Heriot-Watt University

Spatio-temporal Stochastic Hybrid Models of Biological Excitable Membranes

Martin Georg Riedler

July 2011

Submitted for the degree of
Doctor of Philosophy in Mathematics
on completion of research in the
Department of Mathematics,
School of Mathematical and Computer Sciences.

Abstract

A large number of biological systems are intrinsically random, in particular, biological excitable membranes, such as neuronal membranes, cardiac tissue or models for calcium dynamics. The present thesis is concerned with hybrid stochastic models of spatio-temporal dynamics of biological excitable membranes using Piecewise Deterministic Markov Processes (PDMPs). This class of processes allows a precise mathematical description of the internal noise structure of excitable membranes. Overall the aim of the thesis is two-fold: On the one hand, we establish a general hybrid modelling framework for biological excitable membranes and, on the other hand, we are interested in a general advance of PDMP theory which the former necessitates. Regarding the first aim we exemplify the modelling framework on the classical Hodgkin-Huxley model of a squid giant axon. Regarding the latter we present a general PDMP theory incorporating spatial dynamics and present tools for their analysis. Here we focus on two aspects.

Firstly, we consider the approximation of PDMPs by deterministic models or continuous stochastic processes. To this end we derive as general theoretical tools a law of large numbers for PDMPs and martingale central limit theorems. The former establishes a connection of stochastic hybrid models to deterministic models given, e.g., by systems of partial differential equations. Whereas the latter connects the stochastic fluctuations in the hybrid models to diffusion processes. Furthermore, these limit theorems provide the basis for a general Langevin approximation to PDMPs, i.e., certain stochastic partial differential equations that are expected to be similar in their dynamics to PDMPs.

Secondly, we also address the question of numerical simulation of PDMPs. We present and analyse the convergence in the pathwise sense of a class of simulation methods for PDMPs in Euclidean space.

To Ildiko

Acknowledgements

I thank my supervisor Dr Evelyn Buckwar for indispensable advice, helpful discussions and fruitful collaborations. Her constant support made working on PhD thesis on overall enjoyable experience. I further thank Dr Thorsten Sickenberger for valuable mathematical discussions and Dr Michèle Thieullen and Dr Gilles Wainrib for stimulating collaborations.

I acknowledge support from the Department of Mathematics at Heriot-Watt University and the Engineering and Physics Research Council, project EP/E03635X/1, funding my PhD project. I further acknowledge support from the ANR project MANDy (ANR-09-BLAN-0008-01), the BC/DAAD ARC project Nr. 1349/50021880, the ARC project Nr. 1333, the UK Mathematical Neuroscience Network (MNN) and the Cell Signalling Network (SIGNET) funding research visits to various degrees over the course of my PhD project.

Finally, I want to thank the organisers of the Workshop on Stochastic Models in Neuroscience at CIRM, Marseille, in January 2010 who early in the course of my PhD project gave me the opportunity to present my work. The collaborations I established at this workshop resulted in a central part of this thesis and are still ongoing for further research based on the present results.

ACADEMIC REGISTRY
Research Thesis Submission

Name:	MARTIN GEORG RIEDLER		
School/PGI:	MACS		
Version: *(i.e. First, Resubmission, Final)*	FINAL	Degree Sought (Award **and** Subject area)	PhD MATHEMATICS

Declaration

In accordance with the appropriate regulations I hereby submit my thesis and I declare that:

1) the thesis embodies the results of my own work and has been composed by myself
2) where appropriate, I have made acknowledgement of the work of others and have made reference to work carried out in collaboration with other persons
3) the thesis is the correct version of the thesis for submission and is the same version as any electronic versions submitted*.
4) my thesis for the award referred to, deposited in the Heriot-Watt University Library, should be made available for loan or photocopying and be available via the Institutional Repository, subject to such conditions as the Librarian may require
5) I understand that as a student of the University I am required to abide by the Regulations of the University and to conform to its discipline.

* *Please note that it is the responsibility of the candidate to ensure that the correct version of the thesis is submitted.*

Signature of Candidate:		Date:	

Submission

Submitted By *(name in capitals):*	
Signature of Individual Submitting:	
Date Submitted:	

For Completion in the Student Service Centre (SSC)

Received in the SSC by *(name in capitals):*			
Method of Submission *(Handed in to SSC; posted through internal/external mail):*			
*E-thesis Submitted (**mandatory for final theses)***			
Signature:		Date:	

Please note this form should bound into the submitted thesis.

Updated February 2008, November 2008, February 2009, January 2011

Contents

List of Figures

Chapter 1

Introduction

It is widely recognised that a large number of processes occurring in biology are in fact intrinsically random. Particularly, in dealing with ever smaller structures the influence of noise on the behaviour of the systems becomes ever stronger and thus should not be neglected in mathematical modelling. To capture and model the dynamics of biological systems, such as, e.g., neuronal membranes, a wide range of different mathematical descriptions are used. Among these models are deterministic ordinary differential equation (ODE) / partial differential equation (PDE) models, stochastic Markov chain models and stochastic ordinary differential equation (SODE) / stochastic partial differential equation (SPDE) models. A particularly interesting new class of models are stochastic hybrid systems which combine continuous evolution of macroscopic variables with Markov chain models for smaller structures. In recent years the number of applications of hybrid stochastic processes to model systems in mathematical biology, (bio-)chemical reaction kinetics and mathematical neuroscience have increased rapidly. These models either stem from a direct modelling approach, e.g., models of excitable biological membranes [32, 106], or result from a multiscale approximation to more accurate particle models, e.g., in biochemical reaction systems [3, 58, 70, 107] or in gene regulatory networks [123], gene transcription [83], DNA modelling [85] or cell biology [49].

The present thesis is concerned with hybrid stochastic processes and their application to modelling the spatio-temporal evolution of excitable biological membranes such as neuronal membranes, cardiac tissue or membranes involved in calcium dynamics. We focus on models employing *Piecewise Deterministic Markov Processes* (PDMPs), which are an important class of hybrid systems including those considered in most of the aforementioned studies. In particular we show in the present thesis that PDMPs are the appropriate mathematical models to exactly capture the stochastic dynamics arising from smaller building blocks of an excitable membrane model. PDMPs were originally introduced by M. Davis in [36, 37] as a general class of non-diffusive stochastic processes. They are strong Markov processes, taking values in a Euclidean space, which combine continuous deterministic time evolution and discontinuous, instantaneous, random 'jump' events. Specifically, the dynamics of the two components are intrinsically intertwined. On the one hand, the continuous time-evolution, defined by ordinary differential equations (ODEs) in Davis' framework, depends on the outcomes of discrete events, e.g., randomly changing parameters or initial conditions, and, on the other hand, the probability of the discrete events happening, e.g., a random, instantaneous change in a parameter, depends on the time-evolution – the path – of the continuous variables. Davis provided a theoretical framework for the mathematical study of PDMPs as well as applications to stochastic control problems. PDMPs were further investigated in several directions and applications can be found in diverse fields apart from biosciences: stationary distributions [33, 40], communication networks [60], internet traffic [30], financial mathematics [64] and ecology [120] to cite a few.

Regarding neuronal activity in particular, we are interested to model the impact of ion channel stochasticity termed *intrinsic noise* or *channel noise* on action potential generation and propagation. The initial motivation of the present thesis was the observation that PDMPs are particularly well suited for modelling stochastic neuronal membranes. A fact that was simultaneously and independently observed by K. Pakdaman, M. Thieullen and G. Wainrib in [96]. PDMPs naturally combine the stochastic Markov chain models of single ion channels and a deterministic biophysical model of the currents across and along neuronal membranes. Thus they provide an exact description of the channel noise based on the stochasticity of the individual channels. However, we go further than the authors in [96], who take only temporal dynamics into account, by extending the model to include also *spatial dynamics*. In biological excitable membranes a wide range of spatio-temporal effects are observed and they are an essential functional feature of the biological system.
In deterministic PDE models these effects are incorporated as dynamical phenomena and extensively studied. For example, spatio-temporal dynamics are encountered as travelling waves in models of axons modelling the propagation of the action potential from the soma to the dendrites. Further, travelling waves arise in models of calcium dynamics and spiral waves are an important feature in models of cardiac tissue. It is thus a natural question to consider how these effects are affected by channel noise and, as a necessary starting point for these investigations, how spatial dynamics can be incorporated into hybrid stochastic models.

1.1 Aims of the thesis

With respect to modelling of excitable membranes we focus in this thesis on the following aspects:

- Our aim is to present an exact mathematical description of an excitable membrane derived from the building blocks of, on the one hand, continuous-time Markov chain models of ion channels, and, on the other hand, deterministic ODE or PDE models for the flow of current across or across and along the membrane, respectively.
- In particular we show that the underlying physiological model of a neuron incorporating channel noise can be precisely described by the mathematical model class of PDMPs and we provide an explicit example using the Hodgkin-Huxley model for the squid giant axon.

As spatio-temporal models are usually presented in terms of partial differential equations the classical framework of PDMPs as presented by Davis is not sufficient. It only accounts for dynamics of the deterministic component which are generated by systems of ODEs. In general this corresponds to purely temporal models. Therefore a natural starting point of this thesis is to discuss PDMP theory and extend the general theory such that it accounts for the stochastic processes suitable for describing spatio-temporal stochastic hybrid systems. This idea was also stimulated by T. Austin's article [7] wherein the author presents to the best of our knowledge the only spatial hybrid model of a neuronal membrane so far. However, the connection to PDMPs and thus a more general approach is not established in this study. In Section 3.1 we show that the model in [7] is a special case of our PDMP formulation. For models without spatial dynamics the main advantage of the formulation as PDMPs is that there already exists an extensive mathematical theory for its treatment, e. g., [37, 64, 68], in contrast to the case of most other ad-hoc hybrid models in the literature. In particular the theory allows the derivation of diffusion approximations and large deviation results, as presented in [50, 96] and [95], respectively, for the finite-dimensional case. A formulation of spatio-temporal models as PDMPs then provides a suitable starting point to extend the above results to spatially extended models.

This motivates a further aim of this thesis. Thus we may say that overall the aim of the present thesis is two-fold. On the one hand, we are interested in presenting exact models of excitable media. On the other hand, this necessitates a general advance of PDMP theory which is the second general aim of the present thesis. Our contributions to the general theory of PDMPs are the following which are always connected, as they are motivated by, the applicability to excitable membrane models.

- We present a general theory of PDMPs that combine and extend previously published approaches. Whereas often the analysis of PDMPs is restricted to processes taking values in Euclidean space we put a special emphasise on the generality of the state space in view of the application to spatio-temporal hybrid models. An extension of the classical PDMP framework to include spatial dynamics was initially presented by E. Buckwar and the present author in [27]. In this thesis we provide a substantially updated version of this approach. We provide a rigorous framework for a wide class of stochastic spatial models in which discrete random events are globally coupled via continuous space-dependent variables solving PDEs.

– In particular we present as a central result the characterisation of the extended generator of a general class of PDMPs. The extended generator is a central object in the study of Markov processes and allows a precise characterisation of the process. We further discuss the families of PDMPs considered in previous studies as well as those which arise in applications to excitable membranes.

In general an analysis of the effects due to channel noise can be pursued by analytical or numerical means. As PDMPs are rather complex stochastic processes it is generally very difficult to obtain analytical results. Therefore one central aspect is the derivation of approximations to PDMPs by simpler stochastic processes which are expected to preserve the PDMP's qualitative and quantitative features of interest. For numerical studies of the models numerical simulation algorithms have to be derived and an error analysis has to be performed. Within the general PDMP framework we have dealt with these aspects in the following way.

– We present a law of large numbers for PDMPs taking values in Hilbert spaces within the fluid limit setting, i.e., increasing frequency of jumps of decreasing size and variation. It establishes convergence in probability of a sequence of PDMPs to the solution of a deterministic system of abstract evolution equations. Further we present a martingale central limit theorem that provides the convergence of important martingales associated with a PDMP to Hilbert space valued diffusion processes. These theorems allow the derivation of a Langevin approximation to PDMPs in terms of a system of stochastic partial differential equations.
– We present a general class of numerical methods for the pathwise simulation of PDMPs taking values in Euclidean space. We analyse the pathwise error and the asymptotic order of convergence.

With respect to applications to models of neuronal membranes these theoretical findings provide the basis to rigorously derive and analyse analytical and numerical approximations to PDMP models, some of which are already used in the literature. We exemplify the application of these methods to neuron models at the respective places in this thesis.

Finally we briefly comment on the terminology employed throughout this thesis. We use *finite-dimensional PDMPs* to denote PDMPs in the original sense of Davis, i.e., taking values in an Euclidean, i.e., finite-dimensional, space and inter-jump dynamics generated by systems of ODEs. In contrast the PDMPs we establish in this thesis are denoted as *infinite-dimensional PDMPs* as they take values in an infinite-dimensional Hilbert space and inter-jump dynamics generated by abstract evolution equations, e.g., parabolic PDEs. Hence finite-dimensional PDMPs are connected to models of excitable membranes without spatial dynamics and infinite-dimensional PDMPs to models admitting spatio-temporal dynamics.

1.2 Motivation: Stochastic models of neuronal membranes

In this section we discuss the relevance of stochastic hybrid models for neuronal membranes and motivate that PDMPs are the accurate class of stochastic processes to cast these models into a mathematical framework. We briefly review the basic building blocks of membrane models and various approaches that are used to derive a stochastic model of channel noise.

1.2.1 Modelling neuronal excitability

The physiologically important parts of a neuron for its excitability are the cell membrane and various families of ion channels immersed in it, which allow for current flows across the membrane. Then the complex interplay of the transmembrane potential mediated gating of the ion channels gives rise to the excitable dynamics in the transmembrane potential itself. Further, of biological interest are not only local changes in the transmembrane potential but also how these current flows are propagated along spatially extended structures, i.e., how the signal is transmitted along the axon from the soma to the synapses. As the dynamics of ion channels are intrinsically random they introduce variability into the biological system, which is termed channel noise.
For large, homogeneous neuronal structures like the squid giant axon, purely deterministic equations modelling the combined behaviour of the transmembrane potential and the macroscopic average behaviour of channel gating provide a reasonably good description of the excitable dynamics and its propagation. The large number of ion channels involved and the homogeneity of their distribution average out the effect of channel noise. However for small neuronal structures, e.g., as in the central nervous system of vertebrates, purely deterministic modelling is not sufficient, see, e. g. [46], and moreover there appears to be experimental evidence that channel noise plays a functional role [39].
Therefore the task is to find a mathematical model that incorporates both functionally important components and, on the one hand, describes the actual biological system with a sufficient degree of biophysical realism and, on the other hand, is analytically and numerically still tractable.

1.2.2 Review of existing modelling approaches

Efforts have been made to introduce stochasticity into neuronal models from various perspectives. On the one hand from a practical or computational point of view stochastic algorithms have been derived that reproduce a noisy deterministic behaviour. On the other hand from a mere theoretical perspective one wants to derive the deterministic equations as a limit (in some sense) of stochastic models. To derive the intermediate stochastic models we can roughly distinguish between two approaches.

Firstly, an often followed approach, which one might call a top-down approach, consists of adding (small) noise to the deterministic equations, e.g. [117], obtaining an SODE or SPDE model. The original macroscopic system can then easily be retrieved from the stochastic equations by taking the noise intensity to zero. Further, by this approach trial-to-trial variation in the model behaviour and more 'realistic' outcomes of simulations are obtained very simply. Though well suited for external noise, i.e., fluctuations affecting directly the transmembrane potential this ansatz is not suited for modelling channel noise. The drawback is that it does not provide an intrinsic justification for the noise used, its intensity and its type, from physical principles. This means, that introducing noise at this stage of the modelling cannot answer questions such as how in realistic situations microscopic channel noise is transported to and influences the macroscopic level of the system.

In a second, more realistic, bottom-up approach, which we will also follow in our considerations, researchers base their models again partly on deterministic equations for the macroscopic, average behaviour, but also make use of well established microscopic models for the single channels. Single ion channels are successfully modelled by Markov chains with transition rates depending on the transmembrane potential [54]. Now, researchers construct an approximate discrete, stochastic model for a collection of single ion channels interconnected with a deterministic equation for the transmembrane potential where the two parts mutually influence each other. This type of stochastic hybrid models is in general considered for membranes without spatial dynamics [110, 32, 38, 105, 31], except for [7]. Most authors cast their models into so called pseudo-exact algorithms using ad-hoc approximations for the dynamics of the coupled systems. Even if these approximations are avoided, as in [32], the lack of a rigorous mathematical formulation of the stochastic process to be simulated prevents serious investigations of convergence of the simulation algorithm. For example, no proof of the strong Markov property of the resulting process in [32] is given, which is an essential property for the numerical analysis of these processes.

The modelling of the ion channels as individual objects, but using a macroscopic approximation for the flow of ions is well supported by the numbers appearing in biological reality: $10^4 - 10^5$ ions may flow through a single open channel per millisecond

but channel densities are down to less than 10 channels of a certain type per square micrometer [74]. Hybrid models generally take the form of a discrete-time algorithm and it is then shown computationally that a noisy excitable behaviour similar to deterministic models emerges.
Starting from a certain hybrid model proposed in [38] the authors in [50] showed that macroscopic equations can be derived as a deterministic limit by expansion of the generator and taking the number of ion channels to infinity. Another approach using fluid limits and finite dimensional PDMPs has been pursued in [96]. Further, in both works the authors have derived intermediate scale approximations of hybrid models by diffusion processes, i.e., by solutions of SODEs.
We would like to mention that it is possible to extract channel dynamics from macroscopic models of the channel gating and define Markovian kinetics for the distinct classes of ion channels in the membrane. Hence, stochastic hybrid models obtained in this way have the initial macroscopic model as their deterministic limit in the sense of [50, 96]. The rate functions, given in Section A, for the Hodgkin-Huxley model, which we discuss in Section 3.1, are based on this procedure.

We follow this bottom-up approach and propose PDMPs as a precise mathematical formulation of the model of a membrane subject to channel noise. We now motivate this proposition analysing the building blocks of a membrane model in the next subsection.

1.2.3 Derivation of exact models

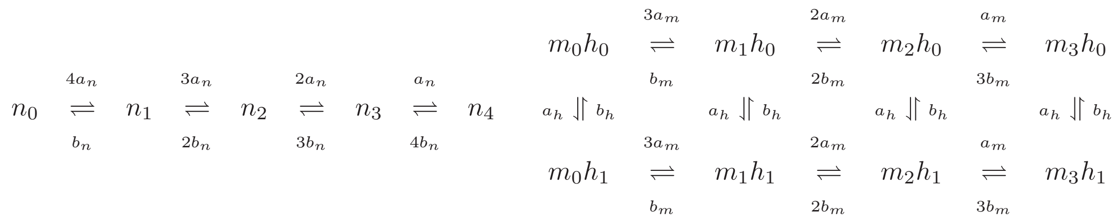
The first building block of the model is a mathematical formulation of the spatio-temporal dynamics of the transmembrane potential u, and as described in [74] this is in general given by the nonlinear cable equation

$$C\dot{u} = D\Delta u + \sum_i g_i(u)(E_i - u)\,, \tag{1.2.1}$$

a second-order parabolic PDE, where Δ is the Laplacian and $\cdot$ denotes differentiation with respect to time. If we assume that the potential is homogeneous over the spatial domain, $\Delta u \equiv 0$, which we refer to as a space-clamped membrane in reference to the experimental technique used by Hodgkin and Huxley, then the cable equation reduces to the membrane balance equation

$$C\dot{u} = \sum_i g_i(u)(E_i - u)\,, \tag{1.2.2}$$

an ordinary differential equation. In both of these equations each summand $g_i(u)(E_i - u)$ corresponds to a different current source with $g_i(u)$ denoting its conductance. Current sources are in particular single or families of ion channels in open, i.e.,



(a) (b)

Figure 1.1: Markov kinetic scheme of (a) potassium and (b) sodium ion channels in the classical Hodgkin-Huxley model of a squid giant axon. The states n_4 and m_3h_1 are the conducting states of the potassium and sodium channel, respectively. The rates $a_x = a_x(u),\ b_x = b_x(u)$ are dependent on the transmembrane potential u.

conducting, states. The conductances $g_i(u)$ depend on an associated model for the channel gating. The stochastic, microscopic modelling of the conductances based on individual channel dynamics gives rise to the second *stochastic* building block. This is in contrast to macroscopic, deterministic models of the conductances where a coupled system of ODEs gives the second building block, cf. Section 3.3.

Single, stochastic ion channels subject to constant transmembrane potential are modelled on a mesoscopic scale of magnitude by continuous-time, discrete-state Markov chains [54] which are most easily described by a state diagram, cf. Fig. 1.1. Therein, e.g., $m_2h_1 \rightharpoonup m_3h_1$ with transition rate a_m indicates that a channel which is in state m_2h_1 changes its state to m_3h_1 at a random time τ, the distribution of which is given by

$$\mathbb{P}[\tau > t] = \exp\big(-a_m t\big), \tag{1.2.3}$$

i.e., an exponential distribution with parameter a_m. An independent collection of ion channels can be simulated using the stochastic simulation algorithm (SSA) [53].

In the next step we combine these building blocks to model a space-clamped membrane, that is, we have to combine equation (1.2.2) and Markov chain models for single channels, where now the time-evolution of the former is coupled to the latter and vice versa. Then, on the one hand, for the conductances in (1.2.2) the dependence of the time-evolution of u on the ion channels is modelled by defining $g_i(u)$ to be the number of open channels n_{state} from the respective family of channels times the conductance of a single channel denoted $\overline{g}_i$. For example, for the current due to sodium (Na) channels defined in Fig. 1.1 we obtain

$$g_{\text{Na}}(u) \equiv \overline{g}_{\text{Na}} n_{m_3h_1}. \tag{1.2.4}$$

On the other hand, the transition rates of the channels now also depend on the evolution of the transmembrane potential u. Thus incorporating this dependence yields the mathematically exact waiting time distribution

$$\mathbb{P}[\tau > t] = \exp\Big(-\int_0^t a_m(u(s))\,\mathrm{d}s\Big) \tag{1.2.5}$$

for the example considered in (1.2.3). Such a model with the time-evolution coupled in this way is what we consider and refer to as an *exact model* for a (space-clamped) membrane.

In Chapter 2 when discussing piecewise deterministic processes it will become apparent that the building blocks of stochastic hybrid ODE models, (1.2.2) and Fig. 1.1, and their coupling via (1.2.4) and (1.2.5), exactly fit the components of finite-dimensional PDMPs. Moreover, by considering PDMPs on infinite-dimensional state spaces we can cast models of spatially extended system, i.e., considering equations (1.2.1) instead of (1.2.2), into a PDMP framework. The coupling of the building blocks for spatial dynamics, in other words, the modelling of the conductances $g_i(u)$ in the PDE (1.2.1), is discussed in Section 3.1.

To conclude this introduction we briefly revisit the class of *pseudo-exact models*, which contains most hybrid models used in the literature, and comment on their connection to exact models. To define pseudo-exact models an ad-hoc approximation concerning the distribution (1.2.5) is employed, see [31, 38, 110]. This approximation is based on the assumption that the transition rates $a_x(u)$, $b_x(u)$ do not change significantly during the time interval between successive state changes in the channel configuration and thus can be considered constant on that interval. Using (1.2.3) instead of (1.2.5) for the distribution of the inter-jump intervals pseudo-exact algorithms become straight-forward to implement and essentially reduce to a version of the SSA. These pseudo-exact algorithms have been used in a vast number of studies, e.g., for the simulation of the transmembrane potential at Nodes of Ranvier in auditory nerves [63, 94]. An exact algorithm has also been proposed in [32]. But to the best of our knowledge this algorithm has only been used in the four studies [12, 56, 88, 105], as it is computationally very expensive compared to pseudo-exact algorithms [93]. Computer experiments comparing simulation statistics for a simple test model suggest that only a reasonable error seems to be introduced when using pseudo-exact algorithms [93]. However, as already mentioned, for the general case a theoretical analysis of the error of this approximation has yet to be performed. As a necessary prerequisite it needs a mathematical framework of the stochastic processes it approximates. Concerning the exact algorithm we have mentioned that it lacks a rigorous justification, as a strong Markov property for stochastic hybrid model is employed which a-priori is not necessarily given. Anticipating a description of exact models by PDMPs we note that this exact algorithm is essentially equivalent to a theoretical method for the exact simulation of finite-dimensional PDMPs which we discuss in Chapter 5.

In that chapter we also present a numerical error analysis for a large class of actual implementations of this theoretical exact method as in practice an exact simulation is rarely possible.

1.3 Outline of the thesis

A broad outline of the thesis is given by the following. In Chapter 2 we present the general theory of PDMPs. We note that a summary of necessary preliminary results to PDMP theory which builds on the theory of random counting measures is presented in Appendix B. The initial Section 2.0.1 of Chapter 2 contains a brief introduction to Markov processes wherein we recall some central aspects of Markov process theory. We define general PDMPs in Section 2.1 and prove the strong Markov property for PDMPs. In Section 2.1.2 we discuss some special cases of PDMPs and derive a characterisation of the extended generator for a large class of PDMPs in Section 2.2.

In Chapter 3 we are concerned with stochastic hybrid models of neuronal membranes given by PDMPs. We first discuss in Section 3.0.1 in more detail the appropriate family of PDMPs which is employed. Then Section 3.1 exemplifies on the Hodgkin-Huxley model how models of excitable membranes are cast into the PDMP framework. We discuss how this approach can be used for more general excitable systems, cf. Section 3.1.1. Further, we also show how spatial models reduce to PDMP models as considered in [96] when they are spatially homogeneous, cf. Section 3.1.2. A second type of PDMP models of spatial neuronal membranes is presented in Section 3.2. Finally, this chapter also contains in Section 3.3 a general existence and uniqueness theorem for deterministic systems of excitable membranes which is a necessary reference for the subsequent chapter.

Next, in Chapter 4 we consider limit theorems for sequences of PDMPs of the form used in spatio-temporal models of excitable membranes. Certain martingales that are associated with PDMPs play a central role for the study of limit theorems. Therefore we present a discussion of these martingales in Section 4.1. Then, the law of large numbers is presented in Section 4.2 which establishes convergence in probability to the solution of a deterministic system of abstract evolution equations. Subsequently martingale central limit theorems are presented in Section 4.3 and the derivation of the Langevin approximation is discussed in Section 4.3.2. We apply these results to compartmental-type models of excitable membranes discussed in Section 3.2. Here the deterministic limit arising in the law of large numbers is the solution to a general excitable membrane system of the type discussed in Section 3.3.

Chapter 5 deals with numerical approximations to PDMPs in Euclidean space. We first present in Section 5.1 a theoretical simulation algorithm which takes the role of an 'exact solution' in numerical analysis. In Section 5.2 we suitably discretise this theoretical algorithm and obtain a class of numerical simulation methods based on continuous ODE methods (methods with dense output). This section also contains the central result which gives conditions that guarantee pathwise convergence of the numerical methods and establishes an asymptotic order of convergence. Then in Section 5.4 we extend this convergence result to a more general class of PDMPs. Finally, in Section we present examples of applications of the numerical methods we presented and analysed to models of neuronal membranes.

The thesis is concluded in Chapter 6 wherein we briefly recapitulate the results we have established and discuss directions for further research motivated by these. Finally, Appendix A contains the precise definition of rate functions and parameter values for particular neuron models employed in this thesis. As already mentioned Appendix B contains an account of theory of random counting measures which is a prerequisite of PDMP theory. Appendix C presents a brief review of existence theory for linear parabolic partial differential equations. It states some classical results that are frequently employed in the present thesis.

Chapter 2

Piecewise Deterministic Markov Processes

In this chapter we present a concise introduction to the general theory of *piecewise deterministic Markov processes* (PDMPs). This class of stochastic processes was introduced by Davis [36, 37] in 1984. Davis' original approach is very much related to switching systems as are the applications that motivated his study of PDMPs. This relation is particularly expressed in the specific structure of the processes he discusses which incorporate a piecewise constant component that jumps among isolated states. Also the discussion is restricted to $\mathbb{R}^d$-valued, homogeneous processes. A recent monograph on PDMPs by Jacobsen [64] overcomes these restrictions to some extent. Therein PDMPs are considered inhomogeneous Markovian jump processes taking values in Borel measurable spaces, cf. Section B.1, that move deterministically in between random jump times. In general, Jacobsen's approach is much more in the sense of general càdlàg Markov processes and more accessible for an analysis. This is a resemblance to differential equations with Markovian switching – a special case of PDMPs – which for analytical purposes are represented as solutions to jump-diffusion equations, cf., e.g., [86] in a more general context. However, a drawback in Jacobsen's presentation is that random initial conditions for PDMPs are neglected and, although a general existence result is stated, with regards to further theory and analysis the class of PDMPs is as in Davis [36, 37] restricted to $\mathbb{R}^d$-valued processes. For the application of PDMPs to model spatially extended excitable membranes, which we pursue in the present thesis, this restricted class of PDMPs is not sufficient as– in anticipation of Chapter 3 – infinite-dimensional function spaces are the appropriate state spaces for PDMPs in this context. We note that further relevant, early publications concerning PDMPs are [29] and [68] which discuss more general classes of piecewise deterministic processes in contrast to Davis allowing for processes that are non-regular, i.e., countably many jumps in a finite time interval may occur. In particular, the authors in [68] present a completely different, complementary approach to

PDMPs. Their approach allows the derivation of necessary conditions for a process to be piecewise deterministic and Markov whereas authors in [36, 37, 64] only present sufficient conditions.

