartheta,T)}(s) \right|^2 ds &= \int_0^t \left| D_\vartheta S_\vartheta \left( \frac{s}{\tilde{T}} \right) - D_\vartheta S_\vartheta \left( \frac{s}{T} \right) \right|^2 ds \\ &\leq \sup_{T' \in (T-\varepsilon, T+\varepsilon)} C\left(\vartheta, \frac{t}{T'}\right)^2 \int_0^t \left| \frac{s}{\tilde{T}} - \frac{s}{T} \right|^{2\delta} ds \\ &\leq \mathrm{cst} \left( \frac{t}{T-\varepsilon} \right)^{2\zeta} \left( \frac{\left|\tilde{T} - T\right|}{(T-\varepsilon)^2} \right)^{2\delta} \int_0^t s^{2\delta} ds \\ &\leq \mathrm{cst}\, t^{2\zeta+2\delta+1} \left|\tilde{T} - T\right|^{2\delta}.\end{aligned}$$

Setting $\alpha := 2\delta$, we can choose

$$\beta := 2(\delta + \zeta) + 1 < 2\left(\delta + \frac{\delta}{2}\right) + 1 = 1 + 3\alpha/2,$$

and hence the Hölder condition (S4) is fulfilled.

3.) As a consequence of the two preceding observations, all of the hypotheses (S1) - (S4) are fulfilled if the mapping $\Theta \times [0,\infty) \ni (\vartheta, s) \mapsto S_\vartheta(s)$ is in $C_b^2\big(\Theta \times [0,\infty); \mathbb{R}^N\big)$ and 1-periodic with respect to $s$. Existence and boundedness of $\partial_s D_\vartheta S_\vartheta(s)$ ensure that we can choose $\delta = 1$ and $\zeta = 0$ above.

4.) Note that the choice of the matrix norm in (S3) and (S4) is of course arbitrary. We decided to go with the Frobenius norm, because it is commonly used and it is convenient to handle in our calculations.

**Example 3.6.** 1.) Let $S_\vartheta(s) = f(\vartheta, \varphi(s))$, where $\varphi \in C^2\big([0,\infty); \mathbb{R}^D\big)$ is 1-periodic and

$$f \colon \Theta \times \mathbb{R}^D \ni (x,y) = (x_1, \ldots, x_d, y_1, \ldots, y_D) \mapsto f(x,y) = \begin{pmatrix} f_1(x,y) \\ \vdots \\ f_N(x,y) \end{pmatrix} \in \mathbb{R}^N$$

is continuously differentiable with respect to $x \in \Theta$ and twice continuously differentiable with respect to $y \in \mathbb{R}^D$. Clearly, the property (S1) holds, and since $\dot{S}_{(\vartheta,T)}(s)$ is given by

$$\begin{pmatrix} (\partial_{x_1} f_1)(\vartheta, \varphi(\frac{s}{T})) & \cdots & (\partial_{x_d} f_1)(\vartheta, \varphi(\frac{s}{T})) & -sT^{-2}(\nabla_y f_1)(\vartheta, \varphi(\frac{s}{T}))^\top \varphi'(\frac{s}{T}) \\ \vdots & & \vdots & \vdots \\ (\partial_{x_1} f_N)(\vartheta, \varphi(\frac{s}{T})) & \cdots & (\partial_{x_d} f_N)(\vartheta, \varphi(\frac{s}{T})) & -sT^{-2}(\nabla_y f_N)(\vartheta, \varphi(\frac{s}{T}))^\top \varphi'(\frac{s}{T}) \end{pmatrix}$$

which is continuous with respect to $\vartheta$, $T$, and $s$, we also have (S2) and (S3). Moreover, we see that the Hölder property from part 2.) of Remark 3.5 is fulfilled if it is fulfilled by the mapping

$$\mathbb{R}^D \ni y \mapsto \begin{pmatrix} (\partial_{x_1} f_1)(\vartheta, y) & \cdots & (\partial_{x_d} f_1)(\vartheta, y) \\ \vdots & & \vdots \\ (\partial_{x_1} f_N)(\vartheta, y) & \cdots & (\partial_{x_d} f_N)(\vartheta, y) \end{pmatrix}.$$

In that case, all of the hypotheses (S1) - (S4) hold.

2.) A special case of the preceding example is a product structure $S_\vartheta(s) = g(\vartheta)\varphi(s)$ with $\varphi \in C^2\big([0,\infty);\mathbb{R}^D\big)$ 1-periodic and $g \in C^1\big(\Theta;\mathbb{R}^{N\times D}\big)$. As for all $s, \tilde{s} \in [0,\infty)$ we have

$$\begin{aligned} |D_\vartheta S_\vartheta(s) - D_\vartheta S_\vartheta(\tilde{s})|^2 &= \sum_{i=1}^N \sum_{j=1}^d \left( \sum_{k=1}^D (\partial_{\vartheta_j} g_{i,k})(\vartheta)\big(\varphi_k(s) - \varphi_k(\tilde{s})\big) \right)^2 \\ &\leq \left( \sum_{i=1}^N \sum_{j=1}^d \sum_{k=1}^D (\partial_{\vartheta_j} g_{i,k})^2(\vartheta) \right) |\varphi(s) - \varphi(\tilde{s})|^2 \\ &\leq \left( \sum_{i=1}^N \sum_{j=1}^d \sum_{k=1}^D (\partial_{\vartheta_j} g_{i,k})^2(\vartheta) \right) \|\varphi'\|_\infty^2 \, |s - \tilde{s}|^2 , \end{aligned}$$

no further conditions are needed to ensure the Hölder property from part 2.) of Remark 3.5 to hold with $\delta = 1$ and $\zeta = 0$.

3.) In particular, the example above includes signals of the form

$$S_\vartheta(s) = \sum_{k=1}^l \big( \sin(2k\pi s) g_k(\vartheta) + \cos(2k\pi s) h_k(\vartheta) \big) \quad \text{for all } s \in [0,\infty)$$

with $l \in \mathbb{N}_0$ and $g_k, h_k \in C^1\big(\Theta;\mathbb{R}^N\big)$ for all $k \in \{1,\dots,l\}$.

As in Chapter 2, the fact that $S_{(\vartheta,T)}$ and therefore the entire drift term of (3.12) is $T$-periodic can be exploited in order to use properties of time-homogeneous substructures of the process $\eta$. The grid chain $\eta^{\mathbf{gr}}$ on $\mathbb{R}^N$ and the path segment chain $\eta^{\mathbf{ps}}$ on $C\big([0,T];\mathbb{R}^N\big)$ can be defined in the same way as we defined them for $\mathbb{X}$ in Lemma 2.1. Next to regularity of the signal with respect to the parameters in the sense of (S1) - (S4), the following recurrence condition is the second fundamental assumption in this section.

- **(A3) Positive Harris recurrence:** For all $(\vartheta, T) \in \Theta \times (0,\infty)$ the grid chain $\eta^{\mathbf{gr}}$ under $\mathbb{Q}^{(\vartheta,T)}$ is positive Harris recurrent with invariant probability measure $\mu^{(\vartheta,T)}$.

Verifiable criteria for this condition can be found for example in [34], a specific example will be given in Example 3.15 at the end of this section. Note that (A3) is weaker than the conclusion of Theorem 2.3.

As we know from [32, Theorem 2.1 (a)][2], the path segment chain $\eta^{\mathbf{ps}}$ inherits positive Harris recurrence under $\mathbb{Q}^{(\vartheta,T)}$ from the grid chain and its invariant distribution $m^{(\vartheta,T)}$ is the unique measure on $\mathcal{B}\big(C([0,T];\mathbb{R}^N)\big)$ such that for all $l \in \mathbb{N}$, $0 = t_0 < t_1 < \ldots < t_l = T$, and $B_0, \ldots, B_l \in \mathcal{B}\big(\mathbb{R}^N\big)$ we have

$$\begin{aligned} &m^{(\vartheta,T)}(\eta_{t_i} \in B_i \text{ for all } i \in \{0,\ldots,l\}) \\ &\qquad = \int_{B_0} \mu^{(\vartheta,T)}(dx_0) \int_{B_1} Q^{(\vartheta,T)}_{t_0,t_1}(x_0, dx_1) \ldots \int_{B_l} Q^{(\vartheta,T)}_{t_{l-1},t_l}(x_{l-1}, dx_l), \end{aligned} \tag{3.13}$$

where $\big(Q^{(\vartheta,T)}_{s,t}\big)_{t>s\geq 0}$ is the transition semi-group of $\eta$ under $\mathbb{Q}^{(\vartheta,T)}$ which is defined in analogy to (2.1).

In order to prove our main result (Theorem 3.11 below), we want to use limiting properties of $\eta$ arising from positive Harris recurrence. The foundation for this is laid by the following strong law of large numbers, which we cite from [32, Theorem 2.1 (b)].

**Proposition 3.7.** *Let (A2) and (A3) hold and fix some $(\vartheta, T) \in \Theta \times (0,\infty)$. Assume that $(A_t)_{t\in[0,\infty)}$ is a $\big(\mathbb{Q}^{(\vartheta,T)}, (\mathcal{G}_t)_{t\in[0,\infty)}\big)$-increasing process. If there is a non-negative function $f \in \mathbb{L}^1\big(m^{(\vartheta,T)}\big)$ such that*

$$A_{kT} = \sum_{j=1}^{k} f\left(\eta^{\mathbf{ps}}_j\right) \quad \mathbb{Q}^{(\vartheta,T)}\text{-almost surely for all } k \in \mathbb{N},$$

*then*

$$\frac{1}{t} A_t \xrightarrow{t\to\infty} \frac{1}{T} \int_{C([0,T];\mathbb{R}^N)} f(\varphi) m^{(\vartheta,T)}(d\varphi) \quad \mathbb{Q}^{(\vartheta,T)}\text{-almost surely.}$$

*Proof.* See Section 2 of [32]. □

Proposition 3.7 is the key to the following Lemma 3.8 which is a slightly modified multi-dimensional version of Lemmas 2.1 and 2.2 from [33].

**Lemma 3.8.** *Grant assumptions (A2) and (A3). Further assume that the measurable mapping $g\colon \mathbb{R}^N \to \mathbb{R}^{N\times N}$ has values only in the set of symmetric matrices and is uniformly elliptic. We define the mapping*

$$\begin{aligned} \mathbb{B}^{(\vartheta,T)}_g \colon \big(\mathbb{L}^2\big([0,1];\mathbb{R}^N\big)\big)^2 &\to \mathbb{R}, \\ (u,v) &\mapsto \int_0^1 u(s)^\top \Big(\mu^{(\vartheta,T)} Q^{(\vartheta,T)}_{0,sT}(g^{-1})\Big) v(s) ds, \end{aligned} \tag{3.14}$$

*where*

$$\mu^{(\vartheta,T)} Q^{(\vartheta,T)}_{0,sT}(g^{-1}) = \int_{\mathbb{R}^N} \mu^{(\vartheta,T)}(dz) \int_{\mathbb{R}^N} Q^{(\vartheta,T)}_{0,sT}(z, d\tilde{z}) g^{-1}(\tilde{z}) \in \mathbb{R}^{N\times N}$$

*is understood as a matrix-valued integral. Then the following statements are true.*

[2]Note that even though this Theorem is only explicitly stated for $\mathbb{R}$-valued processes, the authors remark at the beginning of the section that it remains valid for any polish state space, in particular $\mathbb{R}^N$.