In our presentation of PDMPs in Section 2.1 we follow for the existence results the presentation of Jacobsen [64], but most importantly extend the results to allow for random initial conditions. We note that the existence theory is largely based on the theory of random counting measures a brief account of which we present in the Appendix B.2. The constructive approach to PDMPs, as employed by Davis [37], is extended in Section 2.1.1 to the general class of PDMPs we consider. It is of particular importance for numerical analysis of simulation methods for PDMPs, cf. Chapter 5. Finally, in Section 2.2 we present as a main result a precise characterisation of the extended generator for a general class of PDMPs which extends the corresponding results of [37, 64]. The proofs of the two main results in this chapter, the strong Markov property of a PDMP and the characterisation of the extended generator of a PDMP, are, together with two minor helpful results, deferred to the end of the chapter, Section 2.3

Before we proceed to the precise definition and discussion of PDMPs we briefly recall in Section 2.0.1 some general theory concerning Markov processes. For a general theory of Markov processes we refer to the monographs [20, 41, 109] on which the presentation in the subsequent section is based but the essentials can also be found in [37, 64]. For a more thorough presentation of the connections of semigroups and Markov processes we refer to Ethier and Kurtz [41] wherein also the martingale / local martingale problem is discussed in detail.

2.0.1 Markov processes

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ be a filtered probability space and $(X_t)_{t\geq 0}$ be a càdlàg[1] process defined thereon taking values in a Borel measurable space $(E, \mathcal{E})$. Further, $(\mathcal{F}_t^X)_{t\geq 0}$ denotes the natural filtration generated by a process $(X_t)_{t\geq 0}$, i.e., $\mathcal{F}_t^X = \sigma(X_s; s \leq t)$ for all $t \geq 0$. Clearly, here right-continuity and the existence of left limits for the paths are understood in the sense of metric spaces, i.e., $\mathbb{R}_+$ equipped with the usual Euclidean metric and for the Borel space E we choose any metric compatible with the topology. In view of applications this slight ambiguity in the continuity definition

[1] *càdlàg* is an abbreviation for 'continu à droite, limité à gauche', so these are continuous-time stochastic processes with sample paths that are almost surely everywhere right continuous with limits from the left existing everywhere. We note that the càdlàg property of a stochastic process is not essential for the theory of Markov processes. However, it simplifies some definitions, e.g., we can omit conditions for processes being progressively measurable (with respect to a given filtration) as every right-continuous, adapted process is progressively measurable. A second condition regularly employed is for general processes to be optional (with respect to a given filtration), e.g., [66, 68]. Again, this condition can be omitted as, by definition, cádlág adapted processes are optional. Furthermore, all processes considered in this thesis are càdlàg.

does not arise as E is always a (Borel subset of a) normed vector space and continuity holds with respect to the metric induced by the norm.

Definition 2.0.1. (a) The process $(X_t)_{t\geq 0}$ is called a *Markov process with respect to the filtration* $(\mathcal{F}_t)_{t\geq 0}$ if there exist Markov kernels p_{st} on E for all $s \leq t$ such that for all $C \in \mathcal{E}$

$$\mathbb{P}[X_t \in C|\mathcal{F}_s] = p_{st}(X_s, C) \quad \text{a.s.} \tag{2.0.1}$$

A Markov process is called *homogeneous* (or *with stationary transition probabilities*) if in addition one can choose p_{st} such that it depends on t, s only via the difference $t - s$. In this case (2.0.1) reads for all $0 \leq s \leq t$ and all $C \in \mathcal{E}$

$$\mathbb{P}[X_t \in C|\mathcal{F}_s] = p_{t-s}(X_s, C) \quad \text{a.s.} \tag{2.0.2}$$

In these definitions we always take for the Markov kernels $p_{ss}(x, \cdot) = \delta_x$ or $p_0(x, \cdot) = \delta_x$, respectively, where δ_x denotes the Dirac measure at x.

(b) Markov kernels p_{st}, p_t are also called the *transition functions* of the process. A family of transition functions p_{st}, p_t satisfies the *Chapman-Kolmogorov equations* if for all $s \leq t \leq u$ and all $x \in E$, $C \in \mathcal{E}$

$$p_{su}(x, C) = \int_E p_{tu}(y, C)\, p_{st}(x, \mathrm{d}y), \tag{2.0.3}$$

or, in the homogeneous case, if for all s, t and all $x \in E, C \in \mathcal{E}$

$$p_{s+t}(x, C) = \int_E p_t(y, C)\, p_s(x, \mathrm{d}y)\,. \tag{2.0.4}$$

In particular, the Chapman-Kolmogorov equations are for every $s \geq 0$ and all $C \in \mathcal{E}$ satisfied for $x = X_s$ by the transition functions of a Markov process almost surely. Further, two Markov processes are called *equivalent* if they have the same transition functions.

(c) A Markov process with respect to $(\mathcal{F}_t)_{t\geq 0}$ with transition functions p_{st} or p_t in the homogeneous case satisfies the *strong Markov property* if for all $\mathcal{F}_t$-stopping times τ on the set $[\tau < \infty]$ it holds that for all $C \in \mathcal{E}$

$$\mathbb{P}[X_{\tau+t} \in C|\mathcal{F}_\tau] = p_{\tau,\tau+t}(X_\tau, C) \quad \text{a.s.} \tag{2.0.5}$$

or in the homogeneous case

$$\mathbb{P}[X_{\tau+t} \in C|\mathcal{F}_\tau] = p_t(X_\tau, C) \quad \text{a.s.} \tag{2.0.6}$$

(d) The marginal probability of the initial condition X_0, i.e., the probability measure

ν on E given by $\nu(C) = \mathbb{P}[X_0 \in C]$, is called the *initial law* of the Markov process. Note that equivalent Markov processes which, in addition, possess the same initial law, are versions of each other.[2]

We next comment briefly on the definition of Markov processes presented above. Clearly, if a process is a Markov process with respect to some filtration, then it is also Markov process with respect to any coarser filtration. In particular, every Markov process satisfies the Markov property with respect to its natural filtration. The opposite is in general not true, that is, a Markov process need not be Markov with respect to a finer filtration. Further, there are several equivalent definitions in the literature for a stochastic process to be a Markov process. E.g., Ethier and Kurtz [41] define a Markov process by the property that for all $C \in \mathcal{E}$ the conditional probabilities satisfy

$$\mathbb{P}[X_t \in C \,|\, \mathcal{F}_s] = \mathbb{P}[X_t \in C \,|\, X_s] \quad \text{a.s.} \tag{2.0.7}$$

A second often used definition, e.g., Davis [37], is that for all bounded, measurable real-valued functions $f : E \to \mathbb{R}$ it holds for all $0 \leq s \leq t$

$$\mathbb{E}[f(X_t) \,|\, \mathcal{F}_s] = \mathbb{E}[f(X_t) \,|\, X_s] \quad \text{a.s.} \tag{2.0.8}$$

This last property (2.0.8) is a special case of the generalisation of the Markov property presented in Proposition 2.0.1, which corresponds to the definition of a Markov process in Sharpe [109, p. 3]. Finally, a process $(X_t)_{t\geq 0}$ is Markov if for a family of transition functions it holds that for all bounded, measurable, real-valued f and all $0 \leq s \leq t$ it holds that

$$\mathbb{E}[f(X_t)|\mathcal{F}_s] = \int_E f(y)\, p_{st}(X_s, \mathrm{d}y) \quad \text{a.s.} \tag{2.0.9}$$

It is easily seen that the defining conditions (2.0.1) and (2.0.7) – (2.0.9) are equivalent. Firstly, we find that, on the one hand, (2.0.7) implies (2.0.8) and (2.0.1) implies (2.0.9) due to the monotone convergence theorem [20, 41]. On the other hand, (2.0.7) is a special case of (2.0.8) and (2.0.1) is a special case of (2.0.9). Secondly, as the right hand sides of (2.0.1) and (2.0.9) are measurable with respect to $\sigma(X_s)$ it follows immediately that (2.0.1) implies (2.0.7) and (2.0.9) implies (2.0.8). Finally, (2.0.7) implies (2.0.1) and (2.0.8) implies (2.0.9) as E is a Borel measurable space and thus the existence of Markov kernels p_{st} is guaranteed as regular conditional probabilities always exist. Further, the properties of the conditional distribution imply that the Markov kernels have to satisfy the Chapman-Kolmogorov equations almost surely.

[2] We recall, that two stochastic processes, which need not be defined on the same probability space, are *versions* of each other, if they have the same finite-dimensional distributions. For Markov processes the finite-dimensional distributions are given uniquely by the initial law and the transition functions, see, e.g., [41].

For a more detailed explanation of these arguments we refer to [20, 109].
To conclude, we note that the property (2.0.1) or (2.0.2), respectively, does not necessarily define the Markov kernels p_{st}, p_t uniquely. Nevertheless, as E is a Borel space it is guaranteed that two Markov kernels both satisfying the Markov property define the same probability on E for all x outside of a null set for X_s, cf. [20, p. 15].

Proposition 2.0.1. *$(X_t)_{t\geq 0}$ is a Markov process with respect to $(\mathcal{F}_t)_{t\geq 0}$ if and only if for all $\sigma(X_u, u \geq t)$-measurable, bounded, real-valued random variables U it holds that*

$$\mathbb{E}[U|\mathcal{F}_t] = \mathbb{E}[U|X_t] \quad \text{a.s.}$$

In Sharpe [109, p. 3] such a condition as in Proposition 2.0.1 but for positive, $\sigma(X_u, u \geq t)$-measurable random variables U is stated as a definition of the Markov property. Then, the Proposition 2.0.1 is inferred via a decomposition of bounded random variables into positive and negative part. We further note, that the sufficient part of Proposition 2.0.1, i.e., the condition of the statement implies the Markov property, follows by (2.0.8). This holds as for every bounded, measurable f the random variable $f(X_t)$ is of the form assumed for U. The necessary part of the statement is proven in Jacobsen [64, p. 145].

In the remainder of this section we restrict the presentation to homogeneous Markov processes as this is sufficient for the application of the results following in later sections. Furthermore, it is always possible to transform an inhomogeneous Markov process $(X_t)_{t\geq 0}$ via an extension of the state space into a homogeneous process. This is accomplished considering the associated *space-time process* $(Z_t)_{t\geq 0}$. The process $(Z_t)_{t\geq 0}$ is defined by $Z_t := (t, X_t)$ for all $t \geq 0$.
Hence, first of all, note that the space-time process it is adapted to the same filtration as $(X_t)_{t\geq 0}$ and it satisfies for any Borel set $C \subseteq [0,\infty) \times E$ that for all $s, t, \geq 0$

$$\mathbb{P}[Z_{s+t} \in C|\mathcal{F}_s] = \widehat{p}_t(Z_s, C)\,,$$

where $\widehat{p}_t((s,x), C) := p_{s,s+t}(x, \mathrm{proj}_E(\{t+s\} \times E \cap C))$. Here proj_E denotes the projection of a subset of $[0,\infty) \times E$ on its E-component. One the one hand, $\widehat{p}_t$ is measurable[3] with respect to (s,x) for fixed C and, on the other hand, for fixed (s,x) it defines a probability on $[0,\infty) \times E$. Hence, by Definition 2.0.1(iii) the space-time process is a homogeneous Markov process. It follows immediately that the space-time process is a strong Markov process in case $(X_t)_{t\geq 0}$ satisfies the strong Markov property. Within the general theory of Markov processes space-time processes are

[3] This can be shown by a Dynkin class argument (see the proof of Lemma 2.3.2 for such a method of proof) or we simply tacitly assume that the transition functions p_{st} are jointly measurable as mappings $(s,t,x) \mapsto p_{st}(x, C)$ for any fixed Borel set C, cf. [41] wherein this property is part of the definition of a transition function.

considered in [109, Sect. 16] where the time component is not (fully) deterministic but performs a uniform motion to the right from a random initial condition, i.e., $Z_0 \in [0, \infty) \times E$ instead of $Z_0 \in \{0\} \times E$ as is sufficient for our purposes. In this general case the definition of the process via a transition function follows immediately from the definition of $\widehat{p}_t$.

We repeat and emphasise that *for the remainder of this section* $(X_t)_{t\geq 0}$ *always denotes a homogeneous Markov process*. An important object in the study of homogeneous Markov processes is the *associated semigroup* of operators on the set of bounded, measurable real-valued functions on E, which we denote in the following by $B(E, \mathbb{R})$. This function space is a Banach space when equipped with the supremum norm, denoted by $\|\cdot\|_0$. A family of operators $(P_t)_{t\geq 0}$ forms a semigroup if each operator P_t maps $B(E, \mathbb{R})$ on itself and, in addition, they satisfy the functional relations

$$P_0 = \mathrm{id} \quad \text{and} \quad P_{t+s} = P_s \circ P_t \quad \forall\, t, s \geq 0\,.$$

The semigroup is *strongly continuous* if it is continuous (from the right) at the origin, i.e., if for all $f \in B(E, \mathbb{R})$ it holds that

$$\lim_{t \downarrow 0} \|P_t f - f\|_0 = 0\,.$$

Moreover, a strongly continuous semigroup has continuous paths, i.e., $t \mapsto P_t f$ is continuous for all $f \in B(E, \mathbb{R})$, cf. [103, p. 397]. It is proven in [37] that the set $B_0(E, \mathbb{R}) \subset B(E, \mathbb{R})$ on which any semigroup is strongly continuous is a closed subspace of $B(E, \mathbb{R})$, hence itself a Banach space, and the semigroup P_t maps $B_0(E, \mathbb{R})$ onto $B_0(E, \mathbb{R})$. Finally, a semigroup is a *contraction semigroup* if $\|P_t\|_0 \leq 1$ for all $t \geq 0$, where $\|\cdot\|_0$ denotes the induced operator norm, and the semigroup is measurable if the map $t \mapsto P_t f$ is measurable for all $f \in B(E, \mathbb{R})$.

The connection to Markov processes is established by the following observation. Due to the Chapman-Kolmogorov equations (2.0.4) it follows that the operators P_t, $t \geq 0$ defined by

$$P_t f(x) := \mathbb{E}[f(X_t)|X_0 = x] = \int_E f(y)\, p_t(x, \mathrm{d}y) \quad \forall\, x \geq E \tag{2.0.10}$$

define a measurable contraction semigroup of linear, bounded operators. More generally we define the following.

Definition 2.0.2. Let $(P_t)_{t\geq 0}$ be an arbitrary semigroup defined on a closed subspace $L \subset B(E, \mathbb{R})$. We say that *the Markov process* $(X_t)_{t\geq 0}$ *corresponds to the semigroup* $(P_t)_{t\geq 0}$ or *the semigroup is associated to the Markov process* if for all $t, s \geq 0$ and all

$f \in L$

$$\mathbb{E}[f(X_{s+t})|\mathcal{F}_s^X] = P_t f(X_s). \tag{2.0.11}$$

Obviously, if the semigroup $(P_t)_{t\geq 0}$ is defined by (2.0.10) than condition (2.0.11) is just (2.0.9) and thus satisfied for all $f \in B(E, \mathbb{R})$. The importance of the associated semigroup lies in the fact that it defines a Markov process uniquely as finite-dimensional distributions define a stochastic process uniquely.

Proposition 2.0.2. ([41, Chap. 3, Prop. 1.6]) *Let $(X_t)_{t\geq 0}$ be a Markov process taking values in E with given initial law ν. The process corresponds to a semigroup on a closed subspace L of $B(E, \mathbb{R})$. If L is separating*[4]*, then the semigroup and the initial law determine the finite-dimensional distributions of the process.*

With the transition semigroup of a Markov process we can associate its full generator, which is called the *generator of the Markov process*. In general, the *full generator* of a measurable contraction semigroup is given by the set

$$\mathcal{D}(\widehat{\mathcal{A}}) := \left\{ (f, g) \in B(E, \mathbb{R})^2 \,\middle|\, P_t f - f = \int_0^t P_s g \,\mathrm{d}s \right\}. \tag{2.0.12}$$

The full generator is in general not single-valued. Here, *single-valued* means heuristically that for each f there exists only one g such that the pair (f, g) is in the full generator. Precisely, the full generator is single-valued if $(0, g) \in \mathcal{D}(\widehat{\mathcal{A}})$ implies that $g = 0$. In this case we may understand $\widehat{\mathcal{A}}$ as an operator on $B(E, \mathbb{R})$ mapping f to the unique function g such that $(f, g) \in \mathcal{D}(\widehat{\mathcal{A}})$ and write for a pair $(f, g) = (f, \widehat{\mathcal{A}} f)$. Then, in a slight abuse of notation $\mathcal{D}(\widehat{\mathcal{A}})$ is understood as the domain of the operator $\widehat{\mathcal{A}}$. Additionally we have associated to any semigroup its *infinitesimal strong generator* $\overline{\mathcal{A}}$. Heuristically, the strong generator is the derivative at $t = 0$. Denote by $\mathcal{D}(\overline{\mathcal{A}})$ the set of all bounded measurable functionals f for which the limit $\overline{\mathcal{A}} f$ in the following sense exists

$$\lim_{t \downarrow 0} \left\| \frac{P_t f - f}{t} - \overline{\mathcal{A}} f \right\|_0 = 0. \tag{2.0.13}$$

Thus $\overline{\mathcal{A}} f$ is the Fréchet derivative (from the right) of the map $t \mapsto P_t f$ between the Banach spaces $\mathbb{R}_+$ and $B(E, \mathbb{R})$ and it holds that $\mathcal{D}(\overline{\mathcal{A}}) \subset B_0(E, \mathbb{R})$. Moreover, the infinitesimal generator of a strongly continuous semigroup is contained in the full generator, cf. [41, Chap. 1, Prop. 1.5].

The importance of the generator of a Markov process is contained in the observation that its elements have the following property.

[4] A subset $L \subseteq B(E, \mathbb{R})$ of bounded, real-valued functions on E is called *separating* if for two probabilities $\mathbb{P}^1$, $\mathbb{P}^2$ on E the property

$$\int_E f \,\mathrm{d}\mathbb{P}^1 = \int_E f \,\mathrm{d}\mathbb{P}^2 \quad \forall f \in L$$

implies that $\mathbb{P}^1 = \mathbb{P}^2$, cf., e.g., [14, 37, 41].

Proposition 2.0.3. ([41, Chap. 3, Prop. 1.7]) *Let $(X_t)_{t\geq 0}$ be a Markov process taking values in E with transition functions p_t. Further, let $(P_t)_{t\geq 0}$ be a corresponding semigroup and $\widehat{\mathcal{A}}$ its full generator. Then for $(f,g) \in \mathcal{D}(\widehat{\mathcal{A}})$ the process defined by*

$$t \mapsto M_t := f(X_t) - f(X_0) - \int_0^t g(X_s)\,\mathrm{d}s \tag{2.0.14}$$

is an $\mathcal{F}_t^X$-martingale.[5] *Further, the function g is uniquely defined up to sets of zero potential*[6] *and, in particular, if (f,g) is in the infinitesimal generator, then $g = \overline{\mathcal{A}}f$.*

Since the finite dimensional distributions of a Markov process, which uniquely characterise a process, are determined by a corresponding semigroup due to Proposition 2.0.2 and the semigroup in turn is determined by its full generator $\mathcal{D}(\widehat{\mathcal{A}})$ or a sufficiently large set $A \subset \mathcal{D}(\widehat{\mathcal{A}})$, cf. Hille-Yosida Theorem, see, e.g. [41], it follows that a Markov process is uniquely defined by a sufficiently large subset of the full generator of a semigroup. The mathematical task of characterising a Markov process to a given set of pairs of functions (f,g) such that the process (2.0.14) is a martingale is called the *martingale problem*. We say a stochastic process or, equivalently, a probability measure on the canonical space of càdlàg functions is a *solution to the martingale problem* posed by a set of pairs (f,g) if the process (2.0.14) is a martingale for all such pairs. For a discussion of the martingale problem and in particular existence and uniqueness of solutions we refer to [41, 113].
To conclude this introduction to Markov processes, we note that there may be some measurable functions f not in the domain of the generator for which a property akin to Proposition 2.0.14 holds true. This leads to the following definition.

Definition 2.0.3. Let $\mathcal{D}(\mathcal{A})$ denote the set of all real-valued measurable functions f on E such that there exists a real-valued measurable function g on E with the property that $t \mapsto g(X_t)$ is almost surely integrable and the process

$$t \mapsto M_t := f(X_t) - f(X_0) - \int_0^t g(X_s)\,\mathrm{d}s$$

is an $\mathcal{F}_t^X$-local martingale. If such a g exists it is unique up to sets of zero potential and thus we may write $\mathcal{A}f := g$. We call $\mathcal{A}$ the *extended generator* of the process

[5] We note that in [41, Chap. 3, Prop. 1.7] the martingale M_t is defined without subtracting $f(X_0)$ in the right hand side of (2.0.14). However, the definition we employ is equivalent as $f(X_0)$ is only an addition of an $\mathcal{F}_0^X$-measurable function, i.e., $(M_t + f(X_0))_{t\geq 0}$ is a martingale if and only if $(M_t)_{t\geq 0}$ is a martingale. The advantage of definition (2.0.14) is that the process $(M_t)_{t\geq 0}$ satisfies $M_t = 0$.

[6] A Borel set $A \subset E$ is said to be of *zero potential* if for every $x \in E$

$$\mathbb{P}^{0,x}\Big[\int_0^\infty \mathbb{I}_A(X_s)\,\mathrm{d}s = 0\Big] = 1$$

where the probability $\mathbb{P}^{0,x}$ is such that $\mathbb{P}^{0,x}[X_0 = x] = 1$. This means that the process 'spends' almost surely no time in A regardless of the initial condition.

$(X_t)_{t\geq 0}$ and $\mathcal{D}(\mathcal{A})$ its domain.

The advantage of the extended generator over the full generator is that it is usually easier to specify its domain exactly. Further, as every martingale is a local martingale it holds that $\mathcal{D}(\widehat{\mathcal{A}}) \subset \mathcal{D}(\mathcal{A})$ and the extended generator does 'extend' the full generator. Akin to the martingale problem also the extended generator can be used to characterise a Markov process uniquely which is called the *local martingale problem*, cf. [41, p. 224] or the monograph [66]. A solution to the local martingale problem is understood analogously to solutions to martingale problems. We remark that in Chapter 4 we employ the local martingale problem to uniquely characterise the limit in distribution of a sequence of martingales.

2.1 Definition of PDMPs

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ be a filtered probability space with right-continuous filtration and $(E, \mathcal{E})$ be a Borel measurable space. Further we have that μ is an adapted, regular[7] random counting measure defined on this probability space as defined in Appendix B.2. We refer to Appendix B.2 for an introduction of the notation used in connection with random counting measures throughout the subsequent sections. The random counting measure μ also defines a marked point process $(\tau_n, Y_n)_{n\geq 1}$ and an adapted counting process $(N_t)_{t\geq 0}$, cf. Section B.2.7. That is, $N_t = \mu((0,t], E)$ counts the number of events up to and including time t. In particular, one should keep in mind that μ is an $(H, \mathcal{H})$-valued random variable directly defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and all other random variables, i.e., τ_n, Y_n, $n \geq 1$, N_t, $t \geq 0$ as well as the PDMP defined subsequently, are defined as measurable mappings of μ. Furthermore, the probability space supports an E-valued random variable X_0. We generally assume that the set of events Ω is sufficiently rich, by which is understood that Ω contains enough events to exhaust the state space of random variables defined thereon and thus all information on the random variables is purely contained in their law. Finally, without loss of generality we may assume that $\mathcal{F} = \sigma(X_0) \vee \mathcal{F}^\mu_\infty$, i.e., the σ-field on Ω is generated by the random variables X_0 and μ. It is an almost trivial extension to consider σ-fields $\mathcal{F}$ are generated by which X_0 and μ additional random variables independent of these. Then the filtration $(\mathcal{F}_t)_{t\geq 0}$ is given by $\mathcal{F}_t := \sigma(X_0) \vee \mathcal{F}^\mu_t$ for all $t \geq 0$ which is of the form as discussed in Section B.2.7 in connection with random counting measures on arbitrary probability spaces.

Definition 2.1.1. We call a stochastic process $(X_t)_{t\geq 0}$ *piecewise deterministic* if it can be constructed from a marked point process in the following way:

[7]We use the attribute *regular* to emphasise the fact that the random counting measure has only finitely many events in any finite time interval almost surely, which holds for a canonical random counting measure by definition, see Section B.2.1.

(i) There exists an $\mathcal{F}_0$-measurable initial condition X_0 taking values in E,

(ii) for every $n \geq 0$ there exists a measurable function

$$f^n_{x_0} : K^n \times [0,\infty)\big|_{t_n \leq t,\, t_n < \infty} \to E : (z_n, t) \mapsto f^n_{x_0}(\, t\,|z_n),$$

which may depend on the initial condition $x_0 \in E$, such that the process $(X_t)_{t\geq 0}$ satisfies on the set $[X_0 = x_0]$

$$X_t = f^{N_t}_{X_0}(\, t\,|Z_{N_t})\,, \tag{2.1.1}$$

where $Z_{N_t} = \big((\tau_1, Y_1), \ldots, (\tau_{N_t}, Y_{N_t})\big)$ according to the definition in (B.2.2). Further, we always assume that the functions f are càdlàg[8] and satisfy for all $z_n \in (\mathbb{R}_+ \times E)^n$, $n \geq 0$, the boundary conditions

$$f^0_{x_0}(0) = x_0 \quad \text{and} \quad f^n_{x_0}(\, t_n\,|z_n) = y_n\,.$$

This corresponds to the fact that the marks Y_n give the position of the PDP after the nth jump. Other choices of the interpretations of the marks are also possible, e.g., if E is a vector space one may choose to identify via $Y_n = X_{\tau_n} - X_{\tau_n -}$, i.e., the size of the jump that a process performs at the jump time τ_n. Finally, a piecewise deterministic process is called *piecewise continuous* if $t \mapsto f^n_{x_0}(\, t\,|z_n)$ is continuous on $[t_n, \infty)$ for all $n \geq 0$, $z_n \in K^n$.

Remark 2.1.1. We note that in this thesis by definition a PDP is always a *regular* jump process. Nevertheless, we frequently use the attribute 'regular' in connections with PDPs to emphasise this important fact. Clearly, it is possible to define PDPs also from exploding random counting measures, however, such PDPs cannot constitute a Markov process in the sense of Section 2.0.1 and the additional concept of finite life-times of Markov processes has to be introduced. This class of processes does not occur for the models we are considered with in this thesis, hence, we do not discuss this issue any further.

Clearly, the law of a PDP $(X_t)_{t\geq 0}$ is determined by the joint distribution of μ and the initial condition X_0 via the pushforward measure and $\mathcal{F}^X_t \subset \mathcal{F}_t$ for all $t \geq 0$. However, the possible dependence of the law of the random counting measure on the initial condition X_0 complicates matters for the existence of such a process compared to the presentation in [64] wherein only constant initial conditions are considered. Moreover, the existence theorem of random counting measure in Appendix B.2 only covers laws independent of the initial condition. Thus, in order to guarantee the well-definedness of general PDPs with arbitrary initial law we first have to prove that

[8] Here the càdlàg property is understood with respect to a suitable metric on E which exists as every Borel space is metrizable.

there exists probability measure on $(\Omega, \mathcal{F})$ possessing the given marginal for X_0 and the marginal with respect to μ being a canonical random counting measure. That is,

(W) for every initial law ν on E there exists a probability $\mathbb{P}^\nu$ on $(\Omega, \mathcal{F})$ such that $\mathbb{P}^\nu[X_0 \in C] = \nu(C)$ for all $C \in \mathcal{E}$ and $\mathbb{P}^\nu[\mu \in D | X_0] = \mathbb{P}^{0,X_0}(D)$ for all $D \in \mathcal{H}$ where $\mathbb{P}^{0,x}$, $x \in E$, is a probability on $(H, \mathcal{H})$ defined from Markov kernels $(P_x^n)_{n\geq 0}$ and $(\pi_x^n)_{n\geq 0}$ due to Theorem B.2.1.