*(i)* $\mathbb{B}_g^{(\vartheta,T)}$ *is a non-negative definite and symmetric bilinear form.*

*(ii) If we consider* $u, v \in \mathbb{L}^2\big([0,1];\mathbb{R}^N\big)$ *as 1-periodic functions on* $(0,\infty)$*, then for any* $k \in \mathbb{N}_0$ *we have*

$$\frac{k+1}{t^{k+1}} \int_0^t s^k u(s/T)^\top g^{-1}(\eta_s) v(s/T) ds \xrightarrow{t\to\infty} \mathbb{B}_g^{(\vartheta,T)}[u,v] \tag{3.15}$$

$\mathbb{Q}^{(\vartheta,T)}$*-almost surely.*

*Proof.* For the sake of simplicity and as $(\vartheta,T)$ is fixed anyway, we drop all corresponding superscripts. First, we check that $\mathbb{B}_g$ is indeed a well-defined mapping with values in $\mathbb{R}$. Let the lower bound for the eigenvalues of $g(\cdot)$ be denoted by $g_0 \in (0,\infty)$. Recall that $g^{-1}(\cdot)$ always exists, is positive definite, and $g_0^{-1}$ is an upper bound for its eigenvalues. Then by linearity and contractivity of the operator $\mu Q_{0,sT}$, we can estimate

$$0 \le \mathbb{B}_g[u,u] = \int_0^1 \mu Q_{0,sT}\left(u(s)^\top g^{-1}(\cdot) u(s)\right) ds \le g_0^{-1} \int_0^1 |u(s)|^2 \, ds < \infty.$$

Thanks to the symmetry of $g^{-1}$, we can polarise the integrand and thus the whole expression, which allows us to use the above in order to conclude that

$$|\mathbb{B}_g[u,v]| = \frac{1}{2} \left|\mathbb{B}_g[u,u] + \mathbb{B}_g[v,v] - \mathbb{B}_g[u+v,u+v]\right| < \infty,$$

and hence $\mathbb{B}_g$ is well-defined. It is then trivial to see that it is a non-negative definite and symmetric bilinear form, and the proof for (i) is complete.

We note that the left hand side of (3.15) is bilinear in $u$ and $v$ as well. Thanks to this and (i), the proof of the second statement of the Lemma can be reduced to the case $u = v$, since the general case then follows by polarisation.

Let us define the process $A := (A_t)_{t\in[0,\infty)}$ with

$$A_t := \int_0^t u(s/T)^\top g^{-1}(\eta_s) u(s/T) ds \quad \text{for all } t \in [0,\infty).$$

Since $g^{-1}(\cdot)$ is positive definite, the integrand is non-negative, and therefore $A$ is an increasing process whose trajectories are obviously continuous. Note that the expression on the left hand side of (3.15) can be rewritten as

$$\frac{k+1}{t^{k+1}} \int_0^t s^k dA_s.$$

For $k = 0$ this is simply $\frac{1}{t}A_t$, which we will handle with Proposition 3.7. The general statement then follows from this case by taking $f(t) = (\mathbb{B}_g[u,u])^{-1} A_t$ in Lemma 3.17 (found at the end of this chapter).

In order to establish the functional relation between $A$ and $\eta$ that is needed in Proposition 3.7, we define the function

$$f\colon C\big([0,T];\mathbb{R}^N\big) \to [0,\infty), \quad \varphi \mapsto \int_0^T u(s/T)^\top g^{-1}(\varphi(s))u(s/T)ds,$$

which is bounded by $Tg_0^{-1}\,\|u\|_{\mathbb{L}^2([0,1])}$, and thus it is integrable with respect to the probability measure $m$. Due to the periodicity of $u$, we see that

$$\begin{aligned}\sum_{j=1}^k f\left(\eta_j^{\mathbf{ps}}\right) &= \sum_{j=1}^k \int_0^T u(s/T)^\top g^{-1}(\eta_{(j-1)T+s})u(s/T)ds \\ &= \int_0^{kT} u(s/T)^\top g^{-1}(\eta_s)u(s/T)ds \\ &= A_{kT}\end{aligned}$$

for all $k \in \mathbb{N}$, and consequently Proposition 3.7 allows to deduce $\mathbb{Q}$-almost sure convergence

$$\begin{aligned}\lim_{t\to\infty} \frac{1}{t}A_t &= \frac{1}{T}\int_{C([0,T];\mathbb{R}^N)} f(\varphi)m(d\varphi) \\ &= \frac{1}{T}\int_{C([0,T];\mathbb{R}^N)} \int_0^T u(s/T)^\top g^{-1}(\varphi(s))u(s/T)ds\, m(d\varphi) \\ &= \frac{1}{T}\int_0^T u(s/T)^\top \int_{C([0,T];\mathbb{R}^N)} g^{-1}(\varphi(s))m(d\varphi)u(s/T)ds \\ &= \frac{1}{T}\int_0^T u(s/T)^\top \int_{\mathbb{R}^N} g^{-1}(x)\mu Q_{0,s}(dx)u(s/T)ds \\ &= \mathbb{B}_g[u,u],\end{aligned}$$

where the use of Fubini's Theorem in the third step is justified by the non-negativity of the integrand, and the fourth step makes use of (3.13). This completes the proof. □

For each $(\vartheta,T) \in \Theta\times(0,\infty)$ and $t \in [0,\infty)$ we define the symmetric $(d+1)\times(d+1)$-dimensional block matrix

$$\mathcal{I}_{(\vartheta,T)}(t) := \begin{pmatrix} t\left(\mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}[\partial_{\vartheta_i}S_\vartheta,\partial_{\vartheta_j}S_\vartheta]\right)_{i,j=1,\dots,d} & -\frac{t^2}{2T^2}\left(\mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}[\partial_{\vartheta_i}S_\vartheta,S'_\vartheta]\right)_{i=1,\dots,d} \\ \cdots & \frac{t^3}{3T^4}\mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}[S'_\vartheta,S'_\vartheta]\end{pmatrix} \tag{3.16}$$

and also look at its derivative with respect to $t$,

$$\mathcal{I}'_{(\vartheta,T)}(t) = \begin{pmatrix} \left(\mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}[\partial_{\vartheta_i}S_\vartheta,\partial_{\vartheta_j}S_\vartheta]\right)_{i,j=1,\dots,d} & -tT^{-2}\left(\mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}[\partial_{\vartheta_i}S_\vartheta,S'_\vartheta]\right)_{i=1,\dots,d} \\ \cdots & t^2T^{-4}\mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}[S'_\vartheta,S'_\vartheta]\end{pmatrix}.$$

We make the following assumption.

**(S5) Regularity of the signal with respect to $\mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}$:** For all $(\vartheta, T) \in \Theta \times (0,\infty)$ and $t \in (0,\infty)$ we have

$$\text{(i) } \mathcal{I}_{(\vartheta,T)}(t) \text{ is invertible,} \qquad \text{(ii) } \mathcal{I}'_{(\vartheta,T)}(t) \text{ is invertible.}$$

Since $\mathcal{I}_{(\vartheta,T)}(t)$ and $\mathcal{I}'_{(\vartheta,T)}(t)$ are symmetric real matrices, in this context being invertible means the same thing as being positive definite. While part (ii) of (S5) is merely needed for technical reasons (as will become clear in the proof of Theorem 3.11 below), part (i) is of more general importance, since $\mathcal{I}_{(\vartheta,T)}(1)$ will turn out to be the Fisher Information. We will discuss these conditions in detail in the following remark.

**Remark 3.9.** 1.) A simple sufficient condition for (S5) is orthogonality of the functions $\partial_{\vartheta_1} S_\vartheta, \ldots, \partial_{\vartheta_d} S_\vartheta, S'_\vartheta$ with respect to $\mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}$, which is equivalent to both $\mathcal{I}_{(\vartheta,T)}(t)$ and $\mathcal{I}'_{(\vartheta,T)}(t)$ being diagonal matrices with non-vanishing diagonal entries. As such they are invertible. However, this is not a very likely scenario, since $S_\vartheta$ has $d$ degrees of freedom, determines the $d$ functions $\partial_{\vartheta_1} S_\vartheta, \ldots, \partial_{\vartheta_d} S_\vartheta$, and then $S'_\vartheta$ – while adding no further degree of freedom – would have to be orthogonal to these as well.

2.) Without orthogonality, the situation becomes slightly more delicate. If $\mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}$ is positive definite (and hence an inner product), part (ii) of condition (S5) is equivalent to the assertion that $\partial_{\vartheta_1} S_\vartheta, \ldots, \partial_{\vartheta_d} S_\vartheta, S'_\vartheta$ are linearly independent, since $\mathcal{I}'_{(\vartheta,T)}(t)$ is the Gramian matrix of $\partial_{\vartheta_1} S_\vartheta, \ldots, \partial_{\vartheta_d} S_\vartheta, -tT^{-2} S'_\vartheta$ with respect to $\mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}$, and the factor $-tT^{-2}$ is irrelevant for linear independence. As we will prove now, part (i) is then fulfilled as well, as $\mathcal{I}_{(\vartheta,T)}(t)$ is "almost a Gramian matrix". Indeed, setting

$$u_1 := t^{\frac{1}{2}} \partial_{\vartheta_1} S_\vartheta,\ \ldots,\ u_d := t^{\frac{1}{2}} \partial_{\vartheta_d} S_\vartheta,\ u_{d+1} := -\frac{t^{\frac{3}{2}}}{2T^2} S'_\vartheta,$$

we can write

$$\mathcal{I}_{(\vartheta,T)}(t) = \begin{pmatrix} \mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}[u_1,u_1] & \cdots & \cdots & \mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}[u_1,u_{d+1}] \\ \vdots & \ddots & & \vdots \\ \vdots & & \mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}[u_d,u_d] & \mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}[u_d,u_{d+1}] \\ \mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}[u_{d+1},u_1] & \cdots & \mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}[u_{d+1},u_d] & \frac{4}{3}\mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}[u_{d+1},u_{d+1}] \end{pmatrix},$$

and we see that for all $x \in \mathbb{R}^{(d+1)\times(d+1)}$

$$\begin{aligned} x^\top \mathcal{I}_{(\vartheta,T)}(t) x &= \sum_{i,j=1}^{d+1} x_i \mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}[u_i,u_j] x_j + \frac{1}{3} x_{d+1}^2 \mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}[u_{d+1},u_{d+1}] \\ &\geq \mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}\left[\sum_{i=1}^{d+1} x_i u_i, \sum_{j=1}^{d+1} x_j u_j\right]. \end{aligned}$$

Linear independence of $u_1, \ldots, u_{d+1}$ (which follows directly from linear independence of $\partial_{\vartheta_1} S_\vartheta, \ldots, \partial_{\vartheta_d} S_\vartheta, S'_\vartheta$) and positive definiteness of $\mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}$ imply that this expression is strictly positive unless $x$ is zero. In other words, $\mathcal{I}_{(\vartheta,T)}(t)$ is positive definite and hence invertible.

3.) A simple sufficient condition for $\mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}$ to be positive definite is uniform ellipticity of $\left(\sigma\sigma^\top\right)^{-1}$, i.e. the existence of some $\tilde{\sigma}_0 \in (0, \infty)$ such that

$$x^\top \left(\sigma\sigma^\top(z)\right)^{-1} x \geq \tilde{\sigma}_0 \, |x|^2 \quad \text{for all } x, z \in \mathbb{R}^N.$$

If this is the case, for all $u \in \mathbb{L}^2\big([0,1]; \mathbb{R}^N\big)$ we can estimate

$$\mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}[u,u] = \int_0^1 \mu^{(\vartheta,T)} Q^{(\vartheta,T)}_{0,sT} \left(u(s)^\top \left(\sigma\sigma^\top\right)^{-1}(\cdot) u(s)\right) ds \geq \tilde{\sigma}_0 \int_0^1 |u(s)|^2 \, ds,$$

i.e. $\mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}$ is even coercive and hence positive definite.

4.) In conclusion, the property

**(S5') Regularity of the signal with respect to $\mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}$:** The mapping $\left(\sigma\sigma^\top\right)^{-1}$ is uniformly elliptic, and the functions $\partial_{\vartheta_1} S_\vartheta, \ldots, \partial_{\vartheta_d} S_\vartheta, S'_\vartheta$ are linearly independent for all $(\vartheta, T) \in \Theta \times (0, \infty)$.

is a relatively simple and potentially verifiable sufficient condition for (S5).

**Example 3.10.** 1.) If the signal is of the form

$$S_\vartheta = \sum_{k=1}^d \vartheta_k \varphi_k, \tag{3.17}$$

where $\varphi_1, \ldots, \varphi_d \in \mathbb{L}^2\big([0,\infty); \mathbb{R}^N\big)$ are 1-periodic and orthonormal with respect to $\mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}$, we have

$$\mathcal{I}_{(\vartheta,T)}(t) = \begin{pmatrix} t \cdot 1_{d\times d} & -\frac{t^2}{2T^2} \left(\sum_{j=1}^d \vartheta_j \mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}[\varphi_i, \varphi'_j]\right)_{i=1,\ldots,d} \\ \cdots & \frac{t^3}{3T^4} \sum_{i,j=1}^d \vartheta_i \vartheta_j \mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}[\varphi'_i, \varphi'_j] \end{pmatrix}$$

which is invertible for all $t \in [0, \infty)$ whenever

$$\frac{4}{3} \sum_{i,j=1}^d \vartheta_i \vartheta_j \mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}[\varphi'_i, \varphi'_j] \neq \sum_{i=1}^d \left(\sum_{j=1}^d \vartheta_j \mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}[\varphi_i, \varphi'_j]\right)^2. \tag{3.18}$$

Similarly,

$$\mathcal{I}'_{(\vartheta,T)}(t) = \begin{pmatrix} 1_{d\times d} & -tT^{-2} \left(\sum_{j=1}^d \vartheta_j \mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}[\varphi_i, \varphi'_j]\right)_{i=1,\ldots,d} \\ \cdots & t^2 T^{-4} \sum_{i,j=1}^d \vartheta_i \vartheta_j \mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}[\varphi'_i, \varphi'_j] \end{pmatrix}$$

is invertible for all $t \in [0,\infty)$ whenever

$$\sum_{i,j=1}^{d} \vartheta_i \vartheta_j \mathbb{B}_{\sigma\sigma^\top}^{(\vartheta,T)}[\varphi_i', \varphi_j'] \neq \sum_{i=1}^{d} \left( \sum_{j=1}^{d} \vartheta_j \mathbb{B}_{\sigma\sigma^\top}^{(\vartheta,T)}[\varphi_i, \varphi_j'] \right)^2 . \tag{3.19}$$

2.) For $M = N$ let $\sigma \equiv 1_{N\times N}$, then $\mathbb{B}_{\sigma\sigma^\top}^{(\vartheta,T)}$ is just the standard $\mathbb{L}^2$-inner product with respect to Lebesgue's measure. If $N = 1$, $d$ is even, and the signal has a finite Fourier expansion

$$S_\vartheta(s) = \sum_{k=1}^{\frac{d}{2}} \sqrt{2} \left( \vartheta_k \sin(2k\pi s) + \vartheta_{\frac{d}{2}+k} \cos(2k\pi s) \right) \quad \text{for all } s \in [0,\infty),$$

it is both of the type from the first part of this example and of the type introduced in part 3.) of Example 3.6 (so in particular it satisfies (S1) - (S4)). Elementary calculations show that the conditions (3.18) and (3.19) then become

$$\sum_{k=1}^{\frac{d}{2}} k(\vartheta_k^2 + \vartheta_{k+\frac{d}{2}}^2) \neq \alpha \sum_{k=1}^{\frac{d}{2}} k^2 \vartheta_{k+\frac{d}{2}}^2 \quad \text{for } \alpha \in \{3,4\}.$$

If for example there are no cos-terms involved, i.e. $\vartheta_{\frac{d}{2}+1} = \ldots = \vartheta_d = 0$, these inequalities are valid for all $(\vartheta_1, \ldots, \vartheta_{\frac{d}{2}}) \neq 0_{\frac{d}{2}}$.

Having introduced all relevant notions and assumptions, we can now give the main result of this chapter.

**Theorem 3.11** (Local Asymptotic Normality)**.** *Grant all of the hypotheses (A1) - (A3) and (S1) - (S5) and fix* $(\vartheta, T) \in \Theta \times (0,\infty)$. *Set*

$$\delta_n := \begin{pmatrix} n^{-1/2} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & n^{-1/2} & 0 \\ 0 & \cdots & 0 & n^{-3/2} \end{pmatrix} \in \mathbb{R}^{(d+1)\times(d+1)} \quad \textit{for all } n \in \mathbb{N},$$

*and fix any bounded sequence* $(h_n)_{n\in\mathbb{N}} \subset \mathbb{R}^{d+1}$. *Then* $\mathbb{Q}^{(\vartheta,T)}$*-almost surely we have*

$$\Lambda_n^{(\vartheta,T)+\delta_n h_n/(\vartheta,T)} = h_n^\top \mathcal{S}_n^{(\vartheta,T)} - \frac{1}{2} h_n^\top \mathcal{I}_{(\vartheta,T)} h_n + o_{\mathbb{Q}^{(\vartheta,T)}}(1) \tag{3.20}$$

*with Fisher Information* $\mathcal{I}_{(\vartheta,T)} = \mathcal{I}_{(\vartheta,T)}(1)$ *as introduced in* (3.16) *and score*

$$\mathcal{S}_n^{(\vartheta,T)} = \delta_n \int_0^n \left( (\sigma\sigma^\top)^{-1/2}(\eta_s) \dot{S}_{(\vartheta,T)}(s) \right)^\top dB_s^{(\vartheta,T)} \quad \textit{for all } n \in \mathbb{N} \tag{3.21}$$

*such that weak convergence*

$$\mathcal{L}\left( \mathcal{S}_n^{(\vartheta,T)} \middle| \mathbb{Q}^{(\vartheta,T)} \right) \xrightarrow{n\to\infty} \mathcal{N}\left( 0_{d+1}, \mathcal{I}_{(\vartheta,T)} \right) \tag{3.22}$$

*holds.*

**Remark 3.12.** 1.) Note that the local scale $(\delta_n)_{n\in\mathbb{N}}$ in Theorem 3.11 is the same for every set of parameters $(\vartheta, T) \in \Theta \times (0, \infty)$.

2.) Let us briefly comment on the comparison between equations (3.1) and (3.2) in the general Definition 3.2 of Local Asymptotic Normality on the one hand and equations (3.20) and (3.22) in Theorem 3.11 on the other hand. The log-likelihood ratio is adapted to $(\mathcal{G}_n)_{n\in\mathbb{N}}$ by its very definition, and by (3.21) so is the Score. Since each $\mathbb{Q}_n^{(\vartheta,T)}$ is simply the restriction of $\mathbb{Q}^{(\vartheta,T)}$ to $\mathcal{G}_n$, there is no need to carry the $n$ in (3.20) and (3.22).

*Proof of Theorem 3.11.* We fix $(\vartheta, T) \in \Theta \times (0, \infty)$, and in order to reduce notational complexity we drop corresponding indices whenever there is no risk of ambiguity: We write $\mathbb{Q} := \mathbb{Q}^{(\vartheta,T)}$, $B := B^{(\vartheta,T)}$ (see (3.10)), $\mathcal{S}_n := \mathcal{S}_n^{(\vartheta,T)}$, $\mathcal{I} := \mathcal{I}_{(\vartheta,T)}$, $\mathcal{I}(t) := \mathcal{I}_{(\vartheta,T)}(t)$ for all $t \in [0, \infty)$ (see (3.16)), and $\mathbb{B} := \mathbb{B}_{\sigma\sigma^\top}^{(\vartheta,T)}$ (see (3.14)). Moreover, we set

$$(\vartheta_n, T_n) := (\vartheta, T) + \delta_n h_n \quad \text{for all } n \in \mathbb{N}.$$

We now proceed to give the proof, divided into several steps.

1.) The main idea is to introduce a time step size $t \in (0, \infty)$ into the log-likelihood ratio and then interpret

$$\left(\Lambda_{tn}^{(\vartheta_n,T_n)/(\vartheta,T)}\right)_{t\in[0,\infty)}, \quad n \in \mathbb{N},$$

as a sequence of continuous-time stochastic processes. Splitting them into several parts and applying Lemma 3.8 together with tools from continuous-time martingale theory will eventually lead to the desired quadratic expansion. Indeed, adding and subtracting the term $\dot{S}_{(\vartheta,T)}(s)\delta_n h_n$ to the difference of the signals yields

$$\begin{aligned}
\Lambda_{tn}^{(\vartheta_n,T_n)/(\vartheta,T)} &= \int_0^{tn} \big((\sigma\sigma^\top)^{-1/2}(\eta_s)\big(S_{(\vartheta_n,T_n)} - S_{(\vartheta,T)}\big)(s)\big)^\top dB_s \\
&\quad - \frac{1}{2}\int_0^{tn} \big(S_{(\vartheta_n,T_n)} - S_{(\vartheta,T)}\big)^\top(s)\big(\sigma\sigma^\top(z)\big)^{-1}\big(S_{(\vartheta_n,T_n)} - S_{(\vartheta,T)}\big)(s)ds \\
&= h_n^\top \left(\delta_n \int_0^{tn} \big((\sigma\sigma^\top)^{-1/2}(\eta_s)\dot{S}_{(\vartheta,T)}(s)\big)^\top dB_s\right) \\
&\quad - \frac{1}{2}h_n^\top \left(\delta_n \int_0^{tn} \dot{S}_{(\vartheta,T)}(s)^\top\big(\sigma\sigma^\top(\eta_s)\big)^{-1}\dot{S}_{(\vartheta,T)}(s)ds\,\delta_n\right) h_n \\
&\quad + \int_0^{tn} \Big((\sigma\sigma^\top)^{-1/2}(\eta_s)\big(S_{(\vartheta_n,T_n)} - S_{(\vartheta,T)} - \dot{S}_{(\vartheta,T)}\delta_n h_n\big)(s)\Big)^\top dB_s \\
&\quad - \frac{1}{2}\int_0^{tn} \big(S_{(\vartheta_n,T_n)} - S_{(\vartheta,T)} - \dot{S}_{(\vartheta,T)}\delta_n h_n\big)^\top(s)\big(\sigma\sigma^\top(\eta_s)\big)^{-1} \\
&\qquad\qquad \big(S_{(\vartheta_n,T_n)} - S_{(\vartheta,T)} - \dot{S}_{(\vartheta,T)}\delta_n h_n\big)(s)ds \\
&\quad - \int_0^{tn} \big(S_{(\vartheta_n,T_n)} - S_{(\vartheta,T)} - \dot{S}_{(\vartheta,T)}\delta_n h_n\big)^\top(s)\big(\sigma\sigma^\top(\eta_s)\big)^{-1}\big(\dot{S}_{(\vartheta,T)}\delta_n h_n\big)ds
\end{aligned}$$

$$=: h_n^\top \mathcal{S}_n(t) - \frac{1}{2} h_n^\top \mathcal{I}_n(t) h_n + R_n(t) - \frac{1}{2} U_n(t) - V_n(t),$$

and in order to prove the Theorem, we will study convergence in distribution of $\mathcal{S}_n(t)$ for $n \to \infty$ and show almost sure convergence of $\mathcal{I}_n(1)$ to $\mathcal{I} = \mathcal{I}(1)$. Finally, we show that $R_n(t)$, $U_n(t)$, and $V_n(t)$ converge to zero in probability.