Now, given $\mathbb{P}^\nu$ as in (W) the law of a PDP is defined via the pushforward measure. Thus, the problem (W) poses is to construct a joint probability distribution for (X_0, μ) from given marginals. We show in the following that this is always possible in the current context.

Given the families of Markov kernels $(P_x^n)_{n\geq 0}$ and $(\pi_x^n)_{n\geq 0}$ we argue, in a first step, that a probability as in (W) exists for initial conditions given by the Dirac measure at any point $x \in E$: Clearly, there exists a probability on $(\Omega, \mathcal{F})$ such that $\mathbb{P}^{0,x}$ is the marginal of the random counting measure μ and the initial condition satisfies $X_0 = x$ almost surely. We denote such a probability on $(\Omega, \mathcal{F})$ also by $\mathbb{P}^{0,x}$. That the law of a PDP under this probability is well-defined, i.e., that the marginals fully define the joint distribution of (X_0, μ) is implied by Lemma 2.3.1 in Section 2.3.1.
In a next step we use the family of probabilities $(\mathbb{P}^{0,x})_{x\in E}$ to define a probability as in (W) for any law ν on E by

$$\mathbb{P}^\nu(A) := \int_E \mathbb{P}^{0,x}(A)\, \nu(\mathrm{d}x) \quad \forall\, A \in \mathcal{F}. \tag{2.1.2}$$

The probability defined in (2.1.2) is unique and well-defined due to [13, Prop. 7.28] as $(x, A) \mapsto \mathbb{P}^{0,x}(A)$ is a Markov kernel as $\mathbb{P}^{0,x}$ is a probability for each $x \in E$ and $x \mapsto \mathbb{P}^{0,x}(A)$ is a measurable map for all $A \in \mathcal{F}$ due to Lemma 2.3.2 in Section 2.3.1. Moreover, $\mathbb{P}^{0,x}$ is a regular conditional distribution with respect to the marginal law ν on E.

Hence, having completed the discussion of well-definedness of a PDP we next state sufficient conditions that guarantee that a PDP in the sense of Definition 2.1.1 is a strong Markov process, see Theorem 2.1.1. To this end we have to consider a special structure for, on the one hand, the measurable functions f^n that define the process between successive jumps, and, on the other hand, the Markov kernels $(P_x^n)_{n\geq 0}$ and $(\pi_x^n)_{n\geq 0}$, $x \in E$, defining the random counting measure μ conditional on the initial condition, cf. Theorem B.2.1. For the former,

(S) let D denote the set

$$D = \{(s, t, y) \in [0, \infty)^2 \times E |\, s \leq t\}$$

and $\varphi : D \mapsto E : (s,t,y) \mapsto \varphi_{s,t}(y)$ be a measurable function such that $\varphi_{s,t}$ forms a two-parameter semigroup of (usually nonlinear) operators on E, i.e., for all $0 \le s \le u \le t < \infty$

$$\varphi_{s,s} = \mathrm{id} \qquad \varphi_{s,t} = \varphi_{u,t} \circ \varphi_{u,r}\,. \tag{2.1.3}$$

We set $f^n_{x_0}(t|z_n) = \varphi_{\tau_n,t}(y_n)$ and note that this choice of $f^n_{x_0}$ is independent of n and the initial condition x_0 and depends on z_n only via the most recent jump time and mark. Then a PDP $(X_t)_{t\ge0}$ is defined according to Definition 2.1.1 by

$$X_t := \begin{cases} \varphi_{0,t}(X_0) & \text{if } t < \tau_1, \\ \varphi_{\tau_n,t}(Y_n) & \text{if } \tau_n \le t < \tau_{n+1}. \end{cases} \tag{2.1.4}$$

For the families of Markov kernels $(P^n_x)_{n\ge0}$ and (π^n_x) defining the random counting measure we impose the following structure which guarantees that the jump times and the marks of μ depend on the 'past' only via the most recent jump time and mark.

(P1) In general the Markov kernels $(P^n_x)_{n\ge0}$ from K^n into $[0,\infty]$ are uniquely defined by their survivor functions $\overline{S}^n_{x,z_n}$. That is, $\overline{S}^n_{x,z_n}(t)$ states the probability that the $(n+1)$th jump occurs after time t conditional on the the point process up to the nth event being z_n.

Then, we assume there exists a family of survivor functions $S_{s,y}$, $s \ge 0$, $y \in E$, such that $S_{s,y}(s) = 1$ for all $y \in E$ and every $s \ge 0$. Further it satisfies for all $y \in E$ and $s \le u \le t$ the functional relation

$$S_{s,y}(t) = S_{s,y}(u)\; S_{u,\varphi_{s,u}(y)}(t)\,. \tag{2.1.5}$$

Then, the survivor functions defining $(P^n_x)_{n\ge0}$, satisfy

$$\overline{S}^0_x(t) = S_{0,x}(t), \qquad \overline{S}^n_{z_n,x}(t) = S_{t_n,y_n}(t) \quad \forall\, n \ge 1\,. \tag{2.1.6}$$

(P2) As a condition on the Markov kernels $(\pi^n_x)_{n\ge1}$ from J^n to E we assume that there exists for each $t \in [0,\infty)$ a Markov kernel r_t from E to E such that $(t,x) \mapsto r_t(x,C)$ is measurable for every $C \in \mathcal{E}$. Then, for all $t \ge 0$, $C \in \mathcal{E}$, the kernels π^n satisfy

$$\pi^0_x(t,C) = r_t(\varphi_{0,t}(x),C), \qquad \pi^n_x((z_n,t),C) = r_t(\varphi_{t_n,t}(y_n),C) \quad \forall\, n \ge 1\,. \tag{2.1.7}$$

(P3) It will turn out that a PDP is a *homogeneous* Markov process if the semigroup $\varphi_{s,t}$, the Markov kernels r_t, and the survivor functions $S_{s,x}$ reduce to the following special cases:

$$\varphi_{s,t} = \phi_{t-s} \quad \forall\, t \geq s, \tag{2.1.8}$$

where $(\phi_t)_{t\geq 0}$ is a one-parameter semigroup of operators on E, i.e., $\phi_0 = \mathrm{id}$ and $\phi_{t+s} = \phi_t \circ \phi_s$,

$$r_t(y, C) = r(y, C) \quad \forall\, t \geq 0, \tag{2.1.9}$$

where r is a Markov kernel from E to E independent of t, and

$$S_{s,y}(t) = S_y(t-s) \quad \forall\, t \geq s \tag{2.1.10}$$

for a family of survivor functions $S_y, y \in E$ supported on $[0, \infty]$ for which the identity (2.1.5) reads[9]

$$S_y(t-s) = S_y(u-s)\; S_{\phi_{u-s}(y)}(t-u) \quad \forall\, s \leq u \leq t\,. \tag{2.1.11}$$

We are now in the position to state the central theorem of this section, stating that a PDP (2.1.4) generated by a random counting measure with the specific structure (2.1.5)–(2.1.7) is a strong Markov process with respect to the filtration $(\mathcal{F}_t)_{t\geq 0}$. We call a PDP which is a strong Markov process a *piecewise deterministic Markov process* (PDMP). A proof of the following theorem is deferred to Section 2.3.2.

Theorem 2.1.1. *A PDP* (2.1.4) *generated by a semigroup* $(\varphi_{t,s})_{s\geq t\geq 0}$ *and a random counting measure with regular conditional probabilities satisfying* (2.1.5)–(2.1.7) *is a strong Markov process with initial law* ν *with respect to the filtration* $(\mathcal{F}_t)_{t\geq 0}$ *and the measure* $\mathbb{P}^\nu$*. If the semigroup and the Markov kernels reduce to* (2.1.8)–(2.1.11) *then the PDMP is a homogeneous Markov process.*

We call the triple (φ, S, r) consisting of the semigroup $(\varphi_{s,t})_{0\leq s\leq t}$, the survivor functions $(S_{s,y})_{s\geq 0, y\in E}$ and the Markov kernels $(r_t)_{t\geq 0}$ which uniquely characterises the transition functions of a PDMP the *characteristics of the PDMP*. The characteristics together with an initial law ν on E define a PDMP uniquely up to versions, cf. Definition 2.0.1 (d).

2.1.1 Davis' construction procedure on the Hilbert Cube

Mark Davis [37] uses a constructive method to define the paths of a stochastic process on the Hilbert Cube taking values in a particular subset of an Euclidean space. Here the *Hilbert Cube* denotes the canonical probability space of an independently, identically distributed sequence of standard uniform random variables, cf. [13, 37]. He then proceeds to shows that the process defined in this way is a homogeneous PDMP in the

[9] Equivalently we can also impose for S_y that for $s, t \geq 0$ it holds that $S_y(t+s) = S_y(t)\, S_{\phi_t(y)}(s)$, which, in particular, states that S satisfies a *cocycle property* with respect to the flow ϕ.

sense of the preceding section. In the following we have adapted the original procedure to the current setting, that is, now considering general, inhomogeneous PDMPs on an arbitrary Borel measurable space and allowing for a random initial condition. This constructive definition of a PDMP on the Hilbert Cube is particularly important for numerical analysis of simulation methods for PDMPs. In general sampling algorithms for paths of stochastic processes rely on samples of certain defining random events, e.g., in the case of PDMPs, the jump times and marks. These are usually generated by transformations of standard uniform random variables simulated by a pseudo-random number generator. Thus Davis' construction procedure allows the definition of a version of a PDMP for given characteristics and an initial law from a sequence standard uniform random variables thus providing a theoretical sampling algorithm. In particular, for numerical analysis of approximation methods it provides the notion of an exact solution, see Chapter 5 for a further discussion.

Before we present the construction procedure itself we need the following lemma which allows us to obtain any probability measure on a Borel space via a pushforward measure from the standard uniform distribution. The following lemma is taken from [37, Cor. 23.4] where it is stated only for Borel spaces being Borel subsets of an Euclidean space. However, it obviously holds also for general Borel spaces. For the sake of completeness we also present its proof.

Lemma 2.1.1. *Let ν be a probability measure on the Borel measurable space $(E, \mathcal{E})$. Then there exists a Borel measurable function $\psi : [0,1] \to E$ such that the pushforward measure of the Lebesgue measure with respect to ψ on $(E, \mathcal{E})$ equals ν.*

In particular, note that the Lebesgue measure on the unit interval is the law of the standard uniform distribution. Further, the function ψ is not necessarily uniquely defined and choosing different functions for the construction procedure below results in pathwise different PDMPs which, however, are versions of each other defined on the same probability space.

Proof. As E is a Borel measurable space, there exists by definition a bimeasurable bijection $\widehat{\psi}$ between E and a Borel subset of $[0,1]$. As $\widehat{\psi}$ is a bijection on a subset of $[0,1]$ it is clear that $\widehat{\psi}$ is one-to-one (injective) into $[0,1]$, as well as measurable: It holds for any Borel set $A \subseteq [0,1]$ that $\widehat{\psi}^{-1}(A) = \widehat{\psi}^{-1}(A \cap \psi(E))$ which is a Borel subset of $\widehat{\psi}(E)$, hence as $\widehat{\psi}$ is measurable thereon, it is contained in $\mathcal{E}$. Thus $\widehat{\psi} : E \to [0,1]$ is one-to-one and measurable.
We then define $F(t) := \nu(\widehat{\psi}^{-1}([0,t])) \in [0,1]$, which is a cumulative distribution function on $[0,1]$, and its generalised inverse β is given by $\beta(u) := \inf\{t : F(t) \geq u\}$. Next, define

$$\chi(t) := \begin{cases} \widehat{\psi}^{-1}(t) & t \in \widehat{\psi}(E), \\ y_0 & t \notin \psi(E), \end{cases}$$

where y_0 is an arbitrary point in E. Then, for λ denoting the Lebesgue measure on $[0,1]$ it holds that $\lambda \circ \psi^{-1} = \nu$ where $\psi = \chi \circ \beta$, i.e., $\nu(A) = \lambda(\{u \in [0,1] : \psi(u) \in A\})$ for all $A \in \mathcal{E}$.

□

Definition 2.1.2. (*Davis' Construction procedure*) Let the probability space $(\widehat{\Omega}, \widehat{\mathcal{F}}, \widehat{\mathbb{P}})$ be the Hilbert Cube, that is, the space of all sequences $U_0(\widehat{\omega}), U_1(\widehat{\omega}), U_2(\widehat{\omega}), \dots$ with each random variable U_n taking values in $[0,1]$ such that each U_n is uniformly distributed on $[0,1]$ and U_i, U_j are mutually independent for all $i \neq j$. We note that we index the sequence of uniform random variables starting from 0 instead of the more common notation with starting at 1. However, this is an obvious choice in view of the construction procedure we present in the following.
Further, the semigroup $\varphi_{s,t}$, the Markov kernels r_t and the survivor functions $S_{t,y}$ satisfy (2.1.5)–(2.1.7). Further, let $u \mapsto \psi^\nu(u)$ be a measurable function on $[0,1]$ connected to ν in the sense of Lemma 2.1.1 where ν is the law of the initial condition. Next, we extend r_t for all $t \geq 0$ to a Markov kernel from $[0,\infty]$ onto $E \cup \{\nabla\}$ by setting $r_t(\{\infty\},\{\nabla\}) = 1$. Then, for all $t \geq 0$, $y \in \overline{E}$ we denote by $u \mapsto \psi_t(y,u)$ a measurable function on $[0,1]$ connected to the distribution on $\overline{E}$ given by $r_t(y,\cdot)$ in the sense of Lemma 2.1.1.

Then, we construct a marked point process recursively by the following algorithm:

Step 1. Use U_0 to define a random initial condition by

$$\widehat{X}_0(\widehat{\omega}) := \psi^\nu(U_0(\omega)) .$$

Step 2. Define the first jump time by

$$\widehat{\tau}_1(\widehat{\omega}) := \inf\Big\{ t \geq 0 \,\Big|\, S_{0,\widehat{X}_0(\widehat{\omega})}(t) \leq U_1(\widehat{\omega}) \Big\} .$$

Obviously, $\widehat{\tau}_1$ has a distribution on $[0,\infty]$ as specified by the survivor function $S_{0,X_0(\widehat{\omega})}$. Next, we define the first mark by

$$\widehat{Y}_1(\widehat{\omega}) := \psi_{\widehat{\tau}_1(\widehat{\omega})}\big(\varphi_{0,\widehat{\tau}_1(\widehat{\omega})}(\widehat{X}_0(\widehat{\omega}), U_2(\widehat{\omega}))\big) .$$

Step 3. Let $\widehat{\tau}_i$, $\widehat{Y}_i$ for $i = 1, \dots, n$ be the first n times and marks of a marked point process. Then we define the $(n+1)$th time by

$$\widehat{\tau}_{n+1}(\widehat{\omega}) := \inf\Big\{ t \geq 0 \,\Big|\, S_{\widehat{\tau}_n(\widehat{\omega}),\widehat{Y}_n(\widehat{\omega})}(t) \leq U_{2n+1}(\widehat{\omega}) \Big\}$$

and the $(n+1)$th mark is defined by

$$\widehat{Y}_n(\widehat{\omega}) := \psi_{\widehat{\tau}_n(\widehat{\omega})}\big(\varphi_{\widehat{\tau}_n(\widehat{\omega}),\widehat{\tau}_{n+1}(\widehat{\omega})}(\widehat{Y}_n(\widehat{\omega})), U_{2n+2}(\widehat{\omega})\big) .$$

It is clear that this construction procedure generates a marked point process, cf. Definition B.2.1, on the Hilbert Cube. A stochastic process $(\widehat{X}_t)_{t\geq 0}$ is then defined from the thus constructed marked point process $(\widehat{\tau}_n, \widehat{Y}_n)_{n\geq 1}$ and the initial condition $\widehat{X}_0 \in E$ by 2.1.4, that is,

$$\widehat{X}_t = \begin{cases} \varphi_{0,t}(\widehat{X}_0) & \text{if } t < \widehat{\tau}_1, \\ \varphi_{\widehat{\tau}_n,t}(\widehat{Y}_n) & \text{if } \widehat{\tau}_n \leq t < \widehat{\tau}_{n+1}. \end{cases}$$

Hence, due to Theorem 2.1.1 the process $(\widehat{X}_t)_{t\geq 0}$ is a PDMP on the Hilbert Cube with characteristics (φ, S, r) and initial law ν.

2.1.2 Some special PDMPs

When defining particular classes of PDMPs one has to consider two mechanisms: Firstly, the Markov kernels defining random counting measure, i.e., its jump times and post-jump values and, secondly, the semigroup for the inter-jump dynamics. In the following we discuss some special classes of PDMPs as considered in the relevant literature, e.g., [36, 37, 64], or classes used in mathematical neuroscience, cf. Chapter 3. Here, as this is also the case in the remainder of the thesis, we assume that the semigroups are continuous, i.e., the mappings $t \mapsto \varphi_{s,s+t}(y)$ or $t \mapsto \phi_t(y)$, respectively, are continuous for all y and all $s \geq 0$. We start with a discussion of two special case for the stochastic mechanism determining the jump times.

Example 2.1.1. As a first example we present PDMPs that move in all of the phase space E and admit jumps at instantaneous random times. This is the usual class of PDMPs encountered in models of biophysical neuronal membranes or chemical reaction networks. The instantaneous random jump times correspond to the random times when an ion channel switches its state or, in chemical reaction systems, a reaction occurs changing the number of different molecules. The connection of PDMPs and exact models of neuronal membrane is elucidated in comparing the subsequent discussions to (1.2.5). In particular, in these models the distribution of jumps is governed by a 'local' or 'instantaneous' jump rate, which in chemical reaction system corresponds to the usual reaction rate. In mathematical terms this means that the distribution of the waiting time distribution until the next jump has a density with respect to the Lebesgue measure on $[0,\infty)$. This does not exclude the possibility of a positive probability at $\{\infty\}$ in the case that the process stops jumping after finitely many jumps. For further details we refer to the discussion of intensity processes in connection with random counting measures in Section B.2.6 and hazard rates of survivor functions in Section B.3. We note that the majority of the results in [64] are restricted to this type of PDMP with an Euclidean space as the state space. Let

$\lambda_t(y) \geq 0$ denote the *local intensity* of a jump from $y \in E$ at time t, satisfying that $(t,y) \mapsto \lambda_t(y)$ is $\mathcal{B}\otimes\mathcal{E}$-measurable and locally integrable from the right, i.e., for every $y \in E$ and all $t \geq 0$ there exists an $\epsilon = \epsilon(t,y)$ such that for $h \in [0,\epsilon)$

$$\int_t^{t+h} \lambda_s(\varphi_{t,s}(y))\,\mathrm{d}s < \infty\,.$$

Further, we assume that the intensity along the semigroup, i.e., $t \mapsto \lambda_{s+t}(\phi_{s,s+t}(y))$, is càdlàg for all $y \in E$ and $s \geq 0$. Then we define a family of survivor functions $S_{s,y}$, $s \geq 0$, $y \in E$ by

$$S_{s,y}(t) := \exp\Big(-\int_s^t \lambda_u(\varphi_{s,u}(y))\,\mathrm{d}u\Big)\,. \tag{2.1.12}$$

Note, that $t \mapsto \lambda_t(\varphi_{s,t}(y))\,\mathbb{I}_{[t>s]}$ is the hazard rate corresponding to the survivor function $S_{s,y}$.
For the distribution on $(s,\infty]$ defined by the survivor function (2.1.12) there are, depending on the termination time

$$t^\dagger(s,y) := \inf\Big\{t>s\Big|\int_s^t \lambda_u(\varphi_{s,u}(y))\,\mathrm{d}u = \infty\Big\} = \sup\Big\{t>s\Big|\int_s^t \lambda_u(\varphi_{s,u}(y))\,\mathrm{d}u < \infty\Big\}\,, \tag{2.1.13}$$

where $\inf\emptyset = \infty$, the following possibilities:

(i) If $t^\dagger(s,y) < \infty$ or $t^\dagger(s,y) = \infty$ with $\int_s^\infty \lambda_u(\phi_{s,u}(y))\,\mathrm{d}u = \infty$, then the distribution corresponding to $S_{s,y}$ is supported on $(s, t^\dagger(s,y))$ and absolutely continuous with respect to the Lebesgue measure.

(ii) Otherwise, if $t^\dagger(s,y) = \infty$ with $\int_s^\infty \lambda_u(\varphi_{s,u}(y))\,\mathrm{d}u < \infty$, then the distribution corresponding to $S_{s,y}$ is supported on $(s,\infty]$ with non-zero probability at $\{\infty\}$.

In both cases, however, the survivor functions (2.1.12) satisfies the functional relation (2.1.5). Thus for a family of Markov kernels r_t satisfying condition (P2) and any E-valued initial condition X_0 there exists a PDMP which, in addition, is homogeneous if r_t and λ_t are in fact independent of t and the semigroup satisfies $\varphi_{s,t} = \phi_{t-s}$. In the homogeneous case the survivor functions (2.1.12) read

$$S_y(t-s) = \exp\Big(-\int_0^{t-s} \lambda(\phi_u(y))\,\mathrm{d}u\Big)\,.$$

The definition of the compensator (B.2.18) and the considerations in Example B.3.1 yield for the survivor function (2.1.12) that the random counting measure possesses the compensator

$$\widetilde{N}_t = \sum_{n=0}^{N_t-1} \int_{\tau_n}^{\tau_n+1} \lambda_s(\varphi_{\tau_n,s}(Y_n))\,\mathrm{d}s + \int_{\tau_{N_t}}^{t} \lambda_s(\varphi_{\tau_{N_t},s}(Y_{N_t}))\,\mathrm{d}s\,,$$

which, employing definition (2.1.4) of the PDMP $(X_t)_{t\geq 0}$, equals

$$\widetilde{N}_t = \int_0^t \lambda_s(X_s)\,\mathrm{d}s\,.$$

We note that $\widetilde{N}_t$ is predictable as it is continuous. Moreover, as by assumption the intensity λ_t along the paths of the PDMP $(X_t)_{t\geq 0}$ is càdlàg we can change the evaluation of the intensity along the PDMP to an evaluation along the left limit without changing the value of integral. Thus, consistently to Section B.2.6 we obtain

$$\widetilde{N}_t = \int_0^t \lambda_{s-}(X_{s-})\,\mathrm{d}s\,,$$

where the integrand process is left continuous and thus predictable. Furthermore, we obtain a representation of the $\mathcal{F}_t$-compensating measure $\widetilde{N}$ for the random counting measure μ determining the jumps of the PDMP by

$$\widetilde{N}(\mathrm{d}t, \mathrm{d}y) = \lambda_{t-}(X_{t-})\, r_{t-}(X_{t-}, \mathrm{d}y)\,\mathrm{d}t\,,$$

and each compensator $\widetilde{N}_t^C$, for $C \in \mathcal{E}$ has an $\mathcal{F}_t$-predictable intensity process given by

$$\Lambda_t^C = \lambda_{t-}(X_{t-})\, r_{t-}(X_{t-}, C)\,.$$

Finally, in the homogeneous case the intensity process becomes

$$\Lambda_t^C = \lambda(X_{t-})\, r(X_{t-}, C)\,.$$

and thus the compensator of the counting process $(N_t^C)_{t\geq 0}$ is given by

$$\widetilde{N}_t^C = \int_{(0,t]} \lambda(X_{s-})\, r(X_{s-}, C)\,\mathrm{d}s = \int_0^t \lambda(X_s)\, r(X_s, C)\,\mathrm{d}s\,.$$

Example 2.1.2. We next describe a class of processes which correspond to PDMPs as originally considered in [36, 37] and, in the context of General Stochastic Hybrid Systems, in [9, 10, 28]. However, generalising these authors, we allow for an arbitrary Borel space as a state space of the process instead of restricting ourselves to specific subsets of Euclidean spaces. The difference to Example 2.1.1 is that the PDMP does not move in all of E but is restricted to a subset of the state space. In addition to instantaneous jumps the process also exhibits 'forced' jumps when it hits the boundary

of the set it moves in. This class of PDMPs finds extensive use in control theory where usually a control parameter and the value of the quantity of interest is reset when it reaches a certain threshold. However, this type of PDMP also arises in mathematical neuroscience, e.g., integrate-and-fire neuron models with random reset value. In these neuron models a spike is emitted every time the voltage variable hits a prescribed boundary and is instantaneously reset to a value below the threshold.
Let B be some closed, measurable set in E and for all $y \in E$, $s \geq 0$, we define the hitting time

$$t^B(s,y) := \inf\{t > s \,|\, \varphi_{s,t}(y) \in B\}\,, \tag{2.1.14}$$

where, as always, we set $\inf \emptyset = \infty$. Note that $t^B(s,y) > s$ for all $y \in B^c$ and $t^B(s,y) = s$ for all $y \in B \backslash \partial B$. We define

$$B^+(s) := \big\{y \in E : t^B(s,y) > s\big\} \subseteq B^c \cup \partial B,$$

i.e., starting at $y \in B^+(s)$ the semigroup $\varphi_{s,t}(y)$ leaves into the exterior of B for $t > s$. Now, let for any $y \in B^+(t)$ the measurable mapping $\lambda_t(y) \geq 0$ be an intensity satisfying the conditions as in Example 2.1.1. Then we define for any $y \in B^+(s)$ a family of survivor functions by

$$S_{s,y}(t) := \mathbb{I}_{[t<t^B(s,y)]} \exp\Big(-\int_s^t \lambda_u(\varphi_{s,u}(y))\,\mathrm{d}u\Big)\,. \tag{2.1.15}$$

This family satisfies the functional relation (2.1.5) and the distributions such survivor functions define on $(s,\infty]$ are of one of the following types, where $t^\dagger(t,y)$ is given by (2.1.13):

(i) If $t^B(s,y) < \infty$ and $t^\dagger(s,y) = \infty$, then the distribution corresponding to $S_{s,y}$ is supported on $(s, t^B(s,y)]$ with non-zero probability at $\{t^B(s,y)\}$.

(ii) If $t^\dagger(s,y) < \infty$ and $t^B(s,y) < \infty$ then the distribution corresponding to $S_{s,y}$ is supported on $(s, t^\dagger(s,y) \wedge t^B(s,y)]$ and either absolutely continuous with respect to the Lebesgue measure if $t^\dagger(s,y) < t^B(s,y)$ or with non-zero probability at $\{t^B(s,y)\}$ if $t^B(s,y) \leq t^\dagger(s,y)$.

(iii) If $t^B(s,y) = \infty$ and $t^\dagger(s,y) = \infty$ with $\int_s^\infty \lambda_u(\varphi_{s,u}(y))\,\mathrm{d}u = \infty$, then the distribution corresponding to $S_{s,y}$ is supported on (s,∞) and absolutely continuous with respect to the Lebesgue measure.

(iv) If $t^B(s,y) = \infty$ and $t^\dagger(s,y) = \infty$ with $\int_s^\infty \lambda_u(\varphi_{s,u}(y))\,\mathrm{d}u < \infty$, then the distribution corresponding to $S_{s,y}$ is supported on $(s,\infty]$ with non-zero probability at $\{\infty\}$.

(v) If $t^B(s,y) = \infty$ and $t^\dagger(s,y) < \infty$ then the distribution corresponding to $S_{s,y}$ is supported on $(s, t^\dagger(s,y))$ and absolutely continuous with respect to the Lebesgue measure.