2.) For any fixed $n \in \mathbb{N}$ the process

$$M_n := (\mathcal{S}_n(t))_{t\in[0,\infty)} = \Big(\delta_n \int_0^{tn} \big((\sigma\sigma^\top)^{-1/2}(\eta_s)\dot{S}_{(\vartheta,T)}(s)\big)^\top dB_s\Big)_{t\in[0,\infty)}$$

is obviously an $\mathbb{R}^{d+1}$-valued local martingale with respect to $\mathbb{Q}$. In order to determine its weak limit for $n \to \infty$ in the Skorohod space $\mathcal{D}\big([0,\infty);\mathbb{R}^{d+1}\big)$, we study its quadratic variation process $\langle M_n\rangle := (\langle M_n\rangle_t)_{t\in[0,\infty)}$ with

$$\langle M_n\rangle_t := \begin{pmatrix} \Big\langle M_n^{(1)}, M_n^{(1)}\Big\rangle_t & \cdots & \Big\langle M_n^{(1)}, M_n^{(d+1)}\Big\rangle_t \\ \vdots & \ddots & \vdots \\ \Big\langle M_n^{(d+1)}, M_n^{(1)}\Big\rangle_t & \cdots & \Big\langle M_n^{(d+1)}, M_n^{(d+1)}\Big\rangle_t \end{pmatrix} \in \mathbb{R}^{(d+1)\times(d+1)}.$$

As follows from basic stochastic calculus, $\langle M_n\rangle$ is equal to $(\mathcal{I}_n(t))_{t\in[0,\infty)}$. Consequently, for $i, j \in \{1,\dots,d\}$ we have

$$\begin{aligned}\big\langle M_n^{(i)}, M_n^{(j)}\big\rangle_t =& \frac{1}{n}\int_0^{tn} \big(\partial_{\vartheta_i} S_{(\vartheta,T)}(s)\big)^\top \big(\sigma\sigma^\top(\eta_s)\big)^{-1} \partial_{\vartheta_j} S_{(\vartheta,T)}(s)ds \\ =& t\cdot\frac{1}{tn}\int_0^{tn} \big(\partial_{\vartheta_i} S_\vartheta(s/T)\big)^\top \big(\sigma\sigma^\top(\eta_s)\big)^{-1} \partial_{\vartheta_j} S_\vartheta(s/T)ds,\end{aligned}$$

and due to the periodicity of $S_\vartheta$ and by part (ii) of Lemma 3.8 with $g = \sigma\sigma^\top$ and $k = 0$, this expression converges to

$$t \cdot \mathbb{B}[\partial_{\vartheta_i} S_\vartheta, \partial_{\vartheta_j} S_\vartheta] = \mathcal{I}_{i,j}(t)$$

$\mathbb{Q}$-almost surely for $n \to \infty$. Since

$$\partial_T S_{(\vartheta,T)}(s) = \partial_T S_\vartheta(s/T) = -sT^{-2} S'_\vartheta(s/T) \quad \text{for all } s \in (0,\infty),$$

the same argument with $k = 1$ yields

$$\begin{aligned}\big\langle M_n^{(i)}, M_n^{(d+1)}\big\rangle_t &= \big\langle M_n^{(d+1)}, M_n^{(i)}\big\rangle_t \\ &= \frac{1}{n^2}\int_0^{tn} \big(\partial_{\vartheta_i} S_{(\vartheta,T)}(s)\big)^\top \big(\sigma\sigma^\top(\eta_s)\big)^{-1} \partial_T S_{(\vartheta,T)}(s)ds \\ &= \frac{-t^2}{2T^2}\cdot\frac{1}{\frac{1}{2}(tn)^2}\int_0^{tn} s\cdot\big(\partial_{\vartheta_i} S_\vartheta(s/T)\big)^\top \big(\sigma\sigma^\top(\eta_s)\big)^{-1} S'_\vartheta(s/T)ds \\ &\xrightarrow{n\to\infty} \frac{-t^2}{2T^2}\cdot\mathbb{B}[\partial_{\vartheta_i} S_\vartheta, S'_\vartheta] = \mathcal{I}_{i,d+1}(t) = \mathcal{I}_{d+1,i}(t)\end{aligned}$$

$\mathbb{Q}$-almost surely, and analogously (with $k = 2$)

$$\begin{aligned}
\left\langle M_n^{(d+1)}, M_n^{(d+1)} \right\rangle_t &= \frac{1}{n^3} \int_0^{tn} \left(\partial_T S_{(\vartheta,T)}(s)\right)^\top \left(\sigma\sigma^\top(\eta_s)\right)^{-1} \partial_T S_{(\vartheta,T)}(s) ds \\
&= \frac{t^3}{3T^4} \cdot \frac{1}{\frac{1}{3}(tn)^3} \int_0^{tn} s^2 \cdot S_\vartheta'(s/T)^\top \left(\sigma\sigma^\top(\eta_s)\right)^{-1} S_\vartheta'(s/T) ds \\
&\xrightarrow{n\to\infty} \frac{t^3}{3T^4} \cdot \mathbb{B}[S_\vartheta', S_\vartheta'] = \mathcal{I}_{d+1,d+1}(t)
\end{aligned}$$

$\mathbb{Q}$-almost surely. In other words,

$$\langle M_n \rangle_t \xrightarrow{n\to\infty} \mathcal{I}(t) \quad \mathbb{Q}\text{-almost surely for all } t \in [0, \infty),$$

and hence the Martingale Convergence Theorem [42, Corollary VIII.3.24] implies weak convergence

$$\mathcal{L}(M_n|\mathbb{Q}) \xrightarrow{n\to\infty} \mathcal{L}(M|\mathbb{Q}) \quad \text{in } \mathcal{D}\big([0, \infty); \mathbb{R}^{d+1}\big) \tag{3.23}$$

to some limit martingale $M = (M(t))_{t\in[0,\infty)}$.[3] Since $\mathcal{I}'(t)$ is symmetric and non-negative definite, it possesses a square root $\sqrt{\mathcal{I}'(t)} \in \mathbb{R}^{(d+1)\times(d+1)}$. By (S5), $\mathcal{I}'(t)$ is invertible and hence

$$0 \neq \det \mathcal{I}'(t) = \det \sqrt{\mathcal{I}'(t)}\sqrt{\mathcal{I}'(t)}^\top = \left(\det \sqrt{\mathcal{I}'(t)}\right)^2,$$

which is why $\sqrt{\mathcal{I}'(t)}$ is invertible as well. Thus, the Representation Theorem [40, Theorem II.7.1] yields that $M$ can be expressed as

$$M(t) = \int_0^t \sqrt{\mathcal{I}'(s)} dB_s' \quad \text{for all } t \in [0, \infty)$$

with some $(d + 1)$-dimensional Brownian Motion $B'$. Together with (3.23), this also implies weak convergence

$$\mathcal{L}(M_n(t)|\mathbb{Q}) \xrightarrow{n\to\infty} \mathcal{L}(M(t)|\mathbb{Q}) = \mathcal{N}\left(0_{d+1}, \int_0^t \mathcal{I}'(s) ds\right) = \mathcal{N}\left(0_{d+1}, \mathcal{I}(t)\right)$$

for all $t \in [0, \infty)$. In particular, choosing $t = 1$ yields weak convergence of the score

$$\mathcal{L}(\mathcal{S}_n|\mathbb{Q}) = \mathcal{L}(M_n(1)|\mathbb{Q}) \xrightarrow{n\to\infty} \mathcal{N}(0_{d+1}, \mathcal{I}(1)) = \mathcal{N}(0_{d+1}, \mathcal{I}),$$

which completes this step of the proof.

3.) In the second step, we have shown on the fly that

$$\mathcal{I}_n(1) = \langle M_n \rangle_1 \xrightarrow{n\to\infty} \langle M \rangle_1 = \mathcal{I}(1)$$

[3] To be exact, $M$ is actually defined on some arbitrary probability space, but in order to avoid making things more complicated than necessary, we assume without loss of generality that $M$ is in fact defined on the same probability space as the sequence $(M_n)_{n\in\mathbb{N}}$.

$\mathbb{Q}$-almost surely.

4.) It remains to show convergence to zero in $\mathbb{Q}$-probability of the remainder terms $R_n(t)$, $U_n(t)$, and $V_n(t)$ introduced at the very beginning of this proof. Therefore, we consider the sequence $(R_n)_{n\in\mathbb{N}}$ of the local $\mathbb{Q}$-martingales

$$(R_n(t))_{t\in[0,\infty)} = \left(\int_0^{tn} \left((\sigma\sigma^\top)^{-1/2}(\eta_s)\big(S_{(\vartheta_n,T_n)} - S_{(\vartheta,T)} - \dot{S}_{(\vartheta,T)}\delta_n h_n\big)(s)\right)^\top dB_s\right)_{t\in[0,\infty)}.$$

Their quadratic variation processes are obviously given by $(U_n(t))_{t\in[0,\infty)}$. Exploiting the uniform ellipticity assumption (A1), we can use (3.3) to estimate the quadratic variation by

$$\begin{aligned}
\langle R_n\rangle_t &= \int_0^{tn} \big(S_{(\vartheta_n,T_n)} - S_{(\vartheta,T)} - \dot{S}_{(\vartheta,T)}\delta_n h_n\big)^\top(s)\big(\sigma\sigma^\top(\eta_s)\big)^{-1} \\
&\qquad\qquad\qquad\qquad\qquad\qquad \big(S_{(\vartheta_n,T_n)} - S_{(\vartheta,T)} - \dot{S}_{(\vartheta,T)}\delta_n h_n\big)(s)ds \\
&\leq \sigma_0^{-1}\int_0^{tn} \left|S_{(\vartheta_n,T_n)} - S_{(\vartheta,T)} - \dot{S}_{(\vartheta,T)}\delta_n h_n\right|^2 ds \\
&= \sigma_0^{-1}\int_0^{tn} \left|S_{(\vartheta_n,T_n)} - S_{(\vartheta,T)} - D_\vartheta S_{(\vartheta,T)}(\vartheta_n - \vartheta) - \partial_T S_{(\vartheta,T)}(s)(T_n - T)\right|^2 ds.
\end{aligned} \tag{3.24}$$

Note that this upper bound is entirely deterministic. In order to prove that it in fact converges to zero, we will separate the dependence on the parameters $\vartheta$ and $T$ in such a way that we can use the periodicity and (S1) - (S4) efficiently. This can be achieved by continuing the inequality (3.24) with

$$\begin{aligned}
\langle R_n\rangle_t \leq 3\sigma_0^{-1}\bigg( &\int_0^{tn} \left|S_{(\vartheta_n,T_n)}(s) - S_{(\vartheta,T_n)}(s) - D_\vartheta S_{(\vartheta,T_n)}(s)(\vartheta_n - \vartheta)\right|^2 ds \\
&+ \int_0^{tn} \left|\big(D_\vartheta S_{(\vartheta,T_n)} - D_\vartheta S_{(\vartheta,T)}(s)\big)(\vartheta_n - \vartheta)\right|^2 ds \\
&+ \int_0^{tn} \left|S_{(\vartheta,T_n)}(s) - S_{(\vartheta,T)}(s) - \partial_T S_{(\vartheta,T)}(s)(T_n - T)\right|^2 ds\bigg) \\
=:\ & 3\sigma_0^{-1}(A_n + B_n + C_n).
\end{aligned}$$

We will treat convergence of $A_n$, $B_n$, and $C_n$ step for step. For this purpose, set $H := \sup_{n\in\mathbb{N}}|h_n|$ and note that for all $n\in\mathbb{N}$ we have

$$|\vartheta_n - \vartheta| \leq Hn^{-1/2} \quad \text{and} \quad |T_n - T| \leq Hn^{-3/2},$$

since $(\vartheta_n, T_n) - (\vartheta, T) = \delta_n h_n$.

Starting with $A_n$, we observe that for sufficiently large $n\in\mathbb{N}$ we have $T_n \in [T/2, 2T]$

and thus

$$
\begin{aligned}
A_n &\le \left(\frac{tn}{T_n}+1\right)\int_0^{T_n} \left|S_{(\vartheta_n,T_n)}(s)-S_{(\vartheta,T_n)}(s)-D_\vartheta S_{(\vartheta,T_n)}(s)(\vartheta_n-\vartheta)\right|^2 ds \\
&= \left(\frac{tn}{T_n}+1\right)|\vartheta_n-\vartheta|^2\int_0^{T_n} \left|\frac{S_{(\vartheta_n,T_n)}(s)-S_{(\vartheta,T_n)}(s)-D_\vartheta S_{(\vartheta,T_n)}(s)(\vartheta_n-\vartheta)}{|\vartheta_n-\vartheta|}\right|^2 ds \\
&\le \left(\frac{tn}{T/2}+1\right)H^2 n^{-1}\int_0^{2T} \left|\frac{S_{(\vartheta_n,T_n)}(s)-S_{(\vartheta,T_n)}(s)-D_\vartheta S_{(\vartheta,T_n)}(s)(\vartheta_n-\vartheta)}{|\vartheta_n-\vartheta|}\right|^2 ds,
\end{aligned}
$$

where the factor in front of the integral is obviously convergent, and the integral itself tends to zero because of the $\mathbb{L}^2$-continuity condition (S3) and Lemma 3.18 (to be found at the end of this section).

Next, using the Hölder condition (S4), we obtain for sufficiently large $n \in \mathbb{N}$ that

$$
\begin{aligned}
B_n &\le |\vartheta_n-\vartheta|^2\int_0^{tn}\left|D_\vartheta S_{(\vartheta,T_n)}(s)-D_\vartheta S_{(\vartheta,T)}(s)\right|^2 ds \\
&\le H^2 n^{-1}\left(\int_0^{t_0}\left|D_\vartheta S_{(\vartheta,T_n)}(s)-D_\vartheta S_{(\vartheta,T)}(s)\right|^2 ds + C(tn)^\beta\,|T_n-T|^\alpha\right) \\
&\le H^2 n^{-1}\int_0^{t_0}\left|\dot{S}_{(\vartheta,T_n)}(s)-\dot{S}_{(\vartheta,T)}(s)\right|^2 ds + CH^{2+\alpha}t^\beta n^{\beta-(1+3\alpha/2)}.
\end{aligned}
$$

The particular conditions on $\alpha$ and $\beta$ from (S4) make the second summand vanish for $n \to \infty$, while the first summand converges to zero because of (S3).

In order to estimate $C_n$, we make explicit use of the $C^2$-property (S1) which is readily translated into the condition that the mapping

$$
(0,\infty) \ni T \mapsto S_{(\vartheta,T)}(s)
$$

is twice continuously differentiable for any fixed $s \in (0,\infty)$. Consequently, for every $s \in (0,\infty)$ and any $i \in \{1,\dots,N\}$ Taylor expansion with the Lagrange form of the remainder provides a $\varrho_i = \varrho_i(s,\vartheta,T,T_n,h_n)$ between $T$ and $T_n$ such that for sufficiently large $n \in \mathbb{N}$ we can infer that

$$
\begin{aligned}
&\left|S_{(\vartheta,T_n)}(s)-S_{(\vartheta,T)}(s)-(T_n-T)\partial_T S_{(\vartheta,T)}(s)\right|^2 \\
&= \sum_{i=1}^N\left(S^{(i)}_{(\vartheta,T_n)}(s)-S^{(i)}_{(\vartheta,T)}(s)-(T_n-T)\partial_T S^{(i)}_{(\vartheta,T)}(s)\right)^2 \\
&= \sum_{i=1}^N\left(\frac{1}{2}(T_n-T)^2\partial_T^2 S^{(i)}_{(\vartheta,T)}(s)_{|T=\varrho_i}\right)^2 \\
&= \frac{1}{4}(T_n-T)^4\sum_{i=1}^N\left(\frac{s^2}{\varrho_i^4}\left(S^{(i)}_\vartheta\right)''(s/\varrho_i)+\frac{2s}{\varrho_i^3}\left(S^{(i)}_\vartheta\right)'(s/\varrho_i)\right)^2 \\
&\le \frac{1}{4}H^4 n^{-6}2N\left[\left(s^2\frac{\|S''_\vartheta\|_\infty}{(T-n^{-3/2}H)^4}\right)^2+\left(s\frac{2\,\|S'_\vartheta\|_\infty}{(T-n^{-3/2}H)^3}\right)^2\right] \\
&\le \mathrm{cst}\, n^{-6}(s^4+s^2)
\end{aligned}
$$

for some positive constant not depending on $s$ or $n$. Integrating yields

$$C_n \leq \text{cst}\, n^{-6} \int_0^{tn} (s^4 + s^2) ds \xrightarrow{n\to\infty} 0.$$

So far, we have shown that the sequence of random variables $(U_n(t))_{n\in\mathbb{N}}$ not only vanishes in probability under $\mathbb{Q}$ for $n \to \infty$, but is even bounded by a deterministic sequence which goes to zero. Therefore,

$$\mathbb{E}_{\mathbb{Q}}[R_n(t)^2] = \mathbb{E}_{\mathbb{Q}}[\langle R_n\rangle_t] = \mathbb{E}_{\mathbb{Q}}[U_n(t)] \xrightarrow{n\to\infty} 0, \tag{3.25}$$

and in particular, $R_n(t)$ also vanishes in probability under $\mathbb{Q}$ for $n \to \infty$. Finally, the same is true for the last remainder variable $V_n(t)$, as by the Cauchy-Schwarz inequality we get that

$$|V_n(t)|^2 \leq U_n(t) h_n^\top \mathcal{I}_n(t) h_n \leq U_n(t) H^2 \left|\mathcal{I}_n(t)\right| \xrightarrow{n\to\infty} 0, \tag{3.26}$$

since $\mathcal{I}_n(t)$ converges and $U_n(t)$ goes to zero. Taking $t = 1$ completes the proof. □

**Remark 3.13.** The convergence in probability for $n \to \infty$ of the remainder terms $R_n(t)$, $U_n(t)$, and $V_n(t)$ (which determine the term $o_{\mathbb{Q}^{(\vartheta,T)}}(1)$ in (3.20)) is in fact even uniform with respect to $t \in [0, t_0]$ for every $t_0 \in (0, \infty)$. For $U_n(t)$ this is clear, since it only increases with $t$. Using the Burkholder-Davis-Gundy inequality, the estimation (3.25) can be improved to

$$\mathbb{E}_{\mathbb{Q}}\Big[\sup_{t\in[0,t_0]} |R_n(t)|^2\Big] \leq 4\mathbb{E}_{\mathbb{Q}}[\langle R_n\rangle_{t_0}] = 4\mathbb{E}_{\mathbb{Q}}[U_n(t_0)] \xrightarrow{n\to\infty} 0,$$

which also takes care of $R_n(t)$. For $V_n(t)$ we notice that the bound given in (3.26) only depends on $t$ via $\mathcal{I}_n(t)$ and $U_n(t)$ which are both non-decreasing with respect to $t$.

**Remark 3.14.** In the one-dimensional case $M = N = 1$, variants of Theorem 3.11 are already known in the literature, where shape and periodicity are treated separately and one of them is assumed to be known.

If the shape parameter $\vartheta$ is known, Theorem 3.11 includes [33, Theorem 1.1] as a special case (only (A1) - (A3) and the $C^2$-property from (S1) are actually needed in this situation). If, on the other hand, the periodicity is known and the only parameter of interest is $\vartheta$, then Theorem 3.11 leads to the same conclusion as [31, Theorem 2.1] (note that other than in Theorem 3.11, there Score and Fisher Information are written at a time scale given by multiples of the known periodicity $T$). In [31], the $\mathbb{L}^2$-smoothness conditions on the signal are formulated not with respect to Lebesgue's measure $ds$ as in (S2) and (S3) but with respect to the measure

$$\nu^\vartheta(ds) := \mu^\vartheta P^\vartheta_{0,sT}(\sigma^{-2}) ds$$

(where we dropped the superscript $T$, as the periodicity is known). Under the uniform ellipticity condition (A1), these $\mathbb{L}^2$-smoothness assumptions are slightly weaker than (S2) and (S3). However, if (A1) holds (which in practice is more or less the only verifiable condition for $\nu^\vartheta$ to be finite, as is supposed in [31] anyway) the most obvious way to verify these is using that $\nu^\vartheta$ is thus bounded from above by a multiple of Lebesgue's measure. In this sense, the difference between these assumptions is just of a very theoretical nature.