We continue defining the Markov kernels r_t for the distribution of the mark of the random counting measure μ or, equivalently, the post-jump value of the PDMP. Obviously, for the PDMP to be well defined a jump cannot yield a post-jump value y' in the interior of the set B as then $t^B(s, y') = 0$ and another jump would occur instantaneously. Therefore, we have to impose additional conditions on the family of Markov kernels r_t satisfying condition (P2) restricting their support as probability measures on E. That is, in addition, we assume that for all $t \geq 0$, $y \in B^c \cup \partial B$

$$\operatorname{supp} r_t(y, \cdot) \subseteq B^+(t)\,. \tag{2.1.16}$$

Note that here we have to consider the probabilities defined for all $y \in B^c \cup \partial B$ as a jump can either leave from the exterior of B or its boundary. Further, condition (2.1.16) allows that a jump may yield a position at the boundary of B, however, only to such values $y \in \partial B$ for which there exists an $\epsilon = \epsilon(t, y) > 0$ such that $\varphi_{t,t+h}(y) \notin B$ for all $h \in (0, \epsilon)$, that is, the process leaves from y into the exterior of B, cf. the definition of the sets $B^+(t)$.
Then, with these definitions of survivor functions and Markov kernels there exists a PDMP for any initial condition $x_0 \in B^+(0)$ which does not leave $B^c \cup \partial B$ and, moreover, does not spend any time at the boundary, i.e., the process only touches the boundary or restarts from the boundary after a jump leaving into B^c immediately. In addition, the PDMP is homogeneous if r_t and λ_t are independent of t and the semigroup satisfies $\varphi_{s,t} = \phi_{t-s}$.
The definition of the compensator (B.2.18) and the considerations in Example B.3.2 yield for the survivor function (2.1.15) that the random counting measure possesses the compensator

$$\widetilde{N}_t = \sum_{n=0}^{N_t-1} \int_{\tau_n}^{\tau_n+1} \lambda_s(\varphi_{\tau_n,s}(Y_n))\,\mathrm{d}s + \sum_{n=0}^{N_t-1} \mathbb{I}_{\{t^B(\tau_n,Y_n)\}}(\tau_{n+1}) + \int_{\tau_{N_t}}^{t} \lambda_s(\varphi_{\tau_{N_t},s}(Y_{N_t}))\,\mathrm{d}s\,,$$

which, using the definition (2.1.4) of the PDMP $(X_t)_{t\geq 0}$, equals

$$\widetilde{N}_t = \int_0^t \lambda_s(X_s)\,\mathrm{d}s + \sum_{n=1}^{N_t} \mathbb{I}_B(X_{\tau_n-})\,. \tag{2.1.17}$$

This yields for the counting process $(N_t^C)_{t\geq 0}$ the compensator

$$\widetilde{N}_t^C = \int_0^t \lambda_s(X_s) r_s(X_s, C)\,\mathrm{d}s + \sum_{n=1}^{N_t} r_{\tau_n}(X_{\tau_n-}, C)\,\mathbb{I}_B(X_{\tau_n-}), \tag{2.1.18}$$

which in the homogeneous case reduces to

$$\widetilde{N}_t^C = \int_0^t \lambda(X_s) r(X_s, A)\,\mathrm{d}s + \sum_{n=1}^{N_t} r(X_{\tau_n -}, C)\,\mathbb{I}_B(X_{\tau_n -})\,. \tag{2.1.19}$$

Note that, analogously to Example 2.1.1, we can change to left limits in the integrand processes in the compensators (2.1.17)–(2.1.19) to obtain predictable integrand processes without changing the values of the integral. Moreover, the compensator is a predictable process as, on the one hand, the integral in (2.1.19) is continuous and thus predictable and, on the other hand, the sum is piecewise constant changing its values only at hitting times of the set B which are predictable stopping times.

We now continue describing examples for defining the deterministic motion of a PDMP in between two random jumps. Of major and almost exclusive importance for the mechanism of inter-jump behaviour are semigroups that are generated by dynamical systems. In the following we restrict ourselves to homogeneous PDMPs, thus one-parameter semigroups, and distinguish between finite-dimensional semigroups, i.e., PDMPs taking values in Euclidean space $\mathbb{R}^d$, $d \geq 1$, and semigroups on an infinite-dimensional Hilbert space H. The former corresponds to the classical PDMPs encountered in the general literature [36, 37, 64] and the later are particularly motivated by applications to spatio-temporal models in mathematical neuroscience. In the context of dynamical systems one-parameter semigroups usually arise from autonomous evolution equation whereas, if we consider dynamical systems arising, e.g., as solutions to non-autonomous differential equations, the resulting processes are two-parameter semigroups, cf. [73]. With respect to the applications we consider the restriction to one-parameter semigroups is well justified, as in these the survivor functions S and the Markov kernels r are always autonomous. Further, by an extension of the state space by a time component we can always transform a two-parameter semigroup into a one-parameter semigroup (on a larger space) which allows to cast models into the framework of homogeneous Markov processes. Nevertheless, if explicitly needed all results can also be derived for non-homogeneous processes.

Definition 2.1.3. A *dynamical system* on a metric space E is a one-parameter semigroup $(\phi_t)_{t\geq 0}$ of nonlinear operators such that $\phi_t : E \to E$ is continuous for all $t \geq 0$. If in addition the mappings $t \mapsto \phi_t(y)$ are continuous for all $y \in E$ we call $(\phi_t)_{t\geq 0}$ a *continuous dynamical system*.

A general Borel measurable space $(E, \mathcal{E})$ which is the state space for the PDMP and thus the space the semigroup is acting on is easily made into a metric space choosing a metric which is compatible with the topology. Usually, and always in the present thesis, the Borel measurable space E is a separable Hilbert space or a subset thereof and hence the appropriate metric is given by the norm. We briefly comment on

Definition 2.1.3 as the assumptions on what constitutes a dynamical system slightly vary in the literature. The conditions for a semigroup to be a dynamical system as in Definition 2.1.3 correspond to the minimal assumptions made for a dynamical system in [116, Chap. 1.1]. In [104, Chap. 10, p. 262] only continuous dynamical systems are considered, termed simply dynamical systems therein. The authors in [73] assume joint continuity for $(t, y) \mapsto \phi_t(y)$ for a dynamical system. Further, dynamical systems defined only forward in time, i.e., for $t \geq 0$ in contrast to $t \in \mathbb{R}$, are termed semi-dynamical systems therein. It is noted in [104] that the continuity in t and in y does not necessarily imply joint continuity. However, the necessary condition for a semigroup defining the deterministic motion of a PDMP, cf. condition (S) in Section 2.1, is that the map $(t, y) \mapsto \phi_t(y)$ is jointly measurable with respect to t and y. The connection between continuous dynamical systems and inter-jump dynamics of PDMPs is provided by the following proposition.

Proposition 2.1.1. *A continuous dynamical system $(\phi_t)_{t\geq 0}$ on a Borel measurable space E is jointly measurable in t and y.*

Proof. Equipped with a compatible metric a Borel space E is a separable metric space. Further, $[0, \infty)$ is a measurable space together with its Borel-σ-field. By definition the continuous dynamical system $(\phi_t)_{t\geq 0}$ satisfies that $y \mapsto \phi(t, y)$ is continuous for all $t \geq 0$ and $t \mapsto \phi(t, y)$ is measurable for all $y \in E$ as continuity implies measurability. Such a function is a special case of a *Carathéodory function*, cf. [4, Sect. 4.10]. As E is separable and a metrisable topological space it follows by [4, Lemma 4.51] that $(t, y) \mapsto \phi_t(y)$ is jointly measurable. □

Example 2.1.3. We first consider dynamical system that arise as solutions of ordinary differential equations. This is the setup that is encountered in most PDMP applications as it is part of the original setup of Davis [36, 37]. In mathematical neuroscience this will be used to model the transmembrane potential which follows the membrane balance ODE, cf. (1.2.2), in models neglecting the spatial dimension. In chemical reaction system modelling the evolution of molecule concentration by the reaction Reaction Rate Equation yields an ODE system. Let $E \subseteq \mathbb{R}^d$ be an open Borel set which is a Borel space itself equipped with the d-dimensional Borel σ-field restricted to E. We then consider for a function $g : E \subseteq \mathbb{R}^d \to \mathbb{R}^d$ the initial value problem

$$\dot{y} = g(y), \quad y(0) = y_0, \tag{2.1.20}$$

and assume that it has a global solution for every initial condition $y_0 \in E$ which we denote by $\phi_t(y_0)$. Then ϕ defines a one-parameter semigroup on E if the solution remains in E for all positive times for any initial condition $y_0 \in E$. Furthermore, [114, Sect. 2] note that for most standard assumptions leading to existence and uniqueness of a solution to (2.1.20) also continuity with respect to initial data follows, e.g., if g

is globally Lipschitz or locally Lipschitz and the solutions are bounded. Hence the Carathéodory measurability condition implies that $(t, y) \mapsto \phi_t(y)$ is jointly measurable and the semigroup satisfies the condition (S) of the existence Theorem 2.1.1. The situation is completely analogous if E is a separable Banach space and (2.1.20) is an ordinary differential equation on the Banach space E. However, this does not cover most interesting infinite-dimensional examples, cf. the subsequent Example 2.1.4.
Bearing in mind the considerations in Example 2.1.2 it is not necessary for the semigroup to remain in the set E for all positive times. To define a PDMP one has to 'force' a jump every time the solution leaves the domain E. Clearly, in this case it is sufficient that the semigroup property holds only for $t \geq 0$ such that $t < t^\dagger(y_0)$ for every initial condition $y_0 \in E$. This approach to the definition of a PDMP can also be used for $E = \mathbb{R}^d$ and (2.1.20) allows only for local solutions. In this case one forces a jump every time a solution explodes, i.e., the trajectory hits the 'boundary' $\{\infty\}$. Note that $\mathbb{R}^d \cup \{\infty\}$ is a Borel space and $\{\infty\}$ a closed Borel set. Such a PDMP can be used to describe an *exponential integrate-and-fire neuron model* where a spike occurs every time the voltage variable diverges.

Example 2.1.4. For applications to excitable media we are usually interested in the spatio-temporal evolution that arises from parabolic partial differential equations, cf. (1.2.1). Considered as abstract evolution equations the state space of the PDMP is a suitable function space over some spatial domain $D \subseteq \mathbb{R}^d$. Normally such equations cannot be considered as (strong) ODEs in a Banach or Hilbert space but an appropriate setup is the theory of weak solutions to abstract evolution equations. The choice of considering weak solutions is not based on an intrinsic necessity but allows to consider more general equations as when dealing with mild, strong or classical solutions, cf. a discussion of this aspect in [122, Chap. 23.1]. The important property for the definition of the inter-jump dynamics of a PDMP is the semigroup property of solutions which is, in particular, satisfied for mild solutions. The subsequent presentation reflects standard existence theory of parabolic evolution equations, cf. [43, 122], and, in particular, as it is discussed in infinite-dimensional dynamical systems theory, cf. [104, 116].
Let $X \subset H \subset X^*$ be an *evolution triple*[10] of separable, real Hilbert spaces and

[10] Let X, H be a real, separable Hilbert spaces such that X is continuously and densely embedded in H. Then the triple $X \subset H \subset X^*$ is called an *evolution triple*. Here H^* is identified with H and from the density of X in H it follows that $H^* \hookrightarrow X^*$ and thus $H \hookrightarrow X^*$. Moreover as X is reflexive it is valid that H is dense in X^*. The central property of an evolution triple, however, is the connection of the scalar product in H and the duality pairing of X, i.e.,

$$\langle \phi, u \rangle_X = (\phi, u)_H \quad \forall \phi \in H, u \in X\,. \tag{2.1.21}$$

For example, in the case of abstract evolution equations generated by parabolic partial differential equations on a domain $D \subset \mathbb{R}^d$ equipped with suitable boundary conditions the appropriate evolution triple is given by the spaces $X = H^1(D)$ and $H = L^2(D)$.

consider the abstract, linear, inhomogeneous initial value problem

$$\dot{u} = Au + b, \quad u(0) = u_0 \in H\,, \tag{2.1.22}$$

where $A : X \to X^*$ is a linear, bounded operator and $b \in X^*$. A function $u \in L^2((0,T),X) \cap H^1((0,T),X^*) \subset C^0([0,T],H)$ is called a weak solution to (2.1.22) if $u(0) = u_0$ in H and

$$\int_0^T \langle \dot{u}(t), v(t)\rangle_X \,\mathrm{d}t = \int_0^T \langle Au(t), v(t)\rangle_X + \langle b, v(t)\rangle_X \,\mathrm{d}t$$

for all $v \in L^2((0,T),X)$. A general existence theorem, cf. [104, 116, 122] guarantees under coercivity and suitable boundedness conditions the unique existence of a weak solution depending continuously on the initial condition. Hence, the solution generates a continuous dynamical system in H and the Carathéodory measurability condition implies that it is jointly measurable. Therefore, the semigroup defined by the abstract initial value problem (2.1.22) satisfies the condition of the existence Theorem 2.1.1.

2.2 Extended generator and Itô formula

The generator of a Markov process is an important object in the study of Markov processes as it allows, e.g., a unique characterisation of the process via the Martingale Problem. For certain special classes of PDMPs the infinitesimal and / or the extended generator are derived and discussed in [37, 64, 68]. At this point we present a unified derivation of the extended generator and subsequently of an Itô formula for homogeneous PDMPs with survivor functions as considered in Example 2.1.2 and jointly-measurable semigroups $(\phi_t)_{t\geq 0}$ on E. That is, jumps of the PDMP occur in a random instantaneous fashion as well as upon the trajectory hitting a prescribed boundary. For ease of presentation we briefly repeat the definitions. Firstly, let B be a closed Borel set of a Borel measurable space $(E,\mathcal{E})$. We define the hitting time of the boundary with respect to the initial condition by

$$t^B(y) := \inf\{t > 0 \,|\, \phi_t(y) \in B\}\,,$$

and the set including parts of the boundary which is a target for post jump values is given by

$$B^+ := \big\{y \in E : t^B(y) > 0\big\} \subseteq B^c \cup \partial B.$$

Then for each $y \in B^+$ we define a survivor function by

$$S_y(t) = \mathbb{I}_{[t<t^B(y)]} \exp\Big(-\int_0^t \lambda(\phi_s(y))\,\mathrm{d}s\Big)\,, \tag{2.2.1}$$

where $\lambda : B^c \cup \partial B \subset E \to \mathbb{R}_+$ is a non-negative, measurable function which is locally integrable along the trajectories of the semigroup, cf. Example 2.1.1. For a further discussion of the jump rate we refer to Example 2.1.2. Next, the termination time of the corresponding survivor function S_y is given by

$$t^\dagger(y) := \inf\{t > 0 : S_y(t) = 0\} = t^B(y) \wedge \inf\Big\{t > 0 \,\Big|\, \int_0^t \lambda(\phi_s(y))\,\mathrm{d}s = \infty\Big\}.$$

Finally, for the Markov kernel r from E to E we assume that for all $y \in B^c \cup \partial B$ it holds that $\operatorname{supp} r(y, \cdot) \subseteq B^+$. That is, the distribution of the post-jump value is always supported on the set B^+ of initial conditions such that the trajectories of the semigroup started at such a value leave into the exterior of B.

2.2.1 The generator of piecewise continuous PDMPs

Before we present the theorem fully characterising the extended generator we introduce properties of real, measurable functions used in the following.

Definition 2.2.1. We call a measurable function $f : E \to \mathbb{R}$ *path-continuous* if the mapping $t \mapsto f(\phi_t(y))$ is continuous for all $y \in E$ and if, in addition, this function is continuously differentiable we say f is *path-differentiable*. For path-differentiable f we define the *path derivative* or, alternatively, the *derivative along the flow* by

$$\mathrm{d}_\phi f(y) := \lim_{s \downarrow 0} \frac{f(\phi_s(y)) - f(y)}{s}\,, \tag{2.2.2}$$

where the limit exists by definition and it follows that for all $t \geq 0$

$$\frac{\mathrm{d}}{\mathrm{d}t} f(\phi_t(y)) = \mathrm{d}_\phi f(\phi_t(y))\,, \tag{2.2.3}$$

Finally, a mapping f is called *almost everywhere (a.e.) path-differentiable* if for each $y \in E$ the map $t \mapsto f(\phi_t(y))$ is differentiable with respect to t for Lebesgue-almost every $t \geq 0$.

Note that absolute continuity of the function $t \mapsto f(\phi_t(y))$ is equivalent to a function being almost everywhere path-differentiable, hence stronger than path-continuity but weaker than path-differentiability.

The PDMP $(X_t)_{t \geq 0}$ is a stochastic process on $B^+ \subseteq E$, hence as potential elements of the extended generator we consider measurable, bounded real-valued functions on B^+ which in addition satisfy

$$\lim_{t \to t^B(y)} f(\phi_t(y)) < \infty \quad \forall\, y \in B^+ : \, t^B(x) < \infty\,, \tag{2.2.4}$$

i.e., the evaluated along the trajectories of the semigroups the function possesses a limit when approaching the boundary. Moreover, it is clear that all conditions on f need to hold only on a set of values that can be attained by the PDMP. The proof of the following theorem is deferred to Section 2.3.3.

Theorem 2.2.1. *Let $(X_t)_{t\geq 0}$ be a piecewise continuous PDMP with initial law ν and characteristics (ϕ, S, r) as specified above. Then the domain of the extended generator $\mathcal{D}(\mathcal{A})$ is given by all bounded, measurable functions $f : B^+ \subseteq E \to \mathbb{R}$ satisfying* (2.2.4) *such that*

(i) $t \mapsto f(\phi_t(y))$ *is a.e. path-differentiable on* $[0, t^\dagger(y))$,

(ii) f *satisfies for every* $y \in B^+ \cap \partial B$ *the boundary condition*

$$f(y) = \int_E f(x)\, r(y, \mathrm{d}x)\,, \tag{2.2.5}$$

(iii) $(y, t, \omega) \mapsto f(y) - f(X_{t-}(\omega))$ *is a valid integrand for the compensating measure* $\widetilde{N}$.[11]

Then, for $f \in \mathcal{D}(\mathcal{A})$ *the extended generator is given by*

$$\mathcal{A}f(y) = \mathrm{d}_\phi f(y) + \lambda(y) \int_E f(x) - f(y)\, r(y, \mathrm{d}x)\,. \tag{2.2.6}$$

We remark that the a.s. path-differentiability of f implies that the process in (iii), which is adapted to $(\mathcal{F}_t)_{t\geq 0}$, is left-continuous and thus $\mathcal{F}_t$-predictable for each fixed $y \in E$. Further, we note that the domain of the generator is always non-empty as it contains all constant functions which satisfy $\mathcal{A}f \equiv 0$. Moreover, if $\mathcal{A}f \equiv 0$, then $f(X_t)$ is an $\mathcal{F}_t$-martingale.
An immediate consequence of the extended generator is *Dynkin's formula*. Let f be in the domain of the extended generator such that its integral with respect to the compensated measure $M = N - \widetilde{N}$ is an $\mathcal{F}_t$-martingale, then it holds for all $t \geq 0$

$$\mathbb{E}[f(X_t)|X_0] = f(X_0) + \mathbb{E}\Big[\int_0^t \mathcal{A}f(X_s)\,\mathrm{d}s \,\Big|\, X_0\Big]\,. \tag{2.2.7}$$

Example 2.2.1. As a first example we discuss PDMPs taking values in Euclidean space, i.e., $E = \mathbb{R}^d$, for which the semigroup is generated by a system of ordinary differential equations, cf. Example 2.1.3. Thus $\phi_t(y_0)$ is the solution of the initial value problem

$$\dot{y} = g(y), \quad y(0) = y_0,$$

[11]This means that the compensated stochastic integral is an $\mathcal{F}_t$-local martingale.

where $g : \mathbb{R}^d \to \mathbb{R}^d$. Obviously we obtain that the semigroup has continuously differentiable paths (we tacitly assume that g is continuous). If we assume that $f \in \mathcal{D}(\mathcal{A})$ is continuously differentiable, then the extended generator (2.2.6) is given by

$$\mathcal{A}f(y) = \nabla f(y) \cdot g(y) + \lambda(y) \int_E f(x) - f(y)\, r(y, \mathrm{d}x)$$

where $\nabla f(y)$ denotes the gradient of f evaluated at y. This is precisely the result of [37, Thm. 26.14].

Example 2.2.2. We continue the Example 2.1.4, i.e., we suppose that the separable Hilbert spaces $X \subset H \subset X^*$ form an evolution triple and the semigroup ϕ_t on H is generated by a weak solution of an abstract initial value problem (2.1.22). In contrast to the finite-dimensional case considered in Example 2.2.1 the structure of the generator as in (2.2.8) and (2.2.9) does not follow immediately from the chain rule of differentiation as the trajectories of the semigroup ϕ and thus the paths of the process lack the necessary differentiability. As the solution is only a weak solution in general ϕ is not in $C^1([0,T], H)$. However, under certain conditions on f we can derive an expression for the extended generator which we formulate in the following theorem.

Theorem 2.2.2. *Let f be in the domain of the extended generator and additionally be in $C^1(H, \mathbb{R})$. Further, let $f_y(x)$ denote the unique element of H such that*

$$\frac{\mathrm{d}f}{\mathrm{d}y}[x](z) = \big(z, f_y(x)\big)_H \quad \forall\, z \in H,$$

where $\frac{\mathrm{d}f}{\mathrm{d}y}[x]$ denotes the Fréchet-derivative of f in H at $x \in H$. Further, we assume for f that $f_y(x) \in X$ if $x \in X$ and the composition operator is locally bounded in $L^2((0,T), X)$.[12] *Then for almost every $t \in [0,T]$ the extended generator $\mathcal{A}$ of the Hilbert space valued PDMP is given by*

$$\mathcal{A}f(y) = \langle Ay + b, f_y(y)\rangle_X + \lambda(y) \int_H f(z) - f(y)\, r(y, \mathrm{d}z)\,. \tag{2.2.8}$$

If in addition ϕ is a strong solution, i.e., $\dot{\phi} \in L^2((0,T), H)$, then for any function $f \in C^1(H, \mathbb{R})$ the extended generator is given for almost all t by

$$\mathcal{A}f(y) = \big(Ay + b, f_y(y)\big)_H + \lambda(y) \int_E f(z) - f(y)\, r(y, \mathrm{d}z)\,. \tag{2.2.9}$$

The Fréchet derivative at a point $x \in H$ is a linear functional on H, hence the existence and uniqueness of such an element $f_y(x) \in H$ is guaranteed by the Riesz

[12] An example of such a function f is $\|\cdot\|_H^2$ in which case $\frac{\mathrm{d}\|\cdot\|_H^2}{\mathrm{d}y}[x] = 2(\cdot, x)_H$.

Representation Theorem. We note that Theorem 2.2.2 extends in a straightforward way if we consider an abstract nonlinear equation $\dot{u} = Au + G(u)$, where $G : X \to X^*$ is a nonlinear operator, instead of (2.1.22).

Proof. In order to prove equations (2.2.8) and (2.2.9) we have to show that the path-derivative $d_\phi f$ in (2.2.6) possesses the proposed structure. As already mentioned this structure would follow immediately from the chain rule of differentiation if the trajectories of the semigroup itself were sufficiently smooth. Thus we introduce the concept of mollifiers which regularise the trajectories and then obtain the desired result for the trajectories as a limit of the results for the smooth ones.

We denote by $\mu \in C^\infty(\mathbb{R})$ the mollifier on $\mathbb{R}$ given by

$$\mu(t) := \begin{cases} C\exp\left(\frac{1}{t^2-1}\right) & |t| < 1 \\ 0 & |t| \geq 1, \end{cases}$$

where the constant C is selected such that $\int_{\mathbb{R}} \mu(t)\,\mathrm{d}t = 1$. Further, we define that $\mu_\epsilon(t) := \epsilon^{-1}\mu(t/\epsilon)$. Then for any locally integrable function u on $[-\epsilon, T+\epsilon]$, that is, any $u \in L^1_{\mathrm{loc}}((-\epsilon, T+\epsilon), X)$, the mollification of u is the function u_ϵ defined on $[0,T]$ by the convolution

$$u_\epsilon(t) := \int_{[-\epsilon, T+\epsilon]} \mu_\epsilon(t-s)u(s)\,\mathrm{d}s\,. \tag{2.2.10}$$

It follows that $u_\epsilon \in C^\infty([0,T], X)$.

Considering the path-derivative the initial condition y for the trajectories $t \mapsto \phi_t(y)$ is fixed and thus for simplicity of notation we use in the following $\phi(t) := \phi_t(y)$. Further, we denote by ϕ_ϵ the mollification of the trajectory of the semigroup, i.e., $\phi_\epsilon \in C^\infty([0,T], X)$ as well as in $C^\infty([0,T], H)$. From considerations in [43, p. 287] it follows for $T < \infty$ and $\epsilon \to 0$ that

$$\phi_\epsilon \to \phi \text{ in } L^2((0,T), X), \quad \dot{\phi}_\epsilon \to \dot{\phi} \text{ in } L^2((0,T), X^*). \tag{2.2.11}$$

Moreover, ϕ_ϵ also converges to some element $w \in C([0,T], H)$. Thus we can conclude that $\phi(t) = w(t)$ in H for almost every t, i.e., $\phi = w$ in $L^2((0,T), H)$. These results provide the necessary convergence properties of the smoothed trajectories of the semigroup.

To obtain the representation (2.2.8) for the derivative of $t \mapsto f(\phi(t))$ in (2.2.6) we start with the Fréchet derivative $\frac{\mathrm{d}f}{\mathrm{d}y}(x)$ of f evaluated at $x \in H$. This is a linear, continuous map $H \to \mathbb{R}$, hence it is an element of H^*. Therefore, by the Riesz Representation Theorem, we obtain that we can identify

$$\frac{\mathrm{d}f}{\mathrm{d}y}[x] = (\cdot, f_y(x))_H\,,$$

where $f_y(x)$ denotes the unique element $u \in H$ such that

$$\frac{\mathrm{d}f}{\mathrm{d}y}[x](z) = (z,u)_H \quad \forall\, z \in H.$$

Hence, by the chain rule we obtain

$$\frac{\mathrm{d}f\circ\phi_\epsilon}{\mathrm{d}t}[t] = \frac{\mathrm{d}f}{\mathrm{d}y}[\phi_\epsilon(t)]\circ\dot{\phi}_\epsilon[t] = \big(\dot{\phi}_\epsilon[t], f_y(\phi_\epsilon(t))\big)_H\,.$$

Next we identify $\dot{\phi}_\epsilon[t] \in H$ with an element of X^* in the canonical sense, that is

$$\dot{\phi}_\epsilon[t] : X \to \mathbb{R} : y \mapsto (\dot{\phi}_\epsilon[t], y)_H\,,$$

and therefore we can write

$$\frac{\mathrm{d}f\circ\phi_\epsilon}{\mathrm{d}t}[t] = \langle\dot{\phi}_\epsilon[t], f_y(\phi_\epsilon(t))\rangle_X\,, \tag{2.2.12}$$

where we have used the assumption that $f_y(x) \in X$ for $x \in X$.
Next we fix a point $t_1 \in (0,T)$ such that $\phi_\epsilon(t_1) \to \phi(t_1)$ in X for $\epsilon \to 0$ and integrate (2.2.12) for $t_2 > t_1$ to obtain

$$f(\phi_\epsilon(t_2)) - f(\phi_\epsilon(t_1)) = \int_{t_1}^{t_2} \langle\dot{\phi}_\epsilon[s], f_y(\phi_\epsilon(s))\rangle_X\,\mathrm{d}s. \tag{2.2.13}$$

Thus taking the limit for $\epsilon \to 0$ we find the left hand side of (2.2.13) converges to $f(\phi(t_2)) - f(\phi(t_1))$ for almost all t_2 in $(0,T)$.
For the integrand on the right hand side of (2.2.13) note that for almost all s it is valid due to (2.2.11) that $\dot{\phi}_\epsilon[s]$ converges to $\dot{\phi}[s]$ strongly in X^*. Furthermore, as f is continuously differentiable, i.e., the mapping $x \mapsto \frac{\mathrm{d}f}{\mathrm{d}y}[x]$ is continuous from H into H^*, it holds that

$$\lim_{\epsilon\to 0}\Big\|\frac{\mathrm{d}f}{\mathrm{d}y}[\phi_\epsilon(s)] - \frac{\mathrm{d}f}{\mathrm{d}y}[\phi(s)]\Big\|_{H^*} = 0\,.$$

As strong convergence implies weak*-convergence we have that

$$\frac{\mathrm{d}f}{\mathrm{d}y}[\phi_\epsilon(s)](z) \to \frac{\mathrm{d}f}{\mathrm{d}y}[\phi(s)](z) \quad \forall\, z \in H\,,$$

which by definition of f_y is equivalent to

$$\big(z, f_y(\phi_\epsilon(s))\big)_H \to \big(z, f_y(\phi(s))\big)_H \quad \forall\, z \in H.$$

As H is dense in X^* and $f_y(\phi_\epsilon(s))$ is bounded in X – $\phi_\epsilon[s]$ is strongly convergent in X, thus also weakly convergent, thus bounded and by assumption f_y is bounded in X for bounded arguments – it follows that $f_y(\phi_\epsilon(s))$ converges to $f_y(\phi(s))$ weakly in

X [122, Prop. 21.23(g)]. Hence, [122, Prop. 21.23(j)] implies that

$$\langle(\dot{\phi}_\epsilon[s], f_y(\phi_\epsilon(s))\rangle_X \to \langle(\dot{\phi}[s], f_y(\phi(s))\rangle_X$$

for almost all $s \in (t_1, t_2)$. Finally, dominated convergence yields

$$\int_{t_1}^{t_2} \langle\dot{\phi}_\epsilon[s], f_y(\phi_\epsilon(s))\rangle_X \, \mathrm{d}s \to \int_{t_1}^{t_2} \langle\dot{\phi}[s], f_y(\phi(s))\rangle_X \, \mathrm{d}s\,,$$

and hence for almost every t_2 we have

$$f(\phi(t_1)) = f(\phi(t_2)) + \int_{t_1}^{t_2} \langle\dot{\phi}[s], f_y(\phi(s))\rangle_X \, \mathrm{d}s\,.$$

This implies for almost every t in $(0, T)$ that

$$\frac{\mathrm{d}}{\mathrm{d}t} f(\phi(t)) = \langle A\phi(t) + b, f_y(\phi(t))\rangle_X\,.$$

Hence the representation (2.2.8) of the extended generator is proved.