The key to bringing these results together in Theorem 3.11 is the Hölder condition (S4) which is crucial for dealing with the term $B_n$ in step 3.) of its proof. This is the only instant where (in contrast to the terms $A_n$ and $C_n$) we have to impose more than just "joint smoothness", but a more specific relation of the interplay between $T$ and $\vartheta$. It should also be noted that (A1) is essential for this step, as it removes any randomness from the terms we effectively deal with. Otherwise, even if we would reformulate (S2) - (S4) with $\mathbb{L}^2$-convergence replaced by convergence with respect to the semi-norm induced by $\mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}$, we could not treat this term with the methods used in Lemma 3.8 due to the occurrence of different periodicities in the integrand.

**Example 3.15.** For $M = N = 1$ consider the case

$$b(z) = -\gamma z \quad \text{for some } \gamma \in (0,\infty), \qquad \sigma \equiv 1,$$

i.e. $Z$ is a mean reverting Ornstein-Uhlenbeck process with mean reversion speed $\gamma$ and time-dependent mean reversion level $\gamma^{-1}S_{(\vartheta,T)}(t)$. By [32, Example 2.3], the periodic recurrence assumption (A3) is fulfilled, and we see that $\mathbb{B}^{(\vartheta,T)}_{\sigma\sigma^\top}$ is simply the standard $\mathbb{L}^2$-inner product with respect to Lebesgue's measure. In [13], the authors think of $\gamma$ as another unknown parameter, while they assume the periodicity $T$ to be fixed and known. In order to apply our result, we suppose that both $\gamma$ and $T$ are fixed and known, while $\vartheta$ is to be estimated. The signal the authors consider is then the one introduced in (3.17) in Example 3.10. In this setting, we see that the Fisher Information $\mathcal{I}$ is just the identity matrix $1_{d\times d}$ and the Score is given by

$$\mathcal{S}_n = n^{-1/2}\left(\int_0^n \varphi_i(s)dB_s\right)_{i=1,\dots,d} \qquad \text{for all } n \in \mathbb{N}.$$

Proposition 4.1 of [13] implies that the rescaled estimation error $\sqrt{n}(\hat{\vartheta}_n - \vartheta)$ of the maximum likelihood estimator $\hat{\vartheta}_n$ is exactly the central statistic $\mathcal{Z}_n = \mathcal{I}^{-1}\mathcal{S}_n = \mathcal{S}_n$. Combining this with Theorem 3.11, we see that in the sense of the Local Asymptotic Minimax Theorem [29, Theorem 7.12], $\hat{\vartheta}_n$ is in fact optimal with rate $\sqrt{n}$ (cf. [13, Theorem 2]).

**Example 3.16.** More generally, for $M = N = 1$, $\sigma \equiv 1$, and any drift $b\colon \mathbb{R} \to \mathbb{R}$ the process $\tilde{Z} = (\tilde{Z}_t)_{t\in[0,\infty)}$ defined by

$$\tilde{Z}_t := Z_t - \int_0^t b(Z_s)ds \quad \text{for all } t \in [0, \infty)$$

is obviously a solution to the "signal in white noise" equation

$$d\tilde{Z}_t = S_{(\vartheta,T)}(t)dt + dW_t. \tag{3.27}$$

We will now discuss some known results about this equation. Note that even if $Z$ satisfies the recurrence assumption (A3), $\tilde{Z}$ does not. Ibragimov and Khasminskii treat the case where $\vartheta$ is fixed and known and $T$ is to be estimated (see [39, Sections II.7 and III.5]). They show asymptotic normality and efficiency for the maximum likelihood and Bayesian estimators with a normalisation factor that coincides asymptotically with

$$(\delta_n)^{-1}_{d+1,d+1} \left(\mathcal{I}_{d+1,d+1}\right)^{-1/2},$$

when translated into our notation (note that they use a different parametrisation: "our $T$" takes the place of "their $\theta^{-1}$", explaining the difference in the constants appearing in their presentation). So both rate and limit variance are the right ones in the sense of the Local Asymptotic Minimax Theorem. Golubev (see [21], or compare [7] for a more detailed probabilistic explanation) gives an estimator for $T$ under unknown infinite-dimensional $\vartheta$ (the vector of the Fourier-coefficients of the signal) which he proves to be asymptotically normal and efficient, where the normalisation factor is (when translated into our notation) given by

$$n^{3/2}\left(\frac{1}{12T^4}\int_0^1 (S'_\vartheta(s))^2 ds\right)^{1/2} = (\delta_n)^{-1}_{d+1,d+1}\left(\frac{1}{4}\mathcal{I}_{d+1,d+1}\right)^{-1/2}.$$

So while the rate is indeed $\delta_n$, the limit variance for Golubev's estimator apparently differs from what should be the optimal value by a factor of 4. This is due to the fact that he studies a slightly different model in which the driving Brownian Motion is two-sided, and the process is observed over time intervals $[-n/2, n/2]$ and not $[0, n]$. This can be interpreted as two independent "signal in white noise" models $\tilde{Z}^{(1)}$, $\tilde{Z}^{(2)}$ each being observed over the interval $[0, n/2]$, where $\tilde{Z}^{(1)}$ follows (3.27) and $\tilde{Z}^{(2)}$ follows (3.27) with the signal replaced by the same signal run backwards in time. Obviously, $\tilde{Z}^{(1)}$ and $\tilde{Z}^{(2)}$ both generate the same Fisher Information $\mathcal{I}_{d+1,d+1}(1/2)$, using the notation of the proof of Theorem 3.11. As a consequence of the independence structure, the Fisher Information in the experiment arising from observation of $(\tilde{Z}^{(1)}, \tilde{Z}^{(2)})$ indeed turns out to be

$$2 \cdot \mathcal{I}_{d+1,d+1}(1/2) = 2 \cdot (1/2)^3 \mathcal{I}_{d+1,d+1}(1) = \frac{1}{4}\mathcal{I}_{d+1,d+1},$$

showing that Golubev's estimator in fact has what is the optimal limit variance in our model.

We close this section with the two purely analytical Lemmas 3.17 and 3.18 that have already been referred to above. Lemma 3.17 was used in the proof of Lemma 3.8, and Lemma 3.18 came into play in the final step of the proof of Theorem 3.11 when showing that $A_n$ vanishes in the limit. Both of these Lemmas are elementary, and it is highly unlikely that this is the first time they have ever been stated. However, since our humble efforts of finding them in the literature remained fruitless, we decided to simply provide our own proofs for the sake of the completeness of this thesis.

**Lemma 3.17.** *If $f\colon [0,\infty) \to [0,\infty)$ is a non-decreasing continuous function with*

$$\frac{f(t)}{t} \xrightarrow{t\to\infty} 1, \tag{3.28}$$

*then*

$$\frac{k+1}{t^{k+1}} \int_0^t s^k df(s) \xrightarrow{t\to\infty} 1$$

*for all $k \in \mathbb{N}_0$.*

*Proof.* Using the Stieltjes product formula, we can write

$$\begin{aligned}\frac{k+1}{t^{k+1}} \int_0^t s^k df(s) &= \frac{k+1}{t^{k+1}} \left( t^k f(t) - \int_0^t f(s) ds^k \right) \\ &= (k+1)\frac{f(t)}{t} - \frac{k+1}{t^{k+1}} \int_0^t f(s) k s^{k-1} ds \\ &= (k+1)\frac{f(t)}{t} - k - \frac{k(k+1)}{t^{k+1}} \int_0^t (f(s)-s) s^{k-1} ds.\end{aligned}$$

By (3.28), the first summand converges to $k+1$ for $t \to \infty$, so it remains to prove that

$$g(t) := \frac{k+1}{t^{k+1}} \int_0^t (f(s)-s) s^{k-1} ds$$

vanishes for $t \to \infty$. Let $\varepsilon > 0$. According to (3.28), there is some $t_0 \in (0,\infty)$ such that

$$\left| \frac{f(s)}{s} - 1 \right| \le \varepsilon \quad \text{for all } s \in [t_0, \infty).$$

Then

$$\begin{aligned}|g(t)| &\le \frac{k+1}{t^{k+1}} \int_0^{t_0} |f(s)-s|\, s^{k-1} ds + \frac{k+1}{t^{k+1}} \int_{t_0}^t (s\varepsilon) s^{k-1} ds \\ &\le \frac{k+1}{t^{k+1}} \int_0^{t_0} |f(s)-s|\, s^{k-1} ds + \varepsilon\end{aligned}$$

for all $t \in [t_0, \infty)$. As the integral in the last step is finite and does not depend on $t$, this shows that

$$\limsup_{t\to\infty} |g(t)| \le \varepsilon,$$

which completes the proof, since $\varepsilon$ was chosen arbitrarily. □

**Lemma 3.18.** *Let $U \subset \mathbb{R}^D$ open and $f\colon U \to \mathbb{L}^2_{\mathrm{loc}}\big([0,\infty);\mathbb{R}^N\big)$ continuously differentiable in the sense of (S2) and (S3), i.e. for each $u \in U$ there is a function $\dot f(u) \in \mathbb{L}^2_{\mathrm{loc}}\big([0,\infty);\mathbb{R}^{N\times D}\big)$ such that for every $t \in (0,\infty)$*

$$\int_0^t \left| \frac{f(u) - f(v) - \dot f(u)(u-v)}{|u-v|} \right|^2 dx \to 0, \quad \text{as } v \to u,$$

*and*

$$\int_0^t \left| \dot f(u) - \dot f(v) \right|^2 dx \to 0, \quad \text{as } v \to u.$$

*Then for any two sequences $(u_n)_{n\in\mathbb{N}}, (v_n)_{n\in\mathbb{N}} \subset U$ converging to $u \in U$ and any $t \in (0,\infty)$ we have*

$$\int_0^t \left| \frac{f(u_n) - f(v_n) - \dot f(u_n)(u_n-v_n)}{|u_n-v_n|} \right|^2 dx \xrightarrow{n\to\infty} 0.$$

*Proof.* Let $\varepsilon > 0$ and choose $\delta > 0$ such that $B_\delta(u) \subset U$ and for all $v \in B_\delta(u)$ we have

$$\int_0^t \left| \dot f(u) - \dot f(v) \right|^2 dx < \frac{\varepsilon^2}{4}.$$