If, in addition, $\dot{\phi} \in L^2((0,T), H)$ we obtain for the mollified version the limits

$$\phi_\epsilon \to \phi \text{ in } L^2((0,T), H), \quad \dot{\phi}_\epsilon \to \dot{\phi} \text{ in } L^2((0,T), H).$$

Just as before we find that

$$(z, f_y(\phi_\epsilon(s)))_H \to (z, f_y(\phi(s)))_H \quad \forall\, z \in H,$$

hence $f_y(\phi_\epsilon(s))$ converges weakly to $f_y(\phi(s))$ in H. Then a repetition of the same arguments as above yields

$$(\dot{\phi}_\epsilon[s], f_y(\phi_\epsilon(s)))_H \to (\dot{\phi}[s], f_y(\phi(s)))_H\,,$$

and by dominated convergence it follows that for almost all t_2

$$f(\phi(t_1)) = f(\phi(t_2)) + \int_{t_1}^{t_2} \big(\dot{\phi}[s], f_y(\phi(s))\big)_H \, \mathrm{d}s\,.$$

This completes the proof of the representation (2.2.9). $\square$

2.2.2 Itô formula

The following version of an Itô-formula for PDMPs is a generalisation of the corresponding result in Davis [37, Thm. 31.3] to the state space E being an arbitrary Borel measurable space. Subsequently we also show how the Itô-formulae in Jacobsen [64, Thm. 7.6.1(a), Prop. 7.7.1] are derived as special cases.

Theorem 2.2.3. *We assume that $f \in B(E, \mathbb{R})$ is a valid integrand function for the martingale measure M, cf. Section B.2.4, associated with the PDMP $(X_t)_{t\geq 0}$ and that f is almost surely path-differentiable. Then for all $t \geq 0$ it holds that*

$$\begin{aligned} f(X_t) - f(X_0) &= \int_0^t \mathcal{A}f(X_s)\,\mathrm{d}s + \int_0^t \int_E f(y) - f(X_s)\, M(\mathrm{d}s, \mathrm{d}y) \\ &\qquad + \sum_{\tau_i \leq t} \int_E f(y) - f(X_{\tau_i -})\, r(X_{\tau_i -}, \mathrm{d}y)\, \mathbb{I}_{[X_{\tau_i -} \in B]}, \end{aligned} \tag{2.2.14}$$

where $\mathcal{A}f$ is given by the characterisation of the extended generator (2.2.6)*.*

Proof. Observe that the function f satisfies conditions (i) and (ii) in Theorem 2.2.3 for being in the extended generator. Then, the Itô-formula (2.2.14) follows immediately from the considerations in part (a) of the proof of Theorem 2.2.1, see Section 2.3.3, where the additional last term in (2.2.14) arises as f does not satisfy the boundary condition (2.2.5). □

We note that the stochastic integral in (2.2.14) is a local martingale and under certain conditions, cf. Theorem B.2.6, it is even a martingale. Further by definition of the martingale measure $M = N - \widetilde{N}$ equation (2.2.14) is equivalent to

$$f(X_t) - f(X_0) = \int_0^t \frac{\mathrm{d}}{\mathrm{d}s} f(X_s)\,\mathrm{d}s + \int_0^t \int_E f(y) - f(X_{s-})\, \mu(\mathrm{d}s, \mathrm{d}y)\,. \tag{2.2.15}$$

We remark that one has to be cautious with the interpretation of the formula (2.2.15). The counting measure μ therein is the counting measure associated with the PDMP $(X_t)_{t\geq 0}$ and thus depends on the path of the PDMP via its defining objects, the survivor functions S_y and the Markov kernel r. Hence, there is a clear difference to *stochastic differential equations of jump type* which may take a similar form to (2.2.15). In connection with stochastic differential equations of jump type the occurring counting measure μ is a Poisson random measure. However, there is also some overlap. The Poisson random measure is a random counting measure connected in the sense of Section 2.1 for which the defining Markov kernels and survivor functions are as in (P3) and in addition independent of $y \in E$. Thus, a Poisson random measure is a special type of random counting measure and hence we obtain that solutions to certain stochastic differential equations are also PDMPs, or, conversely, certain PDMPs solve stochastic differential equations. We refer to [64, Sec. 7.4.3] for a further discussion.

2.3 Section Appendix

2.3.1 Two helpful results

Lemma 2.3.1. *For every $x \in E$ it holds that the σ-field $\sigma(X_0)$ is independent of $\mathcal{F}^\mu_\infty$ under the probability $\mathbb{P}^{0,x}$.*

Proof. First note that $X_0^{-1}(\mathcal{E})$ is an $\cap$-stable[13] generator of $\sigma(X_0)$. Therefore, if we show that each $A \in X_0^{-1}(\mathcal{E})$ is independent of $\mathcal{F}^\mu_\infty$, then by [8, Cor. 30.4] it follows that $\sigma(X_0)$ and $\mathcal{F}^\mu_t$ are independent σ-fields.
Let A be an arbitrary set in $\mathcal{F}^\mu_\infty$ and $C \in \mathcal{E}$ then, either $x \notin C$ and the probability

$$\mathbb{P}^{0,x}\big([X_0 \in C] \cap A\big) \;=\; 0 = \mathbb{P}^{0,x}[X_0 \in C]\,\mathbb{P}^{0,x}(A)$$

as the event $[X_0 \in C]$ has probability zero under $\mathbb{P}^{0,x_0}$, or $x \in C$, then $x \notin C^c$ and

$$\begin{aligned}\mathbb{P}^{0,x}\big([X_0 \in C] \cap A\big) \;&=\; \mathbb{P}^{0,x}\big([X_0 \in C] \cap A\big) + \mathbb{P}^{0,x}\big([X_0 \in C^c] \cap A\big)\\ &=\; \mathbb{P}^{0,x}(A) \;=\; 1 \cdot \mathbb{P}^{0,x}(A) \;=\; \mathbb{P}^{0,x}[X_0 \in C] \cdot \mathbb{P}^{0,x}(A)\,.\end{aligned}$$

Thus, overall, it holds for arbitrary $C \in \mathcal{E}$, $A \in \mathcal{F}^\mu_\infty$ and any $x \in E$ that

$$\mathbb{P}^{0,x}([X_0 \in C] \cap A) \;=\; \mathbb{P}^{0,x}[X_0 \in C]\,\mathbb{P}^{0,x}(A),$$

which completes the proof. □

Lemma 2.3.2. *The map $x \mapsto \mathbb{P}^{0,x}(A)$ is measurable for all $A \in \mathcal{F}$.*

Proof. We first note that the set of all $A \in \mathcal{F}$ such that $x \mapsto \mathbb{P}^{0,x}(A)$ is measurable forms a Dynkin system.[14] Hence, if we prove measurability for all A in an $\cap$-stable generator of $\mathcal{F}$ the Dynkin Class Theorem[15] implies Lemma 2.3.2.
Firstly, it is straightforward to see that $x \mapsto \mathbb{P}^{0,x}(A)$ is measurable for all $A \in \sigma(X_0)$: Let $C \in \mathcal{E}$ then $\mathbb{P}^{0,x}[X_0 \in C] = \mathbb{I}_C(x)$ which is measurable in x and hence $x \mapsto \mathbb{P}^{0,x}(A)$ is measurable for all $A \in X_0^{-1}(\mathcal{E})$, which is an $\cap$-stable generator of $\sigma(X_0)$. Hence due to the Dynkin Class Theorem measurability holds for all $A \in \sigma(X_0)$.
Secondly, we consider the σ-field $\mathcal{F}^\mu_\infty$ generated by the random counting measure. For any finite, increasing selection of indices $n_1, \ldots, n_m \in \{1, 2, \ldots\}$, $m \in \mathbb{N}$, it holds that

[13] A family of sets $\mathcal{S}$ is called $\cap$-*stable* if it is closed with respect to intersections, i.e., for A, $B \in \mathcal{S}$ it holds that $A \cap B \in \mathcal{S}$.

[14] A familiy of sets $\mathcal{D}$ is called a *Dynkin system* if (i) $\Omega \in \mathcal{D}$, (ii) A, $B \in \mathcal{D}$ and $A \subset B$ implies $B \backslash A \in \mathcal{D}$ and (iii) $A_n \in \mathcal{D}$ for all $n \geq 1$ and $A_n \subset A_{n+1}$ implies $\bigcup_{n\geq 1} A_n \in \mathcal{D}$ [8, 41].

[15] The *Dynkin Class Theorem* states that if $\mathcal{S}$ is an $\cap$-stable subset of a Dynkin class $\mathcal{D}$ then $\sigma(\mathcal{S}) \subset \mathcal{D}$, see [8, 41].

the mapping

$$x \mapsto \mathbb{P}^{0,x}\big[(\tau_{n_1}, Y_{n_1}) \in I_{n_1} \times C_{n_1}, \dots, \big(\tau_{n_m}, Y_{n_m}) \in I_{n_m} \times C_{n_m}\big] \tag{2.3.1}$$

is measurable, where each I_{n_j} is a Borel subset of $[0,\infty)$ and $B_{n_j} \in \mathcal{E}$ for all $j = 1,\dots,m$, as the probability in (2.3.1) satisfies by definition of the regular conditional distributions

$$\begin{aligned}
&\mathbb{P}^{0,x}\Big[(\tau_{n_1}, Y_{n_1}) \in I_{n_1} \times C_{n_1}, \dots, \big(\tau_{n_m}, Y_{n_m}) \in I_{n_m} \times C_{n_m}\big] = \\
&= \int_{(0,\infty)}\int_E \cdots \int_{I_{n_1}}\int_{C_{n_1}}\int_{(0,\infty)}\int_E \cdots \int_{I_{n_m}} \pi^{n_m}_{t_m}(y_{m-1}, C_{n_m})\, P^{n_m}\big((y_{n_m-1}, t_{n_m-1}), \mathrm{d}t_{n_m}\big) \dots \\
&\qquad \dots\, \pi^{n_1+1}_{t_{n_1+1}}(y_{n_1+1}, \mathrm{d}y_{n_1+1})\, P^{n_1+1}\big((y_{n_1}, t_{n_1}), \mathrm{d}t_{n_1+1}\big) \\
&\qquad\qquad \pi^{n_1}_{t_{n_1}}(y_{n_1}, \mathrm{d}y_{n_1})\, P^{n_1}\big((y_{n_1-1}, t_{n_1-1}), \mathrm{d}t_{n_1}\big) \dots\, \pi^0_{t_1}(x, \mathrm{d}y_1)\, P^0(x, \mathrm{d}t_1)\,,
\end{aligned}$$

which is measurable with respect to x due to the properties of the Markov kernels. Hence, the Dynkin class of all sets such that the measurability condition holds, contains all sets of the form

$$\big[(\tau_{n_1}, Y_{n_1}) \in I_{n_1} \times C_{n_1}, \dots, \big(\tau_{n_m}, Y_{n_m}) \in I_{n_m} \times C_{n_m}\big] \tag{2.3.2}$$

for any $m \in \mathbb{N}$, indices $n_1,\dots,n_m$ arbitrarily in $\{1,2,\dots\}$ and each I_{n_j}, C_{n_j} being any Borel set of $[0,\infty)$ or E, respectively. Clearly, the family of sets of the form (2.3.2) is $\cap$-stable. Further, as in part (c) of the proof of Theorem B.2.8 we find that each counting variable N^t_C, $t \geq 0$ and $C \in \mathcal{E}$, is measurable with respect to the σ-field generated by the sets (2.3.2). Thus, in particular, $\mathcal{F}^\mu_\infty$, which is by definition the smallest σ-field, such that all counting variables are measurable, is contained – actually equal – to the σ-field generated by the sets (2.3.2). Hence it follows by the Dynkin Class Theorem that $x \mapsto \mathbb{P}^{0,x}(A)$ is measurable for all $A \in \mathcal{F}^\mu_\infty$.
Thus, overall, we have shown that the measurability property holds for all sets in $\sigma(X_0)$ and $\mathcal{F}^\mu_\infty$.
Next, it is easy to see that the measurability property also holds for all sets of the form $A_1 \cap A_2$ with $A_1 \in \sigma(X_0)$ and $A_2 \in \mathcal{F}^\mu_\infty$ as due to indepence of the σ-fields $x \mapsto \mathbb{P}^{0,x}(A_1 \cap A_2) = \mathbb{P}^{0,x}(A_1)\mathbb{P}^{0,x}(A_2)$. The latter is a measurable function as it is a product of two measurable functions. Hence, as all such sets form an $\cap$-stable generator of $\mathcal{F}$ we infer that the masurability property holds for all $A \in \mathcal{F}$, which completes the proof. □

2.3.2 Proof of Theorem 2.1.1 (The strong Markov property)

The proof is split into three main parts. First we prove (a) for the process $(X_t)_{t\geq 0}$ the simple Markov property with respect to the filtration $(\mathcal{F}_t)_{t\geq 0}$ under the probability $\mathbb{P}^{0,x}$ for any $x \in E$. Then, (b) we infer the simple Markov property under the probability $\mathbb{P}^{\nu}$, where ν is an arbitrary initial law, using the definition of this probabilities by (2.1.2) and the results established in part (a). For later use, we also show, using the shift operators defined in Definition B.2.3, how the Markov property translates into conditions on the distribution of the random counting measure that defines a PDP. In the last part (c), we then prove the strong Markov property in terms of conditions on the random counting measure using the already established simple Markov property. This then implies the strong Markov property for the PDP.
We note that the central arguments of the proof establishing the simple and strong Markov property in parts (a) and (c), respectively, are adapted from [64, Thm. 7.3.2, 7.5.1] to the present setup. However, in [64] only processes with deterministic initial conditions, i.e., corresponding to laws $\mathbb{P}^{0,x}$, and adapted to the natural filtration of the random measure, i.e., corresponding to the Markov property with respect to the filtration $(\mathcal{F}^{\mu}_t)_{t\geq 0}$, are considered. Therefore, we have added to the proof the considerations at the beginning of part (a) and part (b) which allow to infer the Markov property with respect to the filtration $(\mathcal{F}_t)_{t\geq 0}$ under any law $\mathbb{P}^{\nu}$.

(a) For the simple Markov property to hold we have to show that the distribution of the process $(X_t)_{t\geq s}$ conditioned on $\mathcal{F}_s$ depends only on X_s for all $s \geq 0$. To this end we first recall the following statement from [8, Thm. 54.4]: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathcal{G}_1$ and $\mathcal{G}_2$ are sub-σ-fields of $\mathcal{F}$ and Y be an integrable random variable. Then, if $\sigma(Y) \vee \mathcal{F}_2$ is independent of $\mathcal{G}_1$ it holds almost surely that

$$\mathbb{E}[Y|\mathcal{G}_1 \vee \mathcal{G}_2] = \mathbb{E}[Y|\mathcal{G}_2]\,.$$

We apply this result to the random variable $Y = \mathbb{I}_{[X_t \in C]}$ for $C \in \mathcal{E}$ where $t \geq s$ and the σ-fields $\mathcal{G}_1 = \sigma(X_0)$ and $\mathcal{G}_2 = \mathcal{F}^{\mu}_s$, hence, $\mathcal{G}_1 \vee \mathcal{G}_2 = \mathcal{F}_s$. An analogous argument as employed in the proof of Lemma 2.3.1 shows that $\sigma(X_0)$ is independent of $\sigma(X_t) \vee \mathcal{F}^{\mu}_s$ under the measure $\mathbb{P}^{0,x}$. Hence, we can conclude that $\mathcal{F}^0$ is independent of $\sigma(X_t) \vee \mathcal{F}^{\mu}_s$ under the measure $\mathbb{P}^{0,x}$. Therefore it holds for all $0 \leq s \leq t$ that

$$\mathbb{P}^{0,x}[X_t \in C|\mathcal{F}_s] = \mathbb{P}^{0,x}[X_t \in C|\mathcal{F}^{\mu}_s]\,. \tag{2.3.3}$$

The conditional probability in the right hand side almost surely equals the conditional probability with respect to $\mathcal{F}^{\mu_x}_s$ of a PDP started at x (for all (!) trajectories) and defined via a random counting measure μ_x given by the kernels $(P^n_x)_{n\geq 0}$ and $(\pi^n_x)_{n\geq 0}$. For this additional process we can therefore proceed as in [64] and all results are immediately transferable to the original process $(X_t)_{t\geq 0}$.[16] Due to Corollary B.2.1 conditioning on $\mathcal{F}_s^{\mu_x}$ then amounts to conditioning on the events $[N_s = k, Z_k = z_k]$ for all $k \in \mathbb{N}_0$ and $z_k \in K^k$. Moreover, as the distribution of the process in this case is fully determined by the law of the random counting measure it suffices to consider the conditional distribution of the random counting measure $\theta_s\mu$. Given the event $[N_s = k, Z_k = z_k]$ this conditional distribution is due to Theorem B.2.2 given by the probability defined by the Markov kernels

$$P^0_{x|k,z_k} \;:=\; P^k_x(z_k, \cdot\,|(s,\infty]), \tag{2.3.4}$$

$$\pi^0_{x|k,z_k}(t,\cdot) \;:=\; \pi^k_x((z_k,t),\cdot) \quad \forall\, t \geq s, \tag{2.3.5}$$

and

$$P^n_{x|k,z_k}(\widetilde{z}_n, \cdot\,) \;:=\; P^{k+n}_x(z_k \sqcup \widetilde{z}_n, \cdot\,) \quad \forall\, n \geq 1, \tag{2.3.6}$$

$$\pi^n_{x|k,z_k}((\widetilde{z}_n, t), \cdot\,) \;:=\; \pi^{k+n}_x((z_k \sqcup \widetilde{z}_n, t), \cdot\,) \quad \forall\, n \geq 1,\, t > \widetilde{t}_n. \tag{2.3.7}$$

Note, that in addition to Theorem B.2.2 we may have an a-priori dependence of the Markov kernels on the initial condition x which we have accounted for in the notation employed in (2.3.4)–(2.3.7). Hence, establishing the simple Markov property for a PDP now amounts to showing the following: Firstly, the Markov kernels (2.3.4) and (2.3.5) and the measurable function

$$t \mapsto f^k_x(\,t\,|\,z_k),\, t \geq s \tag{2.3.8}$$

which, on the set of conditioning, defines the value of the process until the first jump after s, depend on $X_s = f^k_x(s|z_k)$ only. Secondly, we need to prove that the Markov kernels (2.3.6) and (2.3.7) and the measurable functions

$$t \mapsto f^{k+n}_x(\,t\,|z_k \sqcup \widetilde{z}_n) \quad \forall\, n \geq 1,\, t \geq \widetilde{t}_n \tag{2.3.9}$$

depend on k, z_k and x through y_k only.

We start with the condition on the measurable function (2.3.8) which by definition

[16] The insertion of this additional process starting at x for all trajectories is a technical necessity in order to employ in part (a) the method of proof of [64]. It presupposes that the process is adapted to $(\mathcal{F}^\mu_t)_{t\geq 0}$ and hence the initial condition is adapted to $\{\emptyset, \Omega\}$. Thus, we have two PDPs, one starting in x for all trajectories and the other starting in x for almost all trajectories and the random counting measure depends only on the initial condition. It is obvious that these two processes possess the same law (up to null sets) and their conditional probabilities with respect to the respective canonical filtration do coincide almost surely.

(2.1.4) satisfies

$$f_x^k(t|z_k) = \varphi_{t_k,t}(y_k) = \varphi_{s,t}(\varphi_{t_k,s}(y_k)) = \varphi_{s,t}(X_s) \tag{2.3.10}$$

and hence depends on X_s only. Next we consider for $t \geq s$ the survivor function for the Markov kernel (2.3.4) which satisfies

$$\begin{aligned}
P^0_{x|k,z_k}((t,\infty]) &= P_x^k\big(z_k,(t,\infty]\big|(s,\infty]\big) = \frac{P_x^k(z_k,(t,\infty]\cap(s,\infty])}{P_x^k(z_k,(s,\infty])} = \frac{\overline{S}^k_{z_n,x}(t)}{\overline{S}^k_{z_n,x}(s)} \\
&= \frac{S_{t_k,y_k}(t)}{S_{t_k,y_k}(s)} = \frac{S_{t_k,y_k}(s)\, S_{s,\varphi_{t_k,s}(y_k)}(t)}{S_{t_k,y_k}(s)} \\
&= S_{s,\varphi_{t_k,s}(y_k)}(t) \\
&= S_{s,X_s}(t)\,.
\end{aligned} \tag{2.3.11}$$

Here we have used the definition of the survivor function (2.1.6) for the Markov kernels $(P_x^n)_{n\geq 0}$ and the functional property of the survivor function (2.1.5). As survivor functions completely define probabilities it follows that $P^0_{x|k,z_k}$ depends on X_s only. Further, for $t \geq s$ the kernel (2.3.4) satisfies for all $C \in \mathcal{E}$ by definition (2.1.7)

$$\pi^0_{x|k,z_k}(t,C) = \pi_x^k((z_k,t),C) = r_t(\varphi_{t_k,t}(y_k),C) = r_t(\varphi_{s,t}(X_s),C)\,. \tag{2.3.12}$$

Therefore, also the conditional probability (2.3.5) depends on X_s only.

Next, completely analogous manipulations show that for all $n \geq 1$ the measurable function in (2.3.9) and the Markov kernels in (2.3.6) depend on $\widetilde{t}_n$ and $\widetilde{y}_n$ only and the Markov kernels in (2.3.7) depend on $\widetilde{t}_n$ and $\widetilde{y}_n$ and t only. In particular, all these quantities are independent of x and z_k which completes the proof of the simple Markov property. Further all considered quantities are independent of the initial condition x. Hence, it follows that the PDP possesses transition probabilities that are independent of the initial condition.

Finally, for a PDP to be a homogeneous Markov process the quantities (2.3.4)–(2.3.9) can depend on s, t only through the difference $t-s$. We immediately find that this holds inserting into (2.3.10)–(2.3.12) the definitions of the one-parameter semigroup $\varphi_{s,t} = \phi_{t-s}$, the survivor function $S_{s,y}(t) = S_y(t-s)$ and the Markov kernels $r_t(y,\cdot) = r(y,\cdot)$, cf. (2.1.8)–(2.1.11).

(b) So far we have shown that the PDP $(X_t)_{t\geq 0}$ satisfies the simple Markov property with respect to the filtration $(\mathcal{F}_t)_{t\geq 0}$ for every probability measure $\mathbb{P}^{0,x}$, $x \in E$. We next extend the simple Markov property to any measure $\mathbb{P}^\nu$. First, the simple Markov property states that for each $x \in E$ and for all $C \in \mathcal{E}$ and $s \leq t$ it holds almost surely with respect to the measure $\mathbb{P}^{0,x}$ that

$$\mathbb{P}^{0,x}[X_t \in C|\mathcal{F}_s] = p_{st}(X_s, C)$$

where $p_{s,t}$ are the transition probabilities of the process. Note that their existence is guaranteed and that they are independent of the initial condition $x \in E$ which was established in part (a) of the proof. In order to show that $(X_t)_{t\geq 0}$ is a Markov process with respect to $\mathbb{P}^\nu$ we have to show that $p_{st}(X_s, C)$ is a version of the conditional probability with respect to $\mathcal{F}_s$ under $\mathbb{P}^\nu$, i.e., for all $F \in \mathcal{F}_s$ it holds that

$$\int_F \mathbb{P}^\nu[X_t \in C|\mathcal{F}_s]\, \mathrm{d}\mathbb{P}^\nu = \int_F p_{st}(X_s, C)\, \mathrm{d}\mathbb{P}^\nu . \tag{2.3.13}$$

On the one hand, by definition of the measure $\mathbb{P}^\nu$ we obtain for the right hand side in (2.3.13) that

$$\int_F p_{st}(X_s, C)\, \mathrm{d}\mathbb{P}^\nu = \int_E \int_F p_{st}(X_s, C)\, \mathrm{d}\mathbb{P}^{0,x}\, \nu(\mathrm{d}x) .$$

On the other hand, due to the Markov property of $(X_t)_{t\geq 0}$ with respect to the measures $\mathbb{P}^{0,x}$ the left hand side in (2.3.13) yields

$$\begin{aligned}
\int_F \mathbb{P}^\nu[X_t \in C|\mathcal{F}_s]\, \mathrm{d}\mathbb{P}^\nu &= \int_F \mathbb{I}_{[X_t\in C]}\, \mathrm{d}\mathbb{P}^\nu \\
&= \int_E \int_F \mathbb{I}_{[X_t\in C]}\, \mathrm{d}\mathbb{P}^{0,x}\, \nu(\mathrm{d}x) \\
&= \int_E \int_F \mathbb{P}^{0,x}[X_t \in C|\mathcal{F}_s]\, \mathrm{d}\mathbb{P}^{0,x}\, \nu(\mathrm{d}x) \\
&= \int_E \int_F p_{st}(X_s, C)\, \mathrm{d}\mathbb{P}^{0,x}\, \nu(\mathrm{d}x) .
\end{aligned}$$

Hence, equality (2.3.13) holds and the Markov property with respect to the measure $\mathbb{P}^\nu$ is established.

Finally, we comment on how the Markov properties are presented in terms of the random counting measure μ the PDP $(X_t)_{t\geq 0}$ is constructed from. In particular, in part (a) it is shown that the conditional probability with respect to $\mathcal{F}_s$ of $\theta_s\mu$ depends on X_s only, where θ_s is the shift operator on H as given in Definition B.2.3.

For all $s \geq 0$ and all $x \in E$ let $\mathbb{P}^{s,x}$ denote the probability for μ which is generated by the survivor functions and Markov kernels

$$\overline{S}^{0}_{s,x}(t) := \begin{cases} 1 & t < s \\ S_{s,x}(t) & s \leq t, \end{cases} \qquad \pi^{0}_{t|s,x} := r_t(\varphi_{s,t}(x), \cdot\,),$$

and for $n \geq 1$, $t \geq t_n$

$$\overline{S}^{n}_{z_n|s,x}(t) := S_{t_n,y_n}(t), \qquad \pi^{n}_{z_n,t|s,x} := r_t(\varphi_{t_n,t}(y_n), \cdot\,).$$

That is $\mathbb{P}^{s,x}$ is the probability of a random counting measure / PDP such that there is almost surely no point / jump up to time $s \geq 0$ and, in addition, the PDP is at position x at time s almost surely, i.e., $\mathbb{P}^{s,x}[X_s = x] = 1$. Then the simple Markov property can be expressed as

$$\mathbb{P}^{\nu}[\theta_s \mu \in A \,|\, \mathcal{F}_s] = \mathbb{P}^{s,X_s}[\mu \in A] \quad \forall\, A \in \mathcal{H}. \tag{2.3.14}$$

Moreover, in the homogeneous case set $\mathbb{P}^{x} := \mathbb{P}^{0,x}$, i.e., for the random counting measure μ this is the distribution defined by the Markov kernels

$$\overline{S}^{0}_{s,x}(t) := \begin{cases} 1 & t < s, \\ S_x(t-s) & s \leq t, \end{cases} \qquad \pi^{0}_{t|s,x} := r(\phi_{t-s}(x), \cdot\,),$$

and for $n \geq 1$, $t \geq t_n$

$$\overline{S}^{n}_{z_n|s,x}(t) := S_{y_n}(t - t_n), \qquad \pi^{n}_{z_n,t|s,x} := r(\phi_{t-t_n}(y_n), \cdot\,).$$

This yields the general, homogeneous Markov property

$$\mathbb{P}^{\nu}[\theta^{*}_{t} \mu \in A | \mathcal{F}_t] = \mathbb{P}^{X_t}[\mu \in A] \quad \forall\, A \in \mathcal{H}, \tag{2.3.15}$$

where θ^{*}_{t} denotes the translated shift operator defined in (B.2.13). Note that the Markov properties (2.3.14) and(2.3.15) hold simultaneously for all $A \in \mathcal{H}$ such that

$$\mathbb{P}^{\nu}[\theta_t \mu \in \cdot\, | \mathcal{F}_t] = \mathbb{P}^{t,X_t}[\mu \in \cdot\,] \quad \text{and} \quad \mathbb{P}^{\nu}[\theta^{*}_{t} \mu \in \cdot\, | \mathcal{F}_t] = \mathbb{P}^{X_t}[\mu \in \cdot\,].$$

Thus in each case these are the same probability measures on $(H, \mathcal{H})$ for all ω outside a null set.

(c) In this part of the proof we show the strong Markov property (2.0.5) for the PDMP $(X_t)_{t \geq 0}$. First, note that the Markov properties in terms of the random counting measure, i.e., (2.3.14) and(2.3.15), yield a straightforward definition of a strong Markov property in terms of the random counting measure: A PDP constructed from a random counting measure μ satisfies the strong Markov property if it holds for any $\mathcal{F}_t$-stopping time T on $[T < \infty]$ almost surely (with respect to $\mathbb{P}^\nu$) that

$$\mathbb{P}^\nu[\theta_T\mu \in A|\mathcal{F}_T] = \mathbb{P}^{T,X_T}[\mu \in A] \quad \forall\, A \in \mathcal{H}\,. \tag{2.3.16}$$

In case of a homogeneous PDP this reduces to

$$\mathbb{P}^\nu[\theta^*_T\mu \in A|\mathcal{F}_T] = \mathbb{P}^{X_T}[\mu \in A] \quad \forall\, A \in \mathcal{H} \tag{2.3.17}$$

almost surely (with respect to $\mathbb{P}^\nu$) on $[T < \infty]$.

Therefore, in order to prove the strong Markov property for the PDP it suffices to show that condition (2.3.16) holds for any $\mathcal{F}_t$-stopping time T. Next we invoke the equivalence of random counting measures and marked point processes, cf. Prop. B.2.1. Then, according to Jacobsen [64, p. 168] the property (2.3.16) is in turn implied if for all $n \in \mathbb{N}$ and all measurable, bounded functions $g_i : [0,\infty] \times \overline{E} \to \mathbb{R}$, $i \leq n$, it holds that

$$\mathbb{E}^\nu\Big[\prod_{i=1}^n g_i(\tau_i, Y_n) \circ \theta_T\mu \,\Big|\, \mathcal{F}_T\Big] \;=\; \mathbb{E}^{T,X_T}\Big[\prod_{i=1}^n g_i(\tau_i, Y_n)\Big]\,, \tag{2.3.18}$$

i.e., the term in the right hand side is a version of the conditional expectation with respect to the stopped σ-field $\mathcal{F}_T$. In (2.3.18) the expectation operators $\mathbb{E}^\nu$ and $\mathbb{E}^{t,x}$ denote integration with respect to the probabilities $\mathbb{P}^\nu$ and $\mathbb{P}^{t,x}$, respectively.

We now prove property (2.3.18) by induction with respect to n. Thus, proceeding along a standard approach to establish a strong Markov property, we first (c.1) prove the strong Markov property directly for stopping times taking only countably many values and approximating an arbitrary stopping time from above. This result is then used to (c.2) derive the initial condition for the induction and finally in (c.3) we prove the induction step. In the following let T be an arbitrary but fixed $\mathcal{F}_t$-stopping time.

(c.1) We define a sequence of stopping times T_n taking only countably many values that approximates T from above by

$$T_n := \sum_{i=1}^\infty \frac{i}{2^n}\, \mathbb{I}_{\left[\frac{i-1}{2^n} \leq T < \frac{i}{2^n}\right]} + \infty\, \mathbb{I}_{[T=\infty]}\,.$$

Obviously, $T \leq T_n$ by definition and even $T_n < T$ on the set $[T < \infty]$. Furthermore, it holds that $[\frac{i-1}{2^n} \leq T] \cap [T < \frac{i}{2^n}] = [T_n = \frac{i}{2^n}] \in \mathcal{F}_{i/2^n}$.

We now show that the strong Markov property holds for all stopping times T_n. To this end we prove that for all $A \in \mathcal{H}$ the random variable $\mathbb{P}^{T_n,X_{T_n}}[\mu \in A]$ is a version of the conditional expectation $\mathbb{E}^\nu[\mathbb{I}_{[\theta_{T_n}\mu\in A]}|\mathcal{F}_{T_n}]$. Let $F \in \mathcal{F}_{T_n}$ and define the sets

$F_i = F \cap \left[T_n = \frac{i}{2^n}\right]$. Then $F_i \in \mathcal{F}_{i/2^n}$ and

$$F = \bigcup_{i=1}^{\infty} F_i$$

is a countable union of disjoint elements of $\mathcal{F}$. Further it holds that

$$\begin{aligned}\int_{F_i} \mathbb{P}^{T_n, X_{T_n}}[\mu \in A] \,\mathrm{d}\mathbb{P}^\nu &= \int_{F_i} \mathbb{P}^{\frac{1}{2^n}, X_{i/2^n}}[\mu \in A] \,\mathrm{d}\mathbb{P}^\nu \\ &= \int_{F_i} \mathbb{I}_{[\theta_{i/2^n}\mu \in A]} \,\mathrm{d}\mathbb{P}^\nu \\ &= \mathbb{P}^\nu(F_i \cap [\theta_{T_n}\mu \in A])\,,\end{aligned}$$

where the second equality is due to the simple Markov property (2.3.14). Hence due to the σ-addidivity of a probability measure it follows that

$$\int_F \mathbb{P}^{T_n, X_{T_n}}(A) \,\mathrm{d}\mathbb{P}^\nu = \mathbb{P}^\nu(F \cap [\theta_{T_n}\mu \in A]) = \int_F \mathbb{I}_{[\theta_{T_n}\mu \in A]} \,\mathrm{d}\mathbb{P}^\nu\,. \tag{2.3.19}$$

Since (2.3.19) is valid for all $F \in \mathcal{F}_{T_n}$ it follows by definition of the conditional expectation that on the set $[T_n < \infty] = [T < \infty]$ it holds

$$\mathbb{P}^\nu\left[\theta_{T_n}\mu \in A \,\middle|\, \mathcal{F}_{T_n}\right] = \mathbb{E}^\nu\left[\mathbb{I}_{[\theta_{T_n}\mu \in A]} \middle| \mathcal{F}_{T_n}\right] = \mathbb{P}^{T_n, X_{T_n}}[\mu \in A]\,. \tag{2.3.20}$$

As (2.3.20) holds simultaneously for all $A \in \mathcal{H}$ we have shown the strong Markov property (2.3.16) for all T_n, $n \geq 1$, i.e.,

$$\mathbb{P}^\nu\left[\theta_{T_n}\mu \in \cdot \,\middle|\, \mathcal{F}_{T_n}\right] = \mathbb{P}^{T_n, X_{T_n}}[\mu \in \cdot\,].$$

(c.2) We next use the strong Markov property with respect to the stopping times T_n, $n \geq 1$, to prove the initial condition of the induction for the stopping time T. On $[T < \infty]$ it is valid that

$$\mathbb{E}^{T, X_T}[g_1(\tau_1, Y_1)] = \int_{(T,\infty]} \int_E g_1(t, y)\, r_t(\varphi_{T,t}(X_T), \mathrm{d}y)\, F_{T, X_T}(\mathrm{d}t)\,, \tag{2.3.21}$$

where $F_{T,X_T} = 1 - S_{T,X_T}$ is the cumulative distribution function of the first jump time. Similarly on $[T_n < \infty] = [T < \infty]$ it holds that

$$\mathbb{E}^{T_n, X_{T_n}}[g_1(\tau_1, Y_1)] = \int_{(T_n,\infty]} \int_E g_1(t, y)\, r_t(\varphi_{T_n,t}(X_{T_n}), \mathrm{d}y)\, F_{T_n, X_{T_n}}(\mathrm{d}t)\,.$$

Further, it holds on the set $[N_T = N_{T_n}] \cap [T < \infty]$ that the path starting in X_T at time T satisfies due to (2.1.4) for all $t \geq T_n$ until the first, subsequent jump

$$\varphi_{T,t}(X_T) = \varphi_{T_n,t}\big(\varphi_{T,T_n}(X_T)\big) = \varphi_{T_n,t}(X_{T_n}),$$

and the survivor function of the first jump after T, which on this set is also after T_n, satisfies due to (2.1.5)

$$S_{T,X_T}(t) = S_{T,X_T}(T_n)\, S_{T_n,X_{T_n}}(t).$$

Therefore, we obtain on the set $[N_T = N_{T_n}] \cap [T < \infty]$ from (2.3.21) that

$$\begin{aligned} \mathbb{E}^{T_n,X_{T_n}}[g_1(\tau_1, Y_1)] &= \int_{(T_n,\infty]} \int_E g_1(t,y)\, r_t(\phi_{T_n,t}(X_{T_n}), \mathrm{d}y)\, F_{T_n,X_{T_n}}(\mathrm{d}t) \\ &= \frac{1}{S_{T,X_T}(T_n)} \int_{(T_n,\infty]} \int_E g_1(t,y)\, r_t(\phi_{T,t}(X_T), \mathrm{d}y)\, F_{T,X_T}(\mathrm{d}t). \end{aligned}$$

As $n \to \infty$ implies $T_n \downarrow T$ and $S_{T,X_T}(T_n) \to 1$ pointwise for all $\omega \in \Omega$, we obtain that on the set $[N_T = N_{T_n}] \cap [T < \infty]$ it holds pointwise

$$\lim_{n\to\infty} \mathbb{E}^{T_n,X_{T_n}}[g_1(\tau_1, Y_1)] = \mathbb{E}^{T,X_T}[g_1(\tau_1, Y_1)]. \tag{2.3.22}$$

Let $F \in \mathcal{F}_T \cap [T < \infty]$ and set $F_n := F \cap [N_{T_n} = N_T]$. Therefore, as $\mathbb{I}_{F_n}$ is dominated by $\mathbb{I}_{[T<\infty]}$, $\mathbb{E}^{T_n,X_{T_n}}[g_1(\tau_1, Y_1)]$ is a bounded random variable as g_1 is bounded and since $\lim_{n\to\infty} \mathbb{I}_{[N_{T_n}=N_t]\cap[T<\infty]} = \mathbb{I}_{[T<\infty]}$ holds pointwise, it follows from (2.3.22) that

$$\lim_{n\to\infty} \mathbb{E}^{T_n,X_{T_n}}[g_1(\tau_1, Y_1)]\, \mathbb{I}_{F_n} = \mathbb{E}^{T,X_T}[g_1(\tau_1, Y_1)]\, \mathbb{I}_F$$

holds pointwise on Ω. Moreover, the sequence $\mathbb{E}^{T_n,X_{T_n}}[g_1(\tau_1, Y_1)]\, \mathbb{I}_{F_n}$ is dominated, thus the dominated convergence theorem yields

$$\begin{aligned} \lim_{n\to\infty} \int_{F_n} \mathbb{E}^{T_n,X_{T_n}}[g_1(\tau_1, Y_1)]\, \mathrm{d}\mathbb{P}^\nu &= \lim_{n\to\infty} \int_\Omega \mathbb{E}^{T_n,X_{T_n}}[g_1(\tau_1, Y_1)]\, \mathbb{I}_{F_n}\, \mathrm{d}\mathbb{P}^\nu \\ &= \int_\Omega \mathbb{E}^{T,X_T}[g_1(\tau_1, Y_1)]\, \mathbb{I}_F\, \mathrm{d}\mathbb{P}^\nu \\ &= \int_F \mathbb{E}^{T,X_T}[g_1(\tau_1, Y_1)]\, \mathrm{d}\mathbb{P}^\nu. \end{aligned} \tag{2.3.23}$$

Next, as we have established the strong Markov property for the stopping times T_n

in part (c.1) and as $\mathcal{F}_T \subset \mathcal{F}_{T_n}$ implies $F_n \in \mathcal{F}_{T_n}$, we obtain due to (2.3.23)

$$\begin{aligned}\lim_{n\to\infty} \int_{F_n} g_1(\tau_{T_n,1}, Y_{T_N,1})\,\mathrm{d}\mathbb{P}^\nu &= \lim_{n\to\infty} \int_{F_n} g_1(\tau_1, Y_1) \circ \theta_{T_n}\mu\,\mathrm{d}\mathbb{P}^\nu \\ &= \int_F \mathbb{E}^{T,X_T}[g_1(\tau_1, Y_1)]\,\mathrm{d}\mathbb{P}^\nu\,. \qquad (2.3.24)\end{aligned}$$

Finally, as $(\tau_{T_n,1}, Y_{T_N,1}) \to (\tau_{T,1}, Y_{T,1})$ holds pointwise on Ω and $g_1(\tau_{T_n,1}, Y_{T_N,1})$ is bounded, the dominated convergence theorem yields

$$\lim_{n\to\infty} \int_{F_n} g_1(\tau_{T_n,1}, Y_{T_N,1})\,\mathrm{d}\mathbb{P}^\nu = \int_F g_1(\tau_{T,1}, Y_{T,1})\,\mathrm{d}\mathbb{P}^\nu\,. \qquad (2.3.25)$$

Therefore it follows due to (2.3.24) and (2.3.25) that

$$\int_F \mathbb{E}^{T,X_T}[g_1(\tau_1, Y_1)]\,\mathrm{d}\mathbb{P}^\nu = \int_F g_1(\tau_{T,1}, Y_{T,1})\,\mathrm{d}\mathbb{P}^\nu\,,$$

which is valid for all $F \in \mathcal{F}_T \cap [T < \infty]$. Hence, due to the definition of the conditional expectation we obtain

$$\mathbb{E}^\nu\big[g_1(\tau_{T,1}, Y_{T,1})\,\big|\,\mathcal{F}_T\big] = \mathbb{E}^{T,X_T}[g_1(\tau_1, Y_1)]\,.$$

Thus the initial condition for the induction is proved, i.e., equality (2.3.18) holds for $n = 1$.

(c.3) We assume the induction hypothesis (2.3.18) holds for $n-1$, $n \geq 2$, that is,

$$\mathbb{E}^\nu\Big[\prod_{i=1}^{n-1} g_i(\tau_i, Y_n) \circ \theta_T\mu\,\Big|\,\mathcal{F}_T\Big] = \mathbb{E}^{T,X_T}\Big[\prod_{i=1}^{n-1} g_i(\tau_i, Y_n)\Big]\,. \qquad (2.3.26)$$

Recall that $T_{T,n-1}$, i.e., the time of the $(n-1)$th jump after time T, is a stopping time and hence for $U := \prod_{i=1}^{n-1} g_i(\tau_i, Y_i)$ it follows due to (2.3.26) that

$$\begin{aligned}\mathbb{E}^\nu\big[(U\circ\theta_T\mu)\,g_n(T_{T,n}, Y_{T,n})\big|\mathcal{F}_{T_{T,n-1}}\big] &= (U\circ\theta_T\mu)\,\mathbb{E}^\nu\big[g_n(T_{T,n}, Y_{T,n})\big|\mathcal{F}_{T_{T,n-1}}\big] \\ &= (U\circ\theta_T\mu)\,\mathbb{E}^{T_{T,n-1},X_{T_{T,n-1}}}\big[g_n(\tau_1, Y_1)\big]. \qquad (2.3.27)\end{aligned}$$

Note that $U\circ\theta_T\mu$ is clearly $\mathcal{F}_{T_{T,n-1}}$-measurable. As $T \leq T_{T,n-1}$ it follows that $\mathcal{F}_T \subset \mathcal{F}_{T_{T,n-1}}$ and thus by (2.3.27)

$$\mathbb{E}^\nu\big[(U\circ\theta_T\mu)\,g_n(T_{T,n}, Y_{T,n})\,\big|\,\mathcal{F}_T\big] = \mathbb{E}^\nu\big[(U\circ\theta_T\mu)\,\mathbb{E}^{T_{T,n-1},X_{T_{T,n-1}}}\big[g_n(\tau_1, Y_1)\big]\,\big|\,\mathcal{F}_T\big]\,. \qquad (2.3.28)$$

Note that the inner expectation in the right hand side of (2.3.28) is a bounded, measurable function of $(\tau_{n-1}, Y_{n-1}) \circ \theta_T \mu$. Hence the induction hypotheses yields that

$$\mathbb{E}^{\nu}\Big[(U \circ \theta_T \mu)\, \mathbb{E}^{T_{T,n-1}, X_{T_{T,n-1}}}\big[g_n(\tau_1, Y_1)\big] \,\Big|\, \mathcal{F}_T\Big] \;=\; \mathbb{E}^{T,X_T}\Big[U\, \mathbb{E}^{\tau_{n-1}, Y_{n-1}}\big[g_n(\tau_1, Y_1)\big]\Big]\,. \tag{2.3.29}$$

Now, the inner expectation in the right hand side of (2.3.29) is a version of the conditional expectation with respect to the probability $\mathbb{P}^{T,X_T}$, i.e., with respect to the conditional probability $\mathbb{P}^{T,X_T}$ it holds that

$$\mathbb{E}^{\tau_{n-1}, Y_{n-1}}\big[g_n(\tau_1, Y_1)\big] \;=\; \mathbb{E}^{T,X_T}\big[g_n(\tau_n, Y_n) \,|\, \mathcal{F}_{T_{n-1}}\big] \quad \text{a.s.},$$

where T_{n-1} is understood as the $(n-1)$th jump time of the probability $\mathbb{P}^{T,X_T}$. Thus overall we obtain

$$\begin{aligned}
\mathbb{E}^{\nu}\Big[(U \circ \theta_T \mu)\, \mathbb{E}^{T_{T,n-1}, X_{T_{T,n-1}}}\big[g_n(\tau_1, Y_1)\big] \,\Big|\, \mathcal{F}_T\Big] \;&=\; \mathbb{E}^{T,X_T}\Big[U\, \mathbb{E}^{T,X_T}\big[g_n(\tau_n, Y_n) \,|\, \mathcal{F}_{T_{n-1}}\big]\Big] \\
&=\; \mathbb{E}^{T,X_T}\Big[U\, g_n(\tau_n, Y_n)\Big] \\
&=\; \mathbb{E}^{T,X_T}\Big[\prod_{i=1}^{n} g_i(\tau_i, Y_n)\Big]\,.
\end{aligned}$$

Hence, the induction step is proved and consequently the proof of Theorem 2.1.1 is completed.

2.3.3 Proof of Theorem 2.2.1 (The extended generator)

To characterise the extended generator we have to show that, on the one hand, for functions f satisfying conditions (i)–(iii) of Theorem 2.2.1 the process defined by

$$M_t^f := f(X_t) - f(X_0) - \int_0^t \mathcal{A}f(X_s)\,\mathrm{d}s \tag{2.3.30}$$

is an $\mathcal{F}_t$-local martingale. And, on the other hand, functions f for which (2.3.30) is an $\mathcal{F}_t$-local martingale necessarily meet the conditions specified in the theorem. The proof we present subsequently follows along the line of the corresponding proof in Davis [37, Thm. 26.14]. We have adapted the method of proof to the more general set-up for PDMPs we employ and to the construction of PDMPs from random counting measures following Jacobsen [64].

We now first provide arguments that allow to simplify the problem. Note that due to analogous considerations as in part (b) of the proof of Theorem 2.1.1 we can first establish the local martingale property with respect to the probability measures $\mathbb{P}^{0,x}$, for all $x \in \operatorname{supp}\nu$. Once this is established, integration with respect to the initial ν law establishes the martingale property with respect to the law $\mathbb{P}^{\nu}$. Moreover, under the laws $\mathbb{P}^{0,x}$ conditional probabilities with respect to elements of $\mathcal{F}_t$ equal almost surely the conditional probabilities with respect to $\mathcal{F}_t^\mu$, cf. part (a) of the proof of Theorem 2.1.1. Hence, in a further reduction of the complexity of the problem it is sufficient to consider $\mathcal{F}_t^\mu$-local martingales. That is, in order to characterise the extended generator it is sufficient to prove that the conditions (i)–(iii) of Theorem 2.2.1 are (a) sufficient and (b) necessary for the process (2.3.30) to be a local martingale with respect to the filtration $(\mathcal{F}_t^\mu)_{t\geq 0}$ under the probability $\mathbb{P}^{0,x}$ for every $x \in \operatorname{supp}\nu$.

(a) In this part of the proof we show that conditions (i)–(iii) imply the local martingale property for the process (2.3.30), i.e., the conditions are sufficient. We assume that (iii) holds, i.e., $(t, x, \omega) \mapsto f(x) - f(X_{t-}(\omega))$ is a valid integrand for the $\mathcal{F}_t^\mu$-predictable compensator $\widetilde{N}_t$ which is for the considered class of survivor functions (2.2.1) given by (2.1.19). Then due to Theorem B.2.6(b) it follows that the compensated stochastic integral (B.2.26) is given by

$$
\begin{aligned}
\int_{(0,t]\times E} f(x) - f(X_{s-})\, M(\mathrm{d}t, \mathrm{d}x) \;&= \\
= \;& \int_{(0,t]\times E} f(x) - f(X_{s-})\, \mu(\mathrm{d}s, \mathrm{d}x) - \int_{(0,t]\times E} f(x) - f(X_{s-})\mathrm{d}\widetilde{N}_s(\mathrm{d}s, \mathrm{d}x) \\
= \;& \sum_{\tau_i \leq t} \big(f(X_{\tau_i}) - f(X_{\tau_i -})\big) - \int_0^t \int_E \big(f(x) - f(X_{s-}\big)\, r(X_s, \mathrm{d}x)\, \lambda(X_s)\mathrm{d}s \\
& - \sum_{\tau_i \leq t} \int_E f(x) - f(X_{\tau_i -})\, r(X_{\tau_i -}, \mathrm{d}x)\, \mathbb{I}_{[X_{\tau_i -} \in B]} \qquad (2.3.31)
\end{aligned}
$$

is an $\mathcal{F}_t^\mu$-local martingale. Note that in the Lebesgue integral in the right hand side of (2.3.31) we can change the integrand from left to right continuous, i.e., we integrate with respect to s instead of $s-$, leaving the value of the integral unchanged as X_s and $\lambda(X_s)$ are càdlàg. Next we show that the right hand side of (2.3.31) is of the form (2.3.30) which implies that $f \in \mathcal{D}(\mathcal{A})$.

First note that due to the boundary condition (2.2.5) the last sum in the right hand side of (2.3.31) vanishes. Further on the set $[N_t = k]$ the first sum in the right hand side of (2.3.31) can be expanded to

$$
f(X_t) - f(X_0) - \Big[f(X_t) - f(X_{\tau_k}) + \sum_{i=1}^{k} \Big(f(X_{\tau_i -}) - f(X_{\tau_{i-1}}) \Big) \Big]. \qquad (2.3.32)
$$

To evaluate the terms in the square brackets recall from the definition of the PDMP (2.1.4) that

$$
X_{\tau_i -} = \phi_{\tau_i - \tau_{i-1}}(X_{\tau_{i-1}}) \,,
$$

and thus, as the map $t \mapsto f(\phi_t(x))$ is a.s. path-differentiable on $[0, t^*(x))$ due to condition (i), we immediately obtain

$$f(X_{\tau_i-}) - f(X_{\tau_{i-1}}) = \int_{\tau_{i-1}}^{\tau_i} \frac{\mathrm{d}}{\mathrm{d}s} f(X_s)\,\mathrm{d}s\,.$$

As we sum integrals over all the intervals between successive transitions in (2.3.32) the right hand side of (2.3.31) finally equals

$$f(X_t) - f(X_0) - \int_0^t \Big[\frac{\mathrm{d}}{\mathrm{d}s} f(X_s) + \lambda(X_s) \int_E \big(f(x) - f(X_s)\big)\, p(X_s, \mathrm{d}x)\Big]\,\mathrm{d}s\,, \quad (2.3.33)$$

which holds on all sets $[N_t = k]$, $k \geq 1$. This is an $\mathcal{F}_t^\mu$-local martingale of the form (2.3.30) and hence, the conditions of Theorem 2.2.1 are sufficient for f to be in the domain of the extended generator with extended generator given by (2.2.6).

(b) In this part we show that the conditions (i)–(iii) are also necessary. Let $f \in \mathcal{D}(\mathcal{A})$ which means that the process

$$f(X_t) - f(X_0) - \int_0^t K(X_s)\,\mathrm{d}s \quad (2.3.34)$$

is an $\mathcal{F}_t^\mu$–local martingale with K denoting the extended generator applied to f. From the Martingale Representation Theorem, Theorem B.2.6(a), it follows that there exists an $\mathcal{F}_t^\mu$-predictable integrand $k(s, x, \omega)$ such that (2.3.34) equals almost surely the stochastic integral

$$\begin{aligned} \int_{(0,t]\times E} k(s,x)\, M(\mathrm{d}s, \mathrm{d}x) \quad &= \quad \sum_{\tau_i \leq t} k(\tau_i, X_{\tau_1}) - \int_0^t \int_E k(s,x)\, r(X_{s-}, \mathrm{d}x)\, \lambda(X_s)\,\mathrm{d}s \\ &\quad - \sum_{\tau_i \leq t} \int_E k(\tau_i, x)\, r(X_{\tau_i-}, \mathrm{d}x)\, \mathbb{I}_{[X_{\tau_i-} \in B]}\,. \end{aligned} \quad (2.3.35)$$

Note that the process (2.3.34) is zero for $t = 0$, hence $M_0 = 0$ in Theorem B.2.6(a). As (2.3.34) and (2.3.35) are equal, their jumps have to be equal. The integral in (2.3.34) is continuous therefore jumps can only occur when X_t jumps, i.e., only at times $t = \tau_i$, $i \geq 1$. In the case of the process (2.3.34) the jump heights are given by

$$f(X_t) - f(X_{t-}) \quad (2.3.36)$$

whereas in the case of the process (2.3.35) they are given by

$$k(t, X_t) - \int_E k(t,x)\, r(X_{t-}, \mathrm{d}x)\, \mathbb{I}_{[X_{t-} \in B]}\,. \quad (2.3.37)$$

Thus on the set $[X_{t-} \notin B]$ we obtain that $k(t, X_t) = f(X_t) - f(X_{t-})$ which implies that $k(t, x) = f(x) - f(X_{t-})$ for all (t, x) outside a set $G \subset [0, \infty) \times E$ to which the process never jumps, i.e., $\mathbb{E}\mu(G) = 0$.
Comparing the jumps heights (2.3.36) and (2.3.37) on the set $[X_{t-} \in B]$ yields

$$f(X_t) - f(X_{t-}) = k(t, X_t) - \int_E k(t, x)\, r(X_{t-}, \mathrm{d}x)\,,$$

which in turn yields

$$f(y) - f(X_{t-}) = k(t, y) - \int_E k(t, x)\, r(X_{t-}, \mathrm{d}x) \tag{2.3.38}$$

for all y except on a set $A \in \mathcal{E}$ such that $r(X_{t-}, A) = 0$. We integrate both sides of equation (2.3.38) with respect to the probability $r(X_{t-}, \mathrm{d}y)$ of the post-jump value and obtain

$$\int_E f(y)\, r(X_{t-}, \mathrm{d}y) - f(X_{t-}) = \int_E k(t, y)\, r(X_{t-}, \mathrm{d}y) - \int_E k(t, x)\, r(X_{t-}, \mathrm{d}x)\,.$$

As the right hand side vanishes it follows that f satisfies the boundary condition (2.2.5). Next we fix $z \in B^+ \cap \partial B$. Then the boundary condition yields

$$\begin{aligned} k(t, x) - (f(x) - f(z)) &= k(t, x) - \big(k(t, x) - \int_E k(t, y)\, r(z, \mathrm{d}y)\big) \\ &= \int_E k(t, y) - (f(y) - f(z))\, r(z, \mathrm{d}y)\,, \end{aligned}$$

where the term in the right hand side is independent of x. Thus there exists a predictable process $\widehat{k}$ such that

$$\widehat{k}(t, z) = k(t, x) - (f(x) - f(z)).$$

and a combination with the result on the set $[X_{t-} \notin B]$ yields that in general

$$k(t, x) = f(x) - f(X_{t-}) + \widehat{k}(t, X_{t-})\, \mathbb{I}_{[X_{t-} \in B]}\,. \tag{2.3.39}$$

We insert expression (2.3.39) into the right hand side of (2.3.35) and obtain

$$
\begin{aligned}
(2.3.35) \quad = \quad & \sum_{\tau_i \le t} \big[f(X_{\tau_i}) - f(X_{\tau_i -}) \big] + \sum_{\tau_i \le t} \widehat{k}(\tau_i, X_{\tau_i -}) \, \mathbb{I}_{[X_{\tau_i -} \in B]} \\
& - \int_0^t \int_E f(x) - f(X_{s-}) \, r(X_{s-}, \mathrm{d}x) \, \lambda(X_s) \, \mathrm{d}s \\
& - \int_0^t \int_E \widehat{k}(s, X_{s-}) \, \mathbb{I}_{[X_{s-} \in B]} \, r(X_{s-}, \mathrm{d}x) \, \lambda(X_s) \, \mathrm{d}s \\
& - \sum_{\tau_i \le t} \int_E f(x) - f(X_{t-}) \, r(X_{\tau_i -}, \mathrm{d}x) \, \mathbb{I}_{[X_{\tau_i -} \in B]} \\
& - \sum_{\tau_i \le t} \int_E \widehat{k}(\tau_i, X_{\tau_i -}) \, \mathbb{I}_{[X_{\tau_i -} \in B]} \, r(X_{\tau_i -}, \mathrm{d}x) \, \mathbb{I}_{[X_{\tau_i -} \in B]} \\
= \quad & \sum_{\tau_i \le t} \big[f(X_{\tau_i}) - f(X_{\tau_i -}) \big] - \int_0^t \int_E f(x) - f(X_{s-}) \, r(X_{s-}, \mathrm{d}x) \, \lambda(X_s) \, \mathrm{d}s \\
& - \sum_{\tau_i \le t} \int_E f(x) - f(X_{t-}) \, r(X_{\tau_i -}, \mathrm{d}x) \, \mathbb{I}_{[X_{\tau_i -} \in B]} \, ,
\end{aligned}
$$

which is independent of $\widehat{k}$. Thus we can choose $\widehat{k} = 0$ and obtain for the structure of the integrand k that

$$
k(x, t, \omega) = f(x) - f(X_{t-}(\omega)) \, ,
$$

which holds for all $x \in E$ except on a set the process does not jump to almost surely. This in particular implies that $(t, x, \omega) \mapsto f(x) - f(X_{t-}(\omega))$ is a valid integrand for the compensating measure. Therefore condition (iii) is satisfied.