There is a natural number $n_0 \in \mathbb{N}$ such that for $n \geq n_0$ all $u_n$ and $v_n$ belong to $B_\delta(u)$. Then necessarily $B_\delta(u)$ also contains any convex combination $\lambda u_n + (1-\lambda)v_n$ with $\lambda \in [0,1]$. As our notion of differentiability is equivalent to Frechét differentiability in $\mathbb{L}^2\big([0,t];\mathbb{R}^N\big)$ of $u \mapsto f(u)|_{[0,t]}$ for every $t \in (0,\infty)$, this allows us to apply the generalised Mean Value Theorem according to which for all $n \geq n_0$ we can write

$$f(u_n) - f(v_n) = \int_0^1 \dot f\big(\lambda u_n + (1-\lambda)v_n\big)(u_n - v_n) d\lambda,$$

where the integral is understood as an $\mathbb{L}^2\big([0,t];\mathbb{R}^N\big)$-valued Riemann integral. Then

$$\begin{aligned}
&\left\| \frac{|f(u_n) - f(v_n) - \dot f(u_n)(u_n-v_n)|}{|u_n-v_n|} \right\|_{\mathbb{L}^2([0,t])} \\
\leq &|u_n - v_n|^{-1} \int_0^1 \left\| \big(\dot f(\lambda u_n + (1-\lambda)v_n) - \dot f(u_n)\big)(u_n - v_n) \right\|_{\mathbb{L}^2([0,t])} d\lambda \\
\leq &\int_0^1 \left( \int_0^t \left| \dot f(\lambda u_n + (1-\lambda)v_n) - \dot f(u_n) \right|^2 dx \right)^{1/2} d\lambda \\
\leq &\sqrt{2} \int_0^1 \left( \int_0^t \left| \dot f(\lambda u_n + (1-\lambda)v_n) - \dot f(u) \right|^2 dx + \int_0^t \left| \dot f(u) - \dot f(u_n) \right|^2 dx \right)^{1/2} d\lambda \\
< &\sqrt{2} \int_0^1 \left( 2 \cdot \frac{\varepsilon^2}{4} \right)^{1/2} d\lambda = \varepsilon,
\end{aligned}$$

which proves the Lemma. $\square$

# List of symbols

| | |
|---|---|
| $\mathbb{N}$ | the set $\{1, 2, 3, \ldots\}$ of all natural numbers |
| $\mathbb{N}_0$ | the set $\{0, 1, 2, 3, \ldots\}$ of all natural numbers including zero |
| $\mathbb{R}$ | the set of all real numbers, the one-dimensional euclidean space |
| $\mathtt{cst}$ | an arbitrary finite positive real constant whose exact value is of no particular interest and may vary even within the same line |
| $U \times V$ | the cartesian product of the sets $U$ and $V$; its elements $x \in U \times V$ are always assumed to be columns $x = (u, v)^\top$ with $u \in U$, $v \in V$; however, when combining $x = (u, v)^\top \in U \times V$ and $x' = (u', v')^\top \in U' \times V'$ into elements of $U \times V \times U' \times V'$, we often use the notation $(x, x') := (u, v, u', v')^\top$ instead of the formally consistent yet cumbersome $(x^\top, x'^\top)^\top$ (these conventions canonically extend to cartesian products of more than two sets) |
| $\mathbb{R}^n$ | the $n$-fold cartesian product of $\mathbb{R}$, the $n$-dimensional euclidean space |
| $\operatorname{span}(U)$ | the smallest linear subspace of $\mathbb{R}^n$ containing the set $U \subset \mathbb{R}^n$ |
| $0_n$ | short for $(0, \ldots, 0)^\top \in \mathbb{R}^n$ |
| $e_i$ | the $i$-th canonical unit vector in $\mathbb{R}^n$; the dimension $n$ is not reflected in the notation, but is always clear from the context |
| $\Box_i$ | some arbitrary element in $\mathbb{R}^i$ where the exact entries are irrelevant for us |
| $\lvert x \rvert$ | the euclidean norm $\sqrt{\sum_{i=1}^n x_i^2}$ of $x \in \mathbb{R}^n$, usually simply referred to as the absolute value; induces the euclidean topology on $\mathbb{R}^n$ |
| $\lvert x \rvert_1$ | the 1-norm $\sum_{i=1}^n \lvert x_i \rvert$ of $x \in \mathbb{R}^n$ |
| $B_\varepsilon(x)$ | the euclidean ball $\{y \in \mathbb{R}^n \mid \lvert x - y \rvert < \varepsilon\} \subset \mathbb{R}^n$ with radius $\varepsilon > 0$ and centre $x \in \mathbb{R}^n$ |
| $\operatorname{int}(U)$, $\bar{U}$, $\partial U$ | the interior, closure and boundary of the set $U$ with respect to the topology of the space in which it is contained |

| | |
|---|---|
| $\mathbb{R}^{n\times m}$ | the set of all real matrices $A = (A_{i,j})_{i=1,\dots,n,\, j=1,\dots,m}$ with $n$ rows and $m$ columns |
| $1_{n\times n}$ | the identity matrix in $\mathbb{R}^{n\times n}$ |
| $\lvert A\rvert$ | the Frobenius norm (or Hilbert-Schmidt norm) $\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m} A_{i,j}^2}$ of a matrix $A \in \mathbb{R}^{n\times m}$, induces the euclidean topology on $\mathbb{R}^{n\times m}$ |
| $\operatorname{tr}(A)$ | the trace $\sum_{i=1}^{n} A_{i,i}$ of a quadratic matrix $A \in \mathbb{R}^{n\times n}$ |
| $\det(A)$ | the determinant of a quadratic matrix $A \in \mathbb{R}^{n\times n}$ |
| $f \equiv c$ | short for saying that the function $f$ is constant with $c$ being its only value |
| $C(U;V)$ | the set of all continuous functions $f\colon U \to V$, where $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$ are equipped with the euclidean topology; $C(U;V)$ equipped with the topology of locally uniform convergence becomes itself a topological space; abbreviated to $C(U)$ for $V = \mathbb{R}$ |
| $C^k(U;V)$ | the set of all $f \in C(U;V)$ which are $k$-times continuously (partially) differentiable, $k \in \mathbb{N} \cup \{\infty\}$; abbreviated to $C^k(U)$ for $V = \mathbb{R}$ |
| $C_b^k(U;V)$ | the set of all $f \in C^k(U;V)$ whose partial derivatives of any order up to $k$ (including order zero, i.e. $f$ itself) are bounded; abbreviated to $C_b^k(U)$ for $V = \mathbb{R}$ |
| $\partial_t^k, \partial_{x_j}^k, \dots$ | the $k$-th (partial) derivative with respect to the variable $t$, $x_j$, $\dots$, the case $k = 0$ corresponds to taking no derivative at all |
| $\partial_t, \partial_{x_k}, \dots$ | short for $\partial_t^1, \partial_{x_j}^1, \dots$ |
| $\partial_x^\alpha$ | short for $\partial_{x_n}^{\alpha_n} \cdots \partial_{x_1}^{\alpha_1}$ with $x \in \mathbb{R}^n$ and $\alpha \in \mathbb{N}_0^n$ |
| $\nabla_x u$ | the gradient $(\partial_{x_1} u, \dots, \partial_{x_n} u)^\top$ with respect to $x \in \mathbb{R}^n$ of a real-valued function $u$ that depends on $x$ but possibly also on other variables; if $u$ does not depend on any other variables and there is no risk of ambiguity, we simply write $\nabla u$ |
| $u'$ | short for the derivative of a function $u$ of only one variable |
| $\mathcal{D}\big([0,\infty);\mathbb{R}^n\big)$ | the space of all càdlàg functions $f\colon [0,\infty) \to \mathbb{R}^n$, equipped with the Skorohod topology (see [6, p. 111] for a definition in the case $n = 1$, the general case is handled by canonically identifying $\mathcal{D}\big([0,\infty);\mathbb{R}^n\big)$ with $\big(\mathcal{D}\big([0,\infty);\mathbb{R}\big)\big)^n$) |

| | |
|---|---|
| $\mathcal{B}(U)$ | the Borel-$\sigma$-field of a topological space $U$; the specific topology used is not reflected in the notation, but it is either clear from the context or it is mentioned in the text; unless stated otherwise, the term measurable (for sets or functions) is always meant in the sense of Borel-measurable |
| $\sigma(\mathcal{F})$ | the $\sigma$-field on $\Omega$ generated by the family $\mathcal{F}$ of functions mapping $\Omega$ to a measurable space $(E, \mathcal{E})$ |
| $1_A$ | the indicator function of the set $A$ |
| $\mathbb{L}^p(\mu; \mathbb{R}^n)$ | the space of all measurable functions $f\colon (E, \mathcal{E}) \to (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ such that $\lvert f\rvert^p$ is integrable with respect to the measure $\mu\colon \mathcal{E} \to [0, \infty]$; analogously used for $\mathbb{R}^{n\times m}$-valued functions; abbreviated to $\mathbb{L}^p(\mu)$ for $\mathbb{R}$-valued functions |
| $\mathbb{L}^p(E; \mathbb{R}^n)$ | alternate writing for $\mathbb{L}^p(\mu; \mathbb{R}^n)$ when $\mu$ is Lebesgue's measure on $E \subset \mathbb{R}^k$; analogously used for $\mathbb{R}^{n\times m}$-valued functions; abbreviated to $\mathbb{L}^p(E)$ for $\mathbb{R}$-valued functions |
| $\mathbb{L}^p_{\mathrm{loc}}(\cdots)$ | analogous to any of the above with the only difference that the respective function is supposed to be integrable only locally, i.e. in restriction to any compact subset of $E$ (which has to be a topological space for this to make sense) |
| $\lVert f\rVert_{\mathbb{L}^2(U)}$ | the $\mathbb{L}^2$ norm $\sqrt{\int_U \lvert f(x)\rvert^2\, dx}$ of a function $f$ that is defined on (some superset of) $U$ |
| $\lVert f\rVert_\infty$ | the supremum norm $\sup_{x\in U} \lvert f(x)\rvert$ of a bounded function $f\colon U \to \mathbb{R}^n$ |
| $\mathrm{Lip}(f)$ | the optimal Lipschitz constant $\sup_{x,y\in U} \frac{\lvert f(x)-f(y)\rvert}{\lvert x-y\rvert}$ of a Lipschitz continuous function $f\colon U \to \mathbb{R}^n$ |
| $f\vert_W$ | the restriction of the mapping $f\colon U \to V$ to some $W \subset U$ |
| $\int \cdots dW_t$ | the Itō integral with respect to the Brownian Motion $W$ |
| $\int \cdots \circ\, dW_t$ | the Stratonovich integral with respect to the Brownian Motion $W$ |
| $\mathcal{L}(X\vert P)$ | the law or image measure under the probability law $P$ of the random variable $X$ on its image space |
| $\mathcal{N}(\mu, \Sigma)$ | the Gaussian law on $\mathbb{R}^n$ with mean vector $\mu \in \mathbb{R}^n$ and covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$ |
| $\mathbb{E}_P[X]$ | the expected value of the real-valued random variable $X$ under the probability law $P$ |
| $\frac{d\mu}{d\nu}$ | the Radon-Nikodym derivative of the measure $\mu$ with respect to the measure $\nu$ |

| | |
|---|---|
| $\langle M, N \rangle$ | the quadratic covariation process of the locally square integrable martingales $M$ and $N$ |
| $\langle M \rangle$ | short for $\langle M, M \rangle$, the quadratic variation process of the locally square integrable martingale $M$ |

# Bibliography


[1] J. Azéma, M. Duflo, D. Revuz: Mesure invariante des processus de Markov récurrents. Séminaire de Probabilités III. In: *Lecture Notes in Mathematics* Vol. 88 (1969).

[2] R. Bass: Diffusions and Elliptic Operators. Springer, 1998.

[3] D. Bell: The Malliavin Calculus. Longman Scientific & Technical, 1987.

[4] N. Berglund, B. Gentz: Stochastic dynamic bifurcations and excitability. In: *Stochastic Methods in Neuroscience, Carlo Laing and Gabriel Lord (Eds)* (2008), pp. 1-29.

[5] N. Berglund, D. Landon: Mixed-mode oscillations and interspike interval statistics in the stochastic FitzHugh-Nagumo model. In: *Nonlinearity* Vol. 25(8) (2012), pp. 2303-2335.