Finally, it remains to show condition (i). To this end we consider the sample paths of the processes (2.3.34) and (2.3.35) for $t \in [0, \tau_1)$ and obtain that almost surely

$$
\begin{aligned}
& f(\phi_t(X_0)) - f(X_0) - \int_0^t K(\phi_s(X_0)) \mathrm{d}s \; = \\
& \qquad = \; - \int_0^t \int_E \big(f(x) - f(\phi_s(X_0)\big) \, r(\phi_s(X_0), \mathrm{d}x) \big) \, \lambda(\phi_s(X_0)) \mathrm{d}s \, .
\end{aligned}
$$

As $\mathbb{P}[\tau_1 > t] > 0$ for all $0 \le t < t^*(X_0)$, we infer that the mapping $t \mapsto f(\phi_t(X_0))$ is absolutely continuous, hence f is a.e. path-differentiable on $[0, t^*(X_0))$. We can repeat this comparison for all inter-jump intervals $[\tau_i, \tau_{i+1})$, $i > 1$, and obtain that $t \mapsto f(\phi_t(x)$ is a.e. path-differentiable on $[0, t^*(x))$ for all x except on a set $A \in \mathcal{E}$ where the process never jumps to, i.e., $\mathbb{E}\mu([0, \infty) \times A) = 0$. Hence, condition (i) characterising the domain of the extended generator is satisfied.

We thus obtain for K that

$$\int_0^t K(X_t)\,\mathrm{d}t =$$

$$= f(X_t) - f(X_{\tau_{N_t}}) + \sum_{\tau_i \le t} f(X_{\tau_i -}) - f(X_{\tau_i - 1}) + \int_0^t \int_E f(x) - f(X_{s-})\, r(X_{s-}, \mathrm{d}x)\, \lambda(X_{s-})\,\mathrm{d}s$$

$$= \int_0^t \Big[\frac{\mathrm{d}}{\mathrm{d}t} f(X_s) + \lambda(X_s) \int_E f(x) - f(X_s)\, r(X_s, \mathrm{d}x)\Big]\mathrm{d}s\,.$$

Therefore K is for Lebesgue almost all $t \ge 0$ of the proposed form for the extended generator (2.2.6) almost surely.

Chapter 3

Hybrid stochastic models of excitable media

The processes and results in the preceding section are derived with the aim to construct an analytically tractable hybrid model of excitable membranes, particularly neuronal membranes. In Section 3.1 we describe as an instructive example how to cast the classical Hodgkin-Huxley model for the squid giant axon into the framework of PDMPs in infinite-dimensions. The main result in Theorem 3.1.1 is that such models form a standard PDMP, cf. Definition 3.0.1. Although illustrated on the classical example of the Hodgkin-Huxley axon equation, the modelling approach extends immediately to similar channel based stochastic models, e.g., stochastic versions of models of cardiac tissue or models in Calcium dynamics [47], e.g., the Kaizer-DeYoung model. At the end, in Section 3.1.2 we also comment how the infinite-dimensional PDMP model reduces to a finite-dimensional PDMP model in the special case of space-clamped patches of membrane and discuss compartmental-type membrane models which play an important role for simulation studies, cf. [45, 94].
In terms of modelling, Section 3.1 picks up where we left off at the end of Section 1.2 in order to discuss the general class of stochastic processes that the specific structure of the modelling problem demands. However, prior to this we discuss briefly the specific class of PDMPs we employ modelling these systems. At the end of this chapter we also very briefly discuss deterministic models of excitable media. In particular, we present an existence theorem of solutions of general excitable media equations. We further present results on the qualitative properties of the solutions to membrane equations. These are important preliminary results for the derivation of limit theorems to PDMP models of excitable media which we pursue in Chapter 4.

3.0.1 An appropriate class of PDMPs

In the following we restrict ourselves to consider homogeneous processes, which correspond to models of excitable media without time-dependent input current. To include time-dependent input does not pose an additional problem resulting in inhomogeneous PDMPs, cf. Section 2.1. These can be easily reduced to homogeneous PDMPs via the usual state space extension, cf. the discussion of the space-time process in Section 2.0.1. Thus, let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ be a filtered probability space and $(X_t)_{t\geq 0}$ be a PDMP on a Borel measurable space E, i.e., a càdlàg strong Markov process the trajectories of which possess the following distinct features: They have discontinuities at random times τ_n and random post-jump values X_{τ_n} but move deterministically in phase space in between jumps as defined by a semigroup of mappings $(\varphi_{t,s})_{s\geq t\geq 0}$. The subclass of PDMPs that is employed for the purpose of modelling biological excitable media in the subsequent sections exhibits this general dynamics in the following form: The homogeneous PDMP $(X_t)_{t\geq 0}$ consists of two components $X_t = (Y_t, \theta_t)$ taking values in the state space $E = H \times K$. The first component, Y_t, taking values in H, possesses continuous sample paths and we call Y_t the *continuous component* of the PDMP. The space H is a separable, real Hilbert space. In applications it is either an infinite-dimensional function space, e.g. $H = L^2(D)$ on some spatial domain D, for models incorporating spatial dynamics, or, in the reduction to models without spatial dynamics, it is finite-dimensional, i,e, $H = \mathbb{R}^d$, $1 \leq d < \infty$. The second component, θ_t, takes values in K, possesses piecewise constant sample paths and we denote it as the *piecewise constant component* of the PDMP. Here K denotes an at most countable set of distinct states. For models of excitable media it is a finite set of integer vectors that denote the collection of physiological states different ion channels are in. Thus, note that for this type of PDMPs jumps occur only in a fixed subset of components which in addition are otherwise constant.
Hence, a full, precise definition of the dynamics of the PDMP requires us to consider two constituents making up the characteristics of the process. One that governs the deterministic evolution of the continuous component Y_t and a second one that defines the random jumps of the piecewise constant component θ_t. These components are as considered in Examples 2.1.4 and 2.1.1 which we now briefly recall.

Firstly, the *deterministic evolution of the continuous component* in between jumps is governed by a family of abstract evolution equations. In the membrane model this component describes the dynamics of the transmembrane potential and is usually given by a reaction-diffusion equation defining an abstract evolution equation. What distinguishes hybrid model from deterministic or SDE models is the description of the gating system using an exact description of the channel dynamics based on their kinetic schemes, cf. Fig. 1.1. Hence, channel states change instantaneously at random times and are constant in between. Let $X \subset H \subset X^*$ be an evolution triple of real, separable Hilbert spaces with the embeddings being continuous and dense. Then, for the construction of the continuous component Y_t of the PDMP we consider nonlinear abstract evolution equations

$$\dot{y} = A(\theta)\, y + B(y, \theta) \tag{3.0.1}$$

for all $\theta \in K$, where $A(\theta) : X \to X^*$ is a linear operator and $B(\cdot, \theta) : X \to X^*$ is a nonlinear operator. For the applications in Section 3.3 the operators $A(\theta)$ are second order elliptic partial differential operators with time-independent coefficients, as in the nonlinear cable equation (1.2.1). The parameter θ corresponds to the coefficients g_i which change every time a channel opens or closes. We denote by $\psi(t, (y, \theta))$ the solution to the abstract initial value problem (3.0.1) with initial condition $u(0) = y$, $y \in H$. Given any initial condition $y \in H$ we assume that there exists a unique global weak solution, i.e., $\psi \in L^2((0,T), X) \cap H^1((0,T), X^*) \subset C^0([0,T], H)$ with $\dot{\psi} \in L^2((0,T), X^*)$ for every $T < \infty$, which forms a dynamical system in H. Moreover, as the component θ_t of a PDMP is constant in between jumps it follows that in between jumps the paths of the complete PDMP $(Y_t, \theta_t)_{t\geq 0}$ generate a dynamical system, i.e., for $x = (y, \theta) \in E$ we define the mapping $\phi_t(x) := (\psi(t, (y, \theta)), \theta)$ and it is easy to see that $(\phi_t)_{t\geq 0}$ is a dynamical system on E.

Secondly, the *stochastic jump mechanism* is given by a jump rate Λ governing the jump times and a probability μ for the distribution of the post-jump values describing the random state transitions of the ion channels. Thus $\Lambda : E \to \mathbb{R}_+$ is such that at any time $t \geq 0$ the distribution of the waiting time until the next jump possesses the survivor function

$$\begin{aligned} S(s, X_t) &= \mathbb{P}\big[\theta_{t+r} = \theta_t,\ \forall\, r \in [0, s] \,\big|\, X_t\big] \\ &= \exp\Big(- \int_0^s \Lambda\big(\phi_r(X_t)\big)\, \mathrm{d}r \Big)\,. \end{aligned} \tag{3.0.2}$$

That is, $S(s, X_t)$ states the probability conditional on X_t that there does not occur a jump in $[t, t+s]$. Further, the transition probability μ takes the form of a Markov kernel[1] $\mu : E \times \mathcal{B}_K \to [0,1]$ from E into K defining the distribution of the PDMP's

[1] We note that we make at this point a fundamental change of notation which remains in place for the remainder of the thesis. We no longer use the symbol 'μ' to denote a random counting measure which is used to define a PDMP as in Chapter 2 and Appendix B.2. In particular, for the remainder of thesis we do not use explicitly the random counting measure underlying (or associated) with a PDMP anymore. Therefore the symbol 'μ' is from now on used to denote the Markov kernel defining the post-jump values of a PDMP. In Chapter 2 this Markov kernel was denoted by r, cf. conditions (P2) and (P3) in Section 2.1. The reason for this change in notation is partly due to an attempt to keep the notation of the individual parts of the thesis easily comparable to the respective main

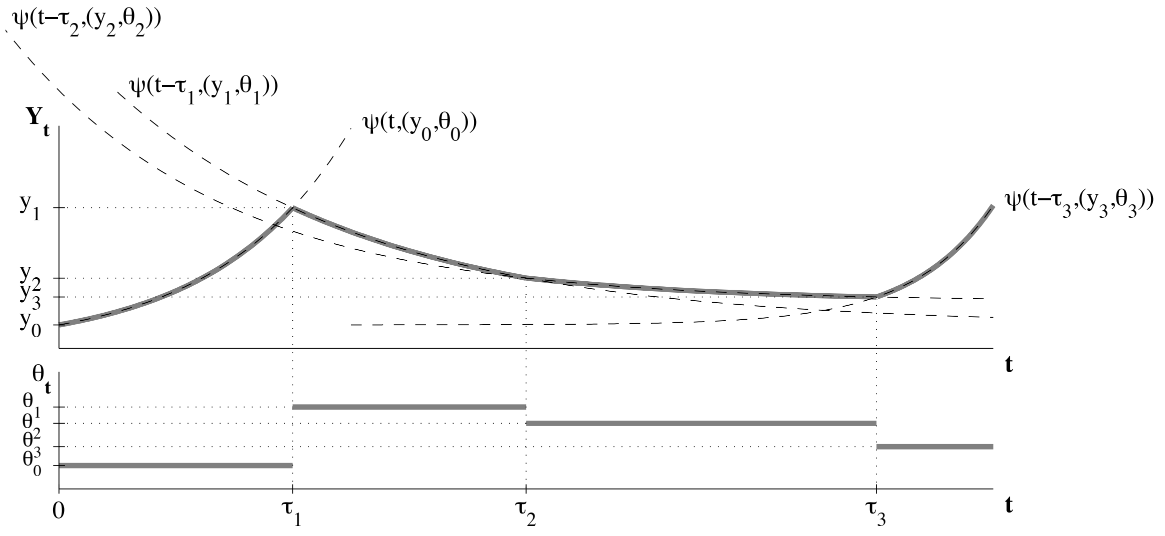


Figure 3.1: Sketch of the construction of a PDMP $(Y_t, \theta_t)_{t\geq 0}$ where the bold grey curves form its trajectory. The continuous component Y_t is shown in the upper part and the piecewise constant component θ_t in the lower part of the figure. The transition times τ_n and the transition targets θ_{τ_n} are random variables depending on the past of the sample path.

post-jump values conditional on the pre-jump values, i.e., for all $A \in \mathcal{B}_K$

$$\mathbb{P}[\theta_t \in A \,|\, \theta_{t-} \neq \theta_t] = \mu(X_{t-}, A)\,. \tag{3.0.3}$$

Here $\mathcal{B}_K$ denotes the Borel sets on K. We note that some care has to be taken defining Λ and μ for a specific model. In general, H is a function space and uniqueness of solutions to (3.0.1) is only given with respect to equivalence classes with respect to the norm on the spaces they belong to, i.e., $L^2((0,T),X)$ or $C^0([0,T],H)$. Therefore, one has to ensure that functions of solutions to (3.0.1), that is, the jump rate Λ and the point probabilities of the Markov kernel μ, are well defined in the sense that they are invariant with respect to equivalence classes.

Using these mechanisms a trajectory of the process is then constructed piecing together the continuous deterministic components at the random jump times as indicated by the definition of a PDMP, see Definition 2.1.1, or Davis' construction procedure, see Section 2.1.1. The structure of trajectories of such PDMPs is illustrated in Figure 3.0.1. The following definition presents conditions on the characteristics (3.0.1) – (3.0.3) which guarantee the well-definedness of the PDMP by Theorem 2.1.1.

Definition 3.0.1. Let H be a real, separable infinite-dimensional Hilbert space. An

references, e.g., [64] for the more theoretical Chapter 2 and Appendix B.2 or [79, 80, 96] for Chapter 4. Further, in our opinion using Greek letters for the closely connected objects 'jump rate' and the 'transition kernel' facilitates their identification and understanding of the formulae, particularly for Chapter 4. Once again, we emphasise, that being aware of this change of notation we are of the opinion that no ambiguities or confusion should arise.

$H \times K$–valued PDMP $(X_t)_{t\geq 0} = (Y_t, \theta_t)_{t\geq 0}$ of the type described in this section is called a *standard* infinite-dimensional PDMP on H if the following conditions are met.

(i) The abstract initial value problem (3.0.1) is well-posed for every $\theta \in K$, i.e., it admits a unique global solution $\psi(\cdot, (y, \theta))$ continuous in H for every initial value $y \in H$ and $\psi(t, (y, \theta))$ depends continuously on the initial condition for all $t \geq 0$.

(ii) The transition rate $\Lambda : E \to \mathbb{R}_+$ is measurable and path-integrable, i.e., for all $x \in E$ it holds that

$$\int_0^t \Lambda(\phi_s(x))\,\mathrm{d}s < \infty \qquad \forall\, t < \infty,$$

but the integral diverges for $t \to \infty$.

(iii) The transition measure μ is a Markov kernel from E to K which satisfies $\mu((y, \theta), \{\theta\}) = 0$ for all $(y, \theta) \in E$.

(iv) The PDMP $(X_t)_{t\geq 0}$ is *regular*, i.e, the number of jumps in the piecewise constant component during any finite time interval is finite almost surely.

Note that the semigroup defining the inter-jump motion of the PDMP is uniquely defined by the operators A, B, hence it is consistent with Section 2.1 to call the quadruple (A, B, Λ, μ) the *characteristics of a standard inifinite-dimensional PDMP*. We briefly comment on the conditions in Definition 3.0.1. First, note that we choose the space H as the phase space of the PDMP as this results in a càdlàg stochastic process by definition of a weak solution. However, we note that regularity results on solutions to (3.0.1) can be carried forward to the paths of PDMPs. That is, e.g., if solutions to (3.0.1) are continuous with respect to a stronger norm than the norm in H, e.g., the norm in X, then also the PDMP is càdlàg with respect to this stronger norm. That is, if the conditions (i)–(iii) are satisfied with respect to the space X then we can obviously say that the process is a PDMP in X. Further, we remark that condition (ii) ensures that for all $x \in E$ the survivor function satisfies $\lim_{t\to\infty} S(t, x) = 0$ which is sufficient for the survivor function to completely define a probability distribution of a non-negative, real random variable. Also the condition ensures that successive jumps cannot occur simultaneously and there is always some positive time in between. Moreover, it guarantees that the piecewise constant component does not stop jumping and is trapped in a final state for all time with positive probability. Finally, measurability of real valued functions on E in conditions (ii) and (iii) is understood as measurability with respect to the product σ-field obtained from the Borel sets on H and the discrete σ-field on K. Also in the image set $\mathbb{R}_+$ and $[0, 1]$, respectively, the Borel σ-field is used to obtain a measurable space. We note that in the remainder of this section measurability is always understood in these terms. Therefore, continuity of the solutions of (3.0.1) with respect to t and with respect to the initial condition (as part of well-posedness) is sufficient for the jointly measurablility of the semigroup, i.e., measurability of the map $(t,x) \mapsto \phi_t(x)$. Finally, we remark on condition (iv). A simple, sufficient condition for regularity of the process, which is easy to show in most relevant cases, is that the transition rates are bounded along the solutions $\phi(\cdot,x)$ for all $x \in E$.

If we consider neuron models without a spatial dimension, then the equations for the excitable media reduce from a partial differential equation to an ordinary differential equation, cf. equations (1.2.2) and (1.2.1) in Section 1.2.3, and the state space is given by $H = \mathbb{R}^d$ with $1 \leq d < \infty$. That is, for the definition of the inter-jump dynamics we consider a family of ordinary differential equations

$$\dot{y} = g(y,\theta) \tag{3.0.4}$$

with constant parameter $\theta \in K$ and $g(\cdot,\theta) : \mathbb{R}^d \to \mathbb{R}^d$. Here each g is such that for any initial condition a unique global solution exists depending continuously on the initial condition. In the membrane model this corresponds to the one-dimensional ODE (1.2.2) with coefficients g_i changing every time a channel opens or closes. However, we consider a multi-dimensional state space $\mathbb{R}^d$, as, e.g., in membrane models which also include calcium gated ion channels, we have a two-dimensional system (3.0.4) also describing the calcium dynamics in addition to the transmembrane potential. In addition, as the component θ_t of a PDMP is constant in between jumps it follows that in between jumps a PDMP's paths satisfy a system of ODEs of the form

$$\begin{pmatrix} \dot{y} \\ \dot{\theta} \end{pmatrix} = \begin{pmatrix} g(y,\theta) \\ 0 \end{pmatrix}. \tag{3.0.5}$$

We denote by $t \mapsto \phi(t,x)$ the unique, global solution of the system (3.0.5) with respect to the initial value $x = (y,\theta) \in E$. This generates a dynamical system.

In the case of a finite-dimensional state space the conditions of Definition 3.0.1 reduce to the following which correspond to the conditions originally given by Davis [37, Def. 24.8]. We note that the statements in the remarks following Definition 3.0.1 remain valid for the finite-dimensional case.

Definition 3.0.2. A finite-dimensional stochastic process $(X_t)_{t\geq 0} = (Y_t,\theta_t)_{t\geq 0}$ of the type described in this section is called a *standard* finite-dimensional PDMP on $\mathbb{R}^d$ if the following conditions are met.

(i) The system of ordinary differential equation (3.0.5) admits a unique global solution $\phi(\cdot,x)$ for every initial conditions $(y,\theta) \in \mathbb{R}^d \times K$ which depends continuously on the initial condition.

(ii) The transition rate $\lambda : \mathbb{R}^d \times K \to \mathbb{R}_+$ is measurable and path-integrable, i.e., for all $x \in E = \mathbb{R}^d \times K$ it holds that

$$\int_0^t \lambda(\phi_s(x))\,\mathrm{d}s < \infty \qquad \forall\, t < \infty,$$

but the integral diverges as $t \to \infty$.

(iii) The transition measure μ is a Markov kernel from E to K which satisfies $\mu((y,\theta),\{\theta\}) = 0$ for all $(y,\theta) \in E$.

(iv) The PDMP $(X_t)_{t\geq 0}$ is *regular*, i.e, the number of jumps in the piecewise constant component during any finite time interval is finite almost surely.

We remark that it is consistent with earlier definitions to call the triple (g, λ, μ) the *characteristics of a standard PDMP*.

3.1 An exact spatio-temporal model of neuronal membranes

In this section we present an exact hybrid model of neuronal membranes using PDMPs. We employ the classical Hodgkin-Huxley model of the squid giant axon to exemplify the modelling approach. The section is concluded with a discussion on how to generalise the approach presented to models of arbitrary neuronal membranes and also on multi-dimensional spatial domains where two- and three-dimensional domains are of particular physical interest. In particular we find that also the model presented in [7], to the best of our knowledge the only spatial hybrid model so far considered for neuronal membranes, can be cast into a PDMP framework. In the following we first consider the appropriate modelling of the ion channels, i.e., the stochastic jump mechanism for the PDMP, and then discuss the modelling of the conductances in the equation for the transmembrane potential.

In the classical Hodgkin-Huxley model the axon is considered as a spatially homogeneous cable in one space dimension, i.e., along the length of the axon [74]. Thus for the spatial domain we take an open interval D in $\mathbb{R}$. Further, in this model two families of voltage-gated ion channels are considered causing the excitable dynamics: sodium (Na) and potassium (K) channels. The kinetic schemes for these channels are given in Fig. 1.1. We provide in Appendix A the precise definition of the rate functions and the numeric values of the constants.
To start with the modelling we define the state space K of the piecewise constant component which we take to be a collection of $n \in \mathbb{N}$ ion channels. Let $\mathcal{C}$ denote the set of all states a single channel can be in, that is, all states of the two kinetic diagrams in Fig. 1.1. Here we consider the two diagrams as one Markov chain, consisting of two irreducible parts. That means, we do not need to differentiate between K– and

Na–channels as this characterisation is done by the state the channels are in and by the convention that transitions from states belonging to an Na–channel to states of a K–channel do not occur with probability one and vice versa. Thus we characterise each channel by a pair $\theta_i = (c_i, x_i),\ \ i = 1, \dots, n$, where $c_i \in \mathcal{C}$ indicates the (time-varying) state of the channel and $x_i \in D$ its (fixed) spatial position. Therefore the elements of the state space $K = \mathcal{C}^n$ of the piecewise constant component of the PDMP are the channel configurations $\theta = (\theta_1, \dots, \theta_n)$, i.e., n–tuples collecting the states the single channels are in. One obtains a new configuration, which corresponds to 'a state' of the piecewise constant component $\theta(t)$ in the membrane model, every time one channel changes its state.

The spatio-temporal dynamics of the transmembrane potential on D are governed by the cable equation

$$\frac{r}{2R}\frac{\partial^2 u}{\partial x^2} = C\frac{\partial u}{\partial t} - \sum_i g_i(u, x)(E_i - u)\,. \tag{3.1.1}$$

Here the positive constants r, R, C denote the axon radius, the resistance of the cytoplasm and the capacitance of the membrane, respectively, [74]. Additional to the currents due to Na– and K–channels, this model also contains a third current due to leakage (L) of charge independent of the ion channels. Therefore, in equation (3.1.1) we have the summation in the right hand side over $i \in \{\mathrm{Na}, \mathrm{K}, \mathrm{L}\}$. We introduce the dependence of the conductances on the channel configuration, i.e., $g_i(u, x) = g_i(\theta, x)$ for $i = \mathrm{Na}, \mathrm{K}$ and $g_\mathrm{L}(u, x) = g_\mathrm{L}(\theta, x) \equiv \overline{g}_\mathrm{L}/|D|$, and rearrange (3.1.1) to

$$\frac{\partial u}{\partial t} = \frac{r}{2CR}\frac{\partial^2 u}{\partial x^2} + C^{-1}\sum_i g_i(\theta, x)(E_i - u)\,. \tag{3.1.2}$$

This is an evolution equation of the form (3.0.1) where the linear operator $A(\theta)$ is given by

$$A(\theta) = \frac{r}{2CR}\frac{\partial^2}{\partial x^2} - C^{-1}\sum_i g_i(\theta, x)$$

and the nonlinearity $B(\theta, u)$ is independent of u and given by the non-autonomous term

$$B(\theta, u) \equiv C^{-1}\sum_i g_i(\theta, x)E_i\,.$$

Next we model the dependence of the conductances $g_\mathrm{Na}(\theta, x),\ g_\mathrm{K}(\theta, x)$ on the channel configuration $\theta \in K$. To the best of our knowledge the only published article dealing with hybrid spatial models in the context of neuron models is [7]. In the context of calcium dynamics the authors in [47, 115] consider a hybrid spatial model which allows a description by PDMPs. These studies provide us with two possibilities of modelling the conductances.

According to [7, 115] a single channel is modelled as a point source generating current if it is in its conducting state, which yields conductances of the form

$$g_{\mathrm{Na}}(\theta,x)=\overline{g}_{\mathrm{Na}}\sum_{j\colon c_j=m_3h_1}\delta(x-x_j),\quad g_{\mathrm{K}}(\theta,x)=\overline{g}_{\mathrm{K}}\sum_{j\colon c_j=n_4}\delta(x-x_j),\tag{3.1.3}$$

where as before the pair (c_j,x_j) denotes the state and position of the jth channel, $\overline{g}_{\mathrm{Na}}$, $\overline{g}_{\mathrm{K}}$ are the conductances of a single Na– or K–channel, respectively, and δ is the Dirac delta function. Using this definition for the conductances the resulting equations (3.1.2) are of the form as considered in [7]. For a second possibility of defining conductances, it can be argued that the concentration of sodium or potassium is essentially homogeneous in a spatially extended domain around an open channel similarly as it is the case for calcium dynamics. Therefore the current through a channel is not modelled as a point source but as being present in a neighbourhood $U(x_j)$ for a channel at point x_j. In this case we obtain conductances defined as

$$\begin{aligned} g_{\mathrm{Na}}(\theta,x) &= \overline{g}_{\mathrm{Na}}\sum_{j\colon r_j=m_3h_1}|U(x_j)|^{-1}\,\mathbb{I}_{U(x_j)}(x)\,,\\ g_{\mathrm{K}}(\theta,x) &= \overline{g}_{\mathrm{K}}\sum_{j\colon r_j=n_4}|U(x_j)|^{-1}\,\mathbb{I}_{U(x_j)}(x)\,. \end{aligned}\tag{3.1.4}$$

Here $\mathbb{I}_A$ denotes the indicator function for the set A. The distinctive effect of choosing (3.1.4) over (3.1.3) as a model for the conductances is that one obtains different spatial regularity for the solutions of (3.1.2). In particular, the smoother $g_i(\theta,x)$ is chosen, the higher is the resulting spatial regularity. Hence, if high regularity of the membrane variable is required from a modelling point of view, it is reasonable to consider smooth enough approximations $d(x,x_j)$ to $\delta(x-x_j)$ to define the conductances, i.e., d are non-negative functions whose integral over D equals one. From this point of view $d(x,x_j)=|U(x_j)|^{-1}\,\mathbb{I}_{U(x_j)}(x)$ can also be considered as an approximation to the Dirac delta function.

For now, we consider conductances defined by (3.1.3) as this is the least smooth choice – hence the most general case – but still sufficiently smooth for the model to be a standard infinite-dimensional PDMP. Further, in accordance with [7]

- we assume Dirichlet boundary conditions, i.e., $u(t,x)=0$ on ∂D for all $t\geq 0$, physiologically corresponding to an axon cut open at its ends, cf. [74];
- we choose for the evolution triple $X\subset H\subset X^*$ the spaces $H=L^2(D)$ and $X=H^1_0(D)$. Note that $\partial/\partial x^2$ maps $H^1_0(D)$ into $H^{-1}(D)$ and also $\delta\in H^{-1}$ due to the Sobolev Embedding Theorem as the spatial domain is one-dimensional. This choice of function spaces guarantees the existence, uniqueness and minimal sufficient smoothness of solutions to (3.1.2), cf. Theorem 3.1.1, and appropriately captures the Dirichlet boundary conditions.