[6] P. Billingsley: Convergence of probability measures. Wiley, 1968.

[7] I. Castillo, C. Lévy-Leduc, C. Matias: Exact Adaptive Estimation of the Shape of a Periodic Function with Unknown Period Corrupted by White Noise. In: *Mathematical Methods of Statistics* Vol. 15 (2006), pp. 1-30.

[8] K. S. Cole, H.A. Antosiewicz, P. Rabinowitz: Automatic Computation of Nerve Excitation. In: *Journal of the Society for Industrial and Applied Mathematics* Vol. 3 (1955), pp. 153-172.

[9] A. Crudu, A. Debussche, A. Muller, O. Radulescu: Convergence of stochastic gene networks to hybrid piecewise deterministic processes. In: *The Annals of Applied Probability* Vol. 22(5) (2012), pp. 1822-1859.

[10] N. Cuneo, J. P. Eckmann: Non-Equilibrium Steady States for Chains of Four Rotors. In: *Communications in Mathematical Physics* Issue 1 (2016), pp. 185-221.

[11] N. Cuneo, J. P. Eckmann, C. Poquet: Non-equilibrium steady state and subgeometric ergodicity for a chain of three coupled rotors. In: *Nonlinearity* Vol. 28 (2015), pp. 2397-2421.

[12] R. Davies: Asymptotic Inference When the Amount of Information Is Random. In: *Proceedings of the Berkeley Symposium in Honour of J. Neyman and J. Kiefer* Vol. II, Wadsworth, 1985.

[13] H. Dehling, B. Franke, T. Kott: Drift Estimation for a Periodic Mean Reversion Process. In: *Statistical Inference for Stochastic Processes* Vol. 13 (2010), pp. 175-192.

[14] M. Desroches, J. Guckenheimer, B. Krauskopf, C. Kuehn, H. M. Osinga, M. Wechselberger: Mixed-mode oscillations with multiple time scales. In: *SIAM Review* Vol. 54(2) (2012), pp. 211-288.

[15] A. Destexhe: Conductance-based integrate and fire models. In: *Neural Computation* Vol. 9 (1997), pp. 503-514.

[16] S. Ditlevsen, P. Lánský: Estimation of the input parameters in the Feller neuronal model. In: *Physical Review E* Vol. 73 (2006), 061910.

[17] S. Ditlevsen, E. Löcherbach: On oscillating systems of interacting neurons. In: *Stochastic Processes and their Applications* Vol. 127 (2017), pp. 1840-1869.

[18] D. Down, S. P. Meyn, R. L. Tweedie: Exponential and Uniform Ergodicity of Markov Processes. In: *Annals of Probability* Vol. 23(4) (1995), pp. 1671-1691.

[19] K. Endler: Periodicities in the Hodgkin-Huxley model and versions of this model with stochastic input. Master Thesis, Institute of Mathematics, University of Mainz (see under http://ubm.opus.hbz-nrw.de/volltexte/2012/3083/), 2012.

[20] A. Faggionato, D. Gabrielli, M. Ribezzi Crivellari: Non-equilibrium Thermodynamics of Piece-wise Deterministic Markov Processes. In: *Journal of Statistical Physics* Vol. 137 (2009), pp. 259-304.

[21] G. Golubev: Estimating the Period of a Signal of Unknown Shape Corrupted by White Noise. In: *Problems in Information Transmission* Vol. 24 (1988), pp. 38-52.

[22] J. Guckenheimer, R. A. Oliva: Chaos in the Hodgkin–Huxley model. In: *SIAM Journal on Applied Dynamical Systems* Vol. 1(1) (2002), pp. 105-114.

[23] M. Hairer: On Malliavin's proof of Hörmander's Theorem. In: *Bulletin des Sciences Mathématiques* Vol. 135 (2011), pp. 650-666.

[24] M. Hairer, J. C. Mattingly: Slow energy dissipation in anharmonic oscillator chains. In: *Communications on Pure and Applied Mathematics* Vol. 62 (2009), pp. 999-1032.

[25] T. Harris: The existence of stationary measures for certain Markov processes. In: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* Vol. II (1956), pp. 113-124.

[26] B. Hassard: Bifurcation of periodic solutions of the Hodgkin-Huxley model for the squid giant axon. In: *Journal of Theoretical Biology* Vol. 71 (1978), pp. 401-420.

[27] A. L. Hodgkin, A. F. Huxley: A Quantitative Description of Membrane Current And Its Application to Conduction And Excitation in Nerve. In: *Journal of Physiology* Volume 117, Vol. 4 (1952), pp. 500-544.

[28] S. Holbach: Local asymptotic normality for shape and periodicity in the drift of a time inhomogeneous diffusion. In: *Statistical Inference for Stochastic Processes*, DOI 10.1007/s11203-017-9157-5.

[29] R. Höpfner: Asymptotic Statistics with a View to Stochastic Processes. de Gruyter, 2014.

[30] R. Höpfner: On a set of data for the membrane potential in a neuron. In: *Mathematical Biosciences* Vol. 207 (2007), pp. 275-301.

[31] R. Höpfner, Y. A. Kutoyants: On LAN for Parametrized Continuous Periodic Signals in a Time Inhomogeneous Diffusion. In: *Statistics & Decisions* Vol. 27 (2009), pp. 309-326.

[32] R. Höpfner, Y. A. Kutoyants: Estimating Discontinuous Periodic Signals in a Time Inhomogeneous Diffusion. In: *Statistical Inference for Stochastic Processes* Vol. 13 (2010), pp. 193-230.

[33] R. Höpfner, Y. A. Kutoyants: Estimating a Periodicity Parameter in the Drift of a Time Inhomogeneous Diffusion. In: *Mathematical Methods of Statistics* Vol. 20 (2011), pp. 58-74.

[34] R. Höpfner, E. Löcherbach: On some ergodicity properties for time inhomogeneous Markov processes with $T$-periodic semigroup. arXiv:1012.4916 [math.PR].

[35] R. Höpfner, E. Löcherbach, M. Thieullen: Ergodicity for a Stochastic Hodgkin-Huxley Model Driven by Ornstein-Uhlenbeck Type Input. In: *Annales de l'Institut Henri Poincaré* Vol. 1 (2016), pp. 483-501.

[36] R. Höpfner, E. Löcherbach, M. Thieullen: Strongly degenerate time inhomogeneous SDEs: densities and support properties. Application to a Hodgkin-Huxley system with periodic input. In: *Bernoulli* Vol. 23(4A) (2017), 2587-2616.

[37] R. Höpfner, E. Löcherbach, M. Thieullen: Ergodicity and Limit Theorems for Degenerate Diffusions with Time Periodic Drift. Application to a Stochastic Hodgkin-Huxley Model. In: *ESAIM P&S* Vol. 20 (2016), pp. 527-554.

[38] R. Hörmander: Hypoelliptic second order differential equations. In: *Acta Mathematica* Vol. 119 (1967), pp. 147-171.

[39] I. A. Ibragimov, R. Z. Khasminskii: Statistical Estimation. Springer, 1981.

[40] N. Ikeda, S. Watanabe: Stochastic Differential Equations and Diffusion Processes. North-Holland Library, 2nd edition, 1989.

[41] E. M. Izhikevich: Dynamical Systems in Neuroscience. The MIT Press, 2007.

[42] J. Jacod, A. Shiryaev: Limit Theorems for Stochastic Processes. Springer, 2nd edition, 2002.

[43] I. Karatzas, S. E. Shreeve: Brownian Motion and Stochastic Calculus. Springer, 1988.

[44] R. Z. Khasminskii: Stochastic Stability of Differential Equations. Second English Edition, Springer, 2012.

[45] Y. A. Kutoyants: Statistical Inference for Ergodic Diffusion Processes. Springer, 2004.

[46] P. Lánský, L. Sacerdote, F. Tomassetti: On the Comparison of Feller and Ornstein-Uhlenbeck Models for Neural Activity. In: *Biological Cybernetics* Vol. 73 (1995), pp. 457-465.

[47] L. LeCam: Locally asymptotically normal families of distributions. In: *University of California Publications in Statistics* Vol. 3 (1960), pp. 37-98.

[48] L. LeCam, G. Yang: Asymptotics in Statistics. Some Basic Concepts. Springer, 1990.

[49] J. Mattingly, A. Stuart, D. Higham: Ergodicity of SDEs and approximations: locally Lipschitz vector fields and degenerate noise. In: *Stoch. Proc. Appl.* Vol. 101 (2002), pp. 185-232.

[50] S. Meyn, R. Tweedie: Stability of Markovian processes I: criteria for discrete-time chains. In: *Advances in Applied Probability* Vol. 24 (1992), pp. 542-574.

[51] A. Millet, M. Sanz-Solé: A simple proof of the support theorem for diffusion processes. In: *Séminaire de probabilités de Strasbourg* Vol. 28 (1994), pp. 36-48.

[52] C. Morris, H. Lecar: Voltage oscillations in the barnacle giant muscle fiber. In: *Biophysical Journal* Vol. 35 (1981), pp. 193-213.

[53] E. Nummelin: A splitting technique for Harris recurrent Markov chains. In: *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* Vol. 43 (1978), pp. 309-318.

[54] E. Nummelin: General irreducible Markov chains and non-negative operators. Cambridge University Press, 1985.

[55] K. Pakdaman, M. Thieullen, G. Wainrib: Fluid limit theorems for stochastic hybrid systems with application to neuron models. In: *Advances in Applied Probability* Vol. 42 (2010), pp. 761-794.

[56] J. Prüss, M. Wilke: Gewöhnliche Differentialgleichungen und dynamische Systeme. Birkhäuser, 2010.

[57] J. Realpe-Gomez, T. Galla, A. J. McKane: Demographic noise and piecewise deterministic Markov processes. In: *Physical Review E* Vol. 86 (2012), 011137.

[58] D. Revuz: Markov Chains. North-Holland, 1975.

[59] D. Revuz, M. Yor: Continuous Martingales and Brownian Motion. Springer, 3rd edition, 2005.

[60] L. Rey-Bellet, L. E. Thomas: Exponential convergence to non-equilibrium stationary states in classical statistical mechanics. In: *Communications in Mathematical Physics* Vol. 225 (2002), pp. 309-329.

[61] J. Rinzel, B. Ermentrout: Analysis of neural excitability and oscillations. In: *Methods in neuronal modeling: from ions to networks, 2nd edition* (1998), pp. 251-291.

[62] J. Rubin, M. Wechselberger: Giant squid-hidden canard: the 3D geometry of the Hodgkin-Huxley model. In: *Biological Cybernetics* Vol. 97(1) (2007), pp. 5-32.

[63] D. W. Stroock, S. R. S. Varadhan: On the support of diffusion processes with applications to the strong maximum principle. In: *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* Vol. 3 (1972), pp. 333-359.

[64] A. Veretennikov: On polynomial mixing bounds for stochastic differential equations. In: *Stochastic Processes and their Applications* Vol. 70 (1997), pp. 115-127.

[65] M. Vidyasagar: Nonlinear Systems Analysis. Second Edition, Prentice Hall, 1993.