To finish the definition of a PDMP it remains to define the transition dynamics of the piecewise constant component, i.e., a transition rate Λ and a transition measure μ. Recall that for single ion channels the transition dynamics are defined by rates a_i, b_i, $i = n, h, m$. These yield waiting time distributions for a transition in a single channel of the form (1.2.5), i.e.,

$$\mathbb{P}[\tau > t] = \exp\Big(-\int_0^t a_i(u(s,x))\mathrm{d}s\Big)$$

for $u(s,x)$ denoting the time varying transmembrane potential at $x \in D$. Obviously rates are additive, i.e., if from a state $c \in \mathcal{C}$ several other states can be reached the rate of leaving c, denoted by q_c, is given by the sum of all rates of possible transitions. For example, a sodium channel from Fig. 1.1 located at $x \in D$ in state m_2h_0 leaves this state with rate $q_{m_2h_0} = 2b_m + a_h + a_m$ and the waiting time distribution of the time until it switches to any target state is given by

$$\mathbb{P}[\tau > t] = \exp\Big(-\int_0^t q_{m_2h_0}(u(s,x))\mathrm{d}s\Big)\,.$$

Moreover, for a collection of independent channels the rate of the first of these channels to change its state is given by the sum of all rates for single channels leaving their present state. Therefore the rate of the first transition of a single channel in a configuration θ is given by

$$\Lambda(u(t,\cdot),\theta) = \sum_{j=1}^{n} q_{c_j}(u(t,x_j))\,, \tag{3.1.5}$$

which yields the transition rates $\Lambda(\cdot,\theta) : H_0^1(D) \to \mathbb{R}_+$ of the PDMP.
Finally we define the transition measure μ. This is a discrete probability distribution on K conditionally on the point in phase space that the process occupies at the transition time τ. Recall that a transition in the configuration θ happens when a single channel switches its state c. Thus to define a distribution on the set of all configurations we associate with each possible transition $c \rightharpoonup \widehat{c}$, c and $\widehat{c}$ in $\mathcal{C}$, that can occur for single channels, i.e., for all transitions that can happen for each single channel θ_i, $i = 1, \dots, n$, a probability

$$\mu\big((\theta, u(\tau,\cdot)), \{\widehat{\theta}\}\big) = \frac{q_{c\rightharpoonup\widehat{c}}(u(\tau,x_j))}{\Lambda(u(\tau,\cdot),\theta)}\,, \tag{3.1.6}$$

where $q_{c\rightharpoonup\widehat{c}}$ is the rate for the transition $c \rightharpoonup \widehat{c}$ for a single channel. Hence the probability (3.1.6) is associated to the transition of configurations $\theta \to \widehat{\theta}$ in K with $\widehat{\theta}$ denoting the configuration one obtains if the channel at position x_j switches from state c to $\widehat{c}$. These probabilities thus define for every $(u,\theta) \in H_0^1(D) \times K$ a probability distribution on K which give the transition measure μ.

The above definition of the transition rates obviously necessitates that the membrane potential is pointwise uniquely defined in $(0,T)\times D$. That is, the special structure of the transition rates (3.1.5) and transition probabilities (3.1.6) evaluates the solution ψ along t at individual points x_j. Therefore, in order to be well-defined the solution necessarily needs to take for all $t\geq 0$ values in the space $H^1_0(D)$ which is embedded in $C^0(\overline{D})$, cf. [7] and also the proof of Theorem 3.1.1 below. However, similar to modelling the conductances by (3.1.4) it can be argued that the transmembrane potential around the location of a channel governs its gating behaviour. This then yields transition rates of the form, e.g.,

$$q_{c\to\hat{c}}\Big(|U(x_j)|^{-1}\int_{U(x_j)} u(t,x)\,\mathrm{d}x\Big). \tag{3.1.7}$$

Such rates are continuous and invariant with respect to values of the solution in the coarser space $L^2(D)$. Using (3.1.7) it is straightforward to define the transition rate Λ and the transition measure analogously to (3.1.5) and (3.1.6). Anticipating the discussion in Section 3.1.1, we note that the approach of modelling the conductances by (3.1.4) and the transition rates by (3.1.7) is better suited for extensions to multi-dimensional spatial domains than the model in [7].

We now state the central result of this section, that is, that the stochastic hybrid membrane models discussed are standard infinite-dimensional PDMPs $(Y_t,\theta_t)_{t\geq 0}$. We further obtain, that under physiological reasonable initial conditions the transmembrane potential component Y_t remains bounded, a fact known for the deterministic Hodgkin-Huxley model. To this end we define the bounds $\overline{u}_-=\min\{E_{\mathrm{Na}},E_{\mathrm{K}},E_{\mathrm{L}}\}$, $\overline{u}_+=\max\{E_{\mathrm{Na}},E_{\mathrm{K}},E_{\mathrm{L}}\}$. The interval $[\overline{u}_-,\overline{u}_+]$ is called a *physiological domain* for the transmembrane potential [42].

Theorem 3.1.1. *The initial condition of the continuous component u_0 satisfies $u_0(x)\in[\overline{u}_-,\overline{u}_+]$ for all $x\in\overline{D}$. Then it holds:*

(a) *For initial conditions $u_0\in H^1_0(D)$ the stochastic hybrid membrane model as discussed above, with any choice of conductance model and transition rates, is an infinite-dimensional standard PDMP on $H^1_0(D)$. Moreover, the continuous component satisfies that $u(t,x)\in[\overline{u}_-,\overline{u}_+]$ for all $(t,x)\in\mathbb{R}_+\times D$.*

(b) *For initial conditions $u_0\in L^2(D)$ the stochastic hybrid membrane model consisting of any choice of conductance models and transition rates as defined in* (3.1.7) *is a standard infinite-dimensional PDMP on $L^2(D)$. Moreover, the continuous component satisfies for all $t\geq 0$ that $u(t,x)\in[\overline{u}_-,\overline{u}_+]$ for almost all $x\in\overline{D}$.*

Proof. We have to prove that the defining properties of the process satisfy the conditions (i)–(iv) of Definition 3.0.1 which we do step by step in parts (A)–(D) in the following.

(A) We can use standard existence theory of linear parabolic equations, cf. Theorem C.0.2, to obtain that there exists a unique weak solution $\psi(t,(y,\theta))$ to (3.1.2) to every initial value $y \in L^2(D)$ for every $T < \infty$ and all configurations $\theta \in K$. In particular, the solutions satisfy $\psi \in L^2((0,T), H_0^1(D)) \cap H^1((0,T), H^{-1}(D))$ which implies $\psi \in C^0([0,T], L^2(D))$ and the dependence on the initial condition is continuous with respect to the norm in $L^2(D)$. Moreover, for conductances (3.1.3) the additional regularity result Proposition 3.4.3 implies that the solution is in fact in $C^0([0,T], H_0^1(D))$ for initial conditions $y \in H_0^1(D)$. For conductances $g_i(\theta) \in L^2(D)$ the same holds due to general existence theory of parabolic PDEs, see Theorem C.0.3 (a). This means that there exists a unique continuous function $\widehat{\psi} \in C^0([0,T], H_0^1(D))$ such that each element of the equivalence class of solutions in $L^2((0,T), H_0^1(D))$ coincides with $\widehat{\psi}$ in $L^2((0,T), H_0^1(D))$. Note, that for each $t \in [0,T]$ the value $\widehat{\psi}(t,(\cdot,\theta))$ is an equivalence class in $H_0^1(D)$. However, the Sobolev Embedding Theorems states that $H_0^1(D)$ is embedded in $C^0(\overline{D})$. Hence, for each t there exists a unique $\widetilde{\psi}(t,(\cdot,\theta)) \in C^0(\overline{D})$ such that $\widehat{\psi}(t,(\cdot,\theta)) = \widetilde{\psi}(t,(\cdot,\theta))$ in $H_0^1(D)$. We always identify $\psi = \widetilde{\psi}_r$ and thus ψ is continuous as a function $[0,T] \times \overline{D} \to \mathbb{R}$. Moreover, in this case it holds that also the dependence on the initial condition is continuous with respect to the norm in $H_0^1(D)$. Thus it holds that condition (i) is satisfied with respect to the space $L^2(D)$ in the case that the initial condition is in $L^2(D)$ and with respect to the space $H_0^1(D)$ in the case that the initial condition is in $H_0^1(D)$.
Further, due to Proposition 3.4.2 the solution ψ satisfies $\psi(t,(y,\theta))(x) \in [\overline{u}_-, \overline{u}_+]$ for all $t \geq 0$, all $x \in \overline{D}$ and every $(y,\theta) \in E$ as long as the initial condition satisfies $y(x) \in [\overline{u}_-, \overline{u}_+]$. We refer to Section 3.4.3 for more details wherein we prove a type of weak maximum principle for excitable membrane PDEs. However, this immediately implies that the membrane component of the hybrid stochastic model remains bounded within the interval $[\overline{u}_-, \overline{u}_+]$ for a suitable initial condition u_0 as the successive initial conditions after jumps in the piecewise constant component are solution values of the preceding PDE and thus also within these bounds.

(B) Next, Λ defined as (3.1.5) is invariant with respect to equivalence classes of solutions which can be identified with a pointwise uniquely defined function, i.e., for solutions in $C^0([0,T], H_0^1(D))$. Further, for continuous, bounded transition rates a_x, b_x, $x = n, m, h$ the transition rate Λ is continuous with respect to the norms in $H_0^1(D) \times K$ and $\mathbb{R}$ and thus Borel-measurable. We remark that the rates for the Hodgkin-Huxley model, cf. Appendix A, are continuous, positive but unbounded. However, as the range of possible values for the transmembrane potential is bounded, see part (A) of the proof, the transition rates can be considered bounded without loss

of generality. Further, as they are strictly positive we can consider them to be bounded away from zero. Hence, Λ can be considered bounded and thus Λ is path-integrable over all finite intervals $[0,T]$, $T \geq 0$. Finally, as we can consider the individual rates a_x, b_x to be bounded away from zero so is the transition rate Λ. Thus, the integral of Λ along the trajectories of the semigroup diverges for $T \to \infty$. Condition (ii) is satisfied for initial conditions $u_0 \in H_0^1(D)$ which are in $[\overline{u}_i, \overline{u}_+]$ pointwise.
On the other hand, Λ as defined by the transition rates (3.1.7) is invariant with respect to equivalence classes of solutions taking values in $C^0([0,T], L^2(D))$. Note that this space possesses insufficient spatial regularity for the rates (3.1.5) being invariant as they necessitate pointwise uniquely defined potential variables, i.e., $u(t) \in H_0^1(D)$ for all t. Just as argued before we find that Λ defined by (3.1.4) is a suitable transition rate for a PDMP with paths continuous in $L^2(D)$. Thus for the above model condition (ii) is satisfied for both conductance models (3.1.3) and $g_i(\theta) \in L^2(D)$ and any initial condition of the membrane variable $u_0 \in L^2(D)$ which is in $[\overline{u}_i, \overline{u}_+]$ pointwise.
This, particularly, shows that initial conditions $u_0 \in L^2(D)$ and transition rates defined by (3.1.5) do not define a PDMP.

(C) By definition there is a close connection between the transition rate Λ and the point probabilities. That is Λ is a finite sum and the point probabilities correspond to the individual summands of Λ weighted by Λ^{-1} such that they sum up to 1. Therefore, making the same distinctions corresponding to part (a) and (b) of Theorem 3.1.1 as in part (B) of the proof we infer that also condition (iii) is satisfied: The point probabilities are continuous functions with respect to the appropriate norms and they are invariant with respect to equivalence classes of solutions. Further, the probabilities of remaining in the same state after a transition are zero as the probabilities of individual channels remaining in the same state upon switching are zero, i.e., $q_{c\to c} = 0$ for all $c \in \mathcal{C}$.

(D) Finally, for condition (iv) to hold it is sufficient that the transition rate Λ is bounded [37, Prop. 24.6]. This follows immediately from having bounded transition rates a_x, b_x, which completes the proof.

□

3.1.1 Generalisation and discussion

We remark again that the above formulation of a spatial hybrid model is, to the best of our knowledge, the second one in the literature in addition to [7]. The results in [7] rely on the very specific structure of the special case treated therein, particularly the explicitly given transition semigroup of the PDE solutions, cf. a discussion of the model in Section 3.4.1. The last section of [7] includes a discussion on desirable extensions and analytical results for spatial hybrid models and comments on the limitations of the author's formulation of the model. We address these topics form the point of view of the formulation as PDMP models and collect some remarks on extensions of the described modelling approach to general conductance based stochastic hybrid models in spatial domains.

(i) The author in [74] gives a good review of why it is sufficient and a reasonably good approximation to consider long and thin nerve fibres as one-dimensional cables, neglecting their actual three-dimensional extent. In contrast to the description of a hybrid axon model as presented in [7] one major advantage of the PDMP formulation is that it can be readily extended to include the actual three-dimensional shape of a nerve fibre. An extension to two-dimensional membranes is of particular interest for modelling calcium dynamics [47, 115] and hybrid versions of cardiac tissue models [48]. Using the infinite-dimensional PDMP formulation we are not only able to capture dynamics in the plane but can also consider arbitrary, sufficiently smooth manifolds, e.g., the surface of a cylinder for the realistic shape of an axon.

(ii) We have so far considered only a spatially homogeneous model for an axon. The model presented, or general multi-dimensional models, respectively, can be extended in a straightforward fashion to include spatial dependence in the coefficients of the partial differential operator. This is achieved by simply replacing the Laplacian by an arbitrary autonomous second order partial differential operator. This can account for, e.g., changing fibre diameter or variations in the resistance or capacitance. Spatial inhomogeneity also includes spatial variation in the distribution of ion channels for which there is no restriction inherent in the model. This allows using the closed description of PDMPs to model extremely inhomogeneous structures such as myelinated nerves with a concentration of channels at the Nodes of Ranvier. It is also possible to consider random distributions of channels which yields random ordinary / partial differential equations for the dynamics of the continuous components of the PDMP between transitions of its piecewise constant component.

(iii) Clearly it is possible to consider different boundary conditions for the PDEs (3.1.2), e.g., Neumann boundary conditions modelling sealed ends of a fibre, which is most realistic for neurons in living tissue [74].

(iv) Finally, for the PDMP framework it does not pose any difficulty to deal with time-dependent input currents into the neurons, i.e., more general, time-dependent right hand sides in the membrane balance equation (3.1.2) or even time-dependent boundary conditions. The resulting process is simply a non-homogeneous PDMP as covered by the general PDMP theory discussed in Section 2.1. Moreover, by a simple state space extension the process can be transformed into a homogeneous

Markov process, cf. the discussion of 'space-time processes' in Section 2.0.1.

Mathematically, the extensions of the model we mentioned above essentially reduce to appropriately modelling the coefficients in the partial differential operator and the conductances of the ionic currents such that a solution exists and the transition dynamics of the piecewise constant component are well defined. That is, on the one hand, using the transition rates (3.1.5) which are evaluated along the membrane variable at specific points in the domain coefficients of the PDE and the conductance model has to be such that the solution space is embedded in $C^0([0,T],C^0(\overline{D}))$. On the other hand, considering the less restrictive transition rates (3.1.7) it is sufficient that the membrane PDE possesses solutions in $C^0([0,T],L^2(D))$.

3.1.2 Models without spatial dynamics

If we consider a space-clamped neuron model, that is, the transmembrane potential is assumed to be homogeneous in space, $u(t,x)\equiv u(t)\in\mathbb{R}$ for all $x\in D$, the abstract evolution equations (3.1.2) reduce to an ordinary differential equation at every point in $x\in D$, cf. (1.2.2). Integrating these equations over the spatial domain yields the family of one-dimensional ODEs

$$\dot u = C^{-1}\sum_i g_i(\theta)\,(E_i-u)\,,\ \theta\in K, \tag{3.1.8}$$

where the conductances are given by $g_{\mathrm{L}}(\theta)=\overline{g}_{\mathrm{L}}$ and

$$g_{\mathrm{Na}}(\theta)=\int_D g_{\mathrm{Na}}(\theta,x)\,\mathrm{d}x=\overline{g}_{\mathrm{Na}}n_{n_4},\quad g_{\mathrm{K}}(\theta)=\int_D g_{\mathrm{K}}(\theta,x)\,\mathrm{d}x=\overline{g}_{\mathrm{K}}n_{m_3h_1}.$$

Here n_c denotes the number of channels in state $c\in\mathcal{C}$ out of all channels in the current configuration θ. Thus the conductances in the space-clamped model are given by the conductance of a single channel times the number of channels in the open state, cf. (1.2.4). This is the prevalent method of modelling conductance in hybrid models and algorithms [110, 32, 38, 105, 31, 96]. We remark that it is easy to see that the boundedness of the membrane component within $[\overline{u}_-,\overline{u}_+]$ also holds for equations (3.1.8) with initial conditions $u(0)\in[\overline{u}_-,\overline{u}_+]$.
Further, as all rates of all single channels in the same state coincide the rate of a transition in the configuration (3.1.5) reduces to the simpler form

$$\lambda(u(t),\theta)=\sum_{c\in\mathcal{C}} n_c q_c(u(t))\,. \tag{3.1.9}$$

Here the summation is over the different states the channels can be in. Moreover, only the different possible transitions for one channel in each state have to be considered independent of the individually identified channels. Thus, the distribution of the occurring configurational change is fully described by the probabilities

$$\mu\big((u(\tau),\theta),\{\widehat{\theta}\}\big) = \frac{n_c q_{c\to\widehat{c}}(u(\tau))}{\lambda(u(\tau),\theta)}. \tag{3.1.10}$$

In this case $\widehat{\theta}$ denotes the configuration one obtains from the initial configuration θ if one arbitrary channel in state c switches to $\widehat{c}$. Hence, the dynamics of the configurational changes are completely described by the number of channels in the individual states alone. It is thus sufficient for space-clamped models to consider a reduced state space $K = \{1,\dots,n\}^{|\mathcal{C}|}$, where each element $\theta \in K$ is a $|\mathcal{C}|$-tuple with every entry recording the number of channels in one specific state $c \in \mathcal{C}$. We remark that simulation algorithms for space-clamped membranes that make use of this reduced description are called *channel number tracking* algorithms in the literature [31]. This is in contrast to algorithms that keep track of all the individual channels called *channel state tracking* algorithms, which are necessary for models including spatial time-evolution.

As finite-dimensional PDMPs can be considered a special case of infinite-dimensional PDMPs we can readily conclude from Theorem 3.1.1 that the above model of a space-clamped membrane is a standard PDMP $(Y_t,\theta_t)_{t\geq 0}$ on the phase space $\mathbb{R}\times\{1,\dots,n\}^{|\mathcal{C}|}$ for all initial conditions $u(0)\in[\overline{u}_-,\overline{u}_+]$ in the sense of Definition 3.0.2.

3.2 Compartmental-type models

In this section we present a second, conceptionally different approach to spatial hybrid membrane models which we call *compartmental-type models*. To the best of our knowledge this class of models have not been analytically described in the literature so far, however, there is a close connection to ad-hoc models used in numerical simulation studies, cf. a discussion at the end of this section.

As for exact spatial models discussed in the preceding section, we consider each single channel, however, group together those channels that are physically close to each other. Further, we assume that for channels close to each other their switching rates are the same and are governed by the average membrane potential over an area around the channels. Moreover, we suppose that the membrane current due to the single channels can be treated as the total current over the group of channels. These assumptions are motivated from considering channel current and switching rates as modelled by (3.1.4) and (3.1.7). For two channels very close to each other the neighbourhoods overlap in a large area and essentially coincide. Hence, the current over this patch of membrane is essentially given by the sum of the two single currents and the switching rates of the channels also effectively coincide. In particular for myelinated neurons where ion channels are closely packed at Nodes of Ranvier but effectively absent at the inter-nodal, myelinated areas this approach provides a reduction in complexity compared to exact models but are still expected to retain a high accuracy. Further, these classes of spatial models are also of particular interest to calcium dynamics for which it is a prominent feature that calcium channels form clusters on the membrane, i.e., isolated spots where individual channels are packed together very closely.

We now define the model components. As before let $D \in \mathbb{R}$ be a bounded interval for which we consider a partition D_k, $k = 1, \ldots, n$ by mutually disjoint, Lebesgue measurable subintervals such that $D = \bigcup_{k=1}^{n} D_k$. Further, each compartment D_k contains a fixed number of channels which are in their specific states $c \in \mathcal{C}$. As channels in the same compartment possess the same stochastic dynamics there is no need to track the state of each individual channel, i.e., use channel tracking, but it is sufficient to record the number of channels in the individual states for each compartment. Hence, θ_c^k denotes the number of channels in compartment D_k which are in state c. Thus, the total channel configuration is described by the $n|\mathcal{C}|$–dimensional vector $\theta = (\theta_c^k)_{k=1,\ldots,n, c\in\mathcal{C}}$ taking values in a finite state space K. Note that the number of channels in each compartment is fixed over time and it is clearly possible that there are no channels in certain compartments. To finish the discussion of the channels it remains to consider the switching rates and the transition measure which we define analogously to (3.1.7). Thus the rate of a channel in compartment D_k to switch from state c to state $\widehat{c}$ is given by

$$q_{c\to\widehat{c}}\Big(|D_k|^{-1}\int_{D_k} u(t,x)\,\mathrm{d}x\Big)\,.$$

Hence the rate such that a change in the configuration θ occurs of one channel in compartment D_k to switch from state c to state $\widehat{c}$ is given by

$$\theta_c^k\, q_{c\to\widehat{c}}\Big(|D_k|^{-1}\int_{D_k} u(t,x)\,\mathrm{d}x\Big)\,.$$

Clearly the total rate Λ is thus given by the sum over all individual rates, i.e.,

$$\Lambda(u,\theta) = \sum_{k=1}^{m}\sum_{c\in\mathcal{C}} \theta_c^k\, q_{c\to\widehat{c}}\Big(|D_k|^{-1}\int_{D_k} u(t,x)\,\mathrm{d}x\Big)\,, \tag{3.2.1}$$

and the transition measure μ is defined by the point probabilities accordingly. That is, the probability that a certain event occurs is given by the rate of this event divided by the total rate.

Finally, it remains to state an appropriate definition of the conductances g_i for the membrane equation (3.1.2) which in the case of the Hodgkin-Huxley model are given by

$$g_{\mathrm{Na}}(x,\theta) = \overline{g}_{\mathrm{Na}} \sum_{k=1}^{n} \theta^k_{m_3h_1}\, \mathbb{I}_{D_k}(x), \quad g_{\mathrm{K}}(x,\theta) = \overline{g}_{\mathrm{K}} \sum_{k=1}^{n} \theta^k_{n_4}\, \mathbb{I}_{D_k}(x)\,, \tag{3.2.2}$$

which satisfy $g_{\mathrm{Na}}(x,\theta)$, $g_{\mathrm{K}}(x,\theta) \in L^2(D)$. All remaining unspecified properties are assumed to be as for Theorem 3.1.1.

Theorem 3.2.1. *For initial conditions $u_0 \in L^2(D)$ compartmental-type hybrid membrane models are standard infinite-dimensional PDMPs on $L^2(D)$. Moreover, the continuous component satisfies for all $t \geq 0$ that $u(t,x) \in [\overline{u}_-, \overline{u}_+]$ for almost all $x \in \overline{D}$.*

Proof. This follows by the same arguments as employed to prove Theorem 3.1.1. □

We conclude this section by a brief discussion of compartmental-type models. Firstly, we note that these type of models can be immediately generalised to higher spatial dimensions, particular domain geometries and more general membrane equations with different boundary conditions. The decisive property is that the solutions of the membrane property for any possible conductance generate a semigroup in $L^2(D)$ which is usually implied by the most basic existence and uniqueness theorems for weak solutions.

In some sense compartmental-type models occupy an intermediate position between exact models as discussed in Section 3.1 and models without spatial domain, see Section 3.1.2. For compartmental models we can think, heuristically speaking, of space being 'discretised' in the channel model and ion channels in the same space compartment being 'lumped together'. Regarding the channels each compartment is then treated as a model without spatial extension and the transmembrane potential at this 'point' in space is given by the average potential across the compartment. However, the channel dynamics of the different compartments are coupled by the membrane equation which extends across all compartments and influences the local dynamics. Further, spatially discretising also the membrane equation, e.g., using finite differences, yields a finite-dimensional PDMP system which is the description of axon models that are used in numerical studies on the effects of channel noise on axon reliability, see, e.g., [45, 44, 94]. However, we note that in implementations these authors always employ the pseudo-exact formulation, i.e., substituting the exact waiting time distributions by exponentially distributed waiting times, cf. Section 1.2.3. Finally, similar models are also employed for stochastic models of chemically reacting particles which also perform a diffusive motion, see, e.g., [6] and references therein.

3.3 Deterministic models of excitable media

In this section we first briefly discuss the Hodgkin-Huxley model to illustrate the class of excitable media models. Then in Section 3.3.1 we prove *well-posedness*[2] for a general excitable media system. This class of equations includes as special cases further biophysical realistic neuron models, e.g., the Morris-Lecar model, models of cardiac cells, see [48] for a comprehensive collection of models, and models in calcium dynamics, e.g., the DeYoung-Kaizer model. The common feature of all these models based on the common underlying cell physiology is that the dynamics of channel proteins are coupled and influenced by a macroscopic variable, the transmembrane potential or the calcium concentration, which itself is influenced by the dynamics of the channels.

The Hodgkin-Huxley model was introduced by A. L. Hodgkin and A. F. Huxley [62] as a model for the experimentally observed electrical properties of the squid giant axon. Hodgkin and Huxley received for their work the *Nobel Price in Medicine and Physiology* in 1963. The model incorporates electrical currents across the neuronal membrane due to sodium (Na) and potassium (K) ions and a constant leakage current (L). In its spatially homogeneous form the model is given by a four-dimensional system of coupled nonlinear ordinary differential equations. Based on physical principles they derived the membrane balance equation

$$C\dot{u} = \overline{g}_{\mathrm{Na}}\, m^3 h\, (E_{\mathrm{Na}} - u) + \overline{g}_{\mathrm{K}}\, n^4\, (E_{\mathrm{K}} - u) + \overline{g}_{\mathrm{L}}\, (E_{\mathrm{L}} - u) + I(t) \tag{3.3.1}$$

describing the time evolution of the transmembrane potential of a space-clamped membrane. That is, the variable u measures the displacement of the transmembrane potential from an equilibrium value which is usually set to zero. The parameter values in (3.3.1) were experimentally fitted ($\overline{g}_i$ denoting the maximal conductance of the membrane with respect to the specific current and E_i is the *Nernst equilibrium*). Further, the variables m, n, h in the nonlinear terms m^3h, n^4 denoting the strength of the conductance ($\sim$ the fraction of open channels) satisfy the coupled equations

$$\begin{aligned} \dot{m} &= a_m(u)(1-m) - b_m(u)\, m, \\ \dot{h} &= a_h(u)(1-h) - b_h(u)\, h, \\ \dot{n} &= a_n(u)(1-n) - b_n(u)\, n\,. \end{aligned} \tag{3.3.2}$$

Here m corresponds to an activation of the sodium current, h an inactivation of the sodium current and n is the potassium current. The rates a_i, b_i along with the powers

[2] As usual *well-posedness* (of an initial-value problem) is understood as the existence of a unique solution which depends continuously on the initial data.

in m^3h, n^4 were experimentally fitted to current measurements, however, they proved to have a physiological interpretation. The powers correspond to four subunits of a single channel. For the channel to be in an open state each subunit has to be in one of two possible states. We remark that the Hodgkin-Huxley model also arises as the limit of more accurate stochastic models, see, e.g., [50] for a derivation from an master equation expansion but, in particular, [96, 119] for a mathematically precise limit theorem. The parameter values for the standard Hodgkin-Huxley model are reported in Appendix A.

In order to quantitatively describe and explain the propagation of an impulse along the axon, Hodgkin and Huxley also introduced in [62] a spatial version of their model based on the cable equation. Due to the physical shape of an axon, a very long and extremely thin object, it is sufficient to consider only one space dimension. For a further discussion of the approximation that yields this model, its limitations and references to primary literature we refer to [74]. The equation is given as follows

$$C\dot{u} = \frac{r}{2R}\,\Delta u + \overline{g}_{\mathrm{Na}}\, m^3 h\,(E_{\mathrm{Na}} - u) + \overline{g}_{\mathrm{K}}\, n^4\,(E_{\mathrm{K}} - u) + \overline{g}_{\mathrm{L}}\,(E_{\mathrm{L}} - u) + I(t)\,, \quad (3.3.3)$$

where m, h, n are again given by the coupled system (3.3.2). Here the potential variable u is dependent on space and time, i.e., $u : [0,\infty) \times D \to \mathbb{R}$ where $D \subseteq \mathbb{R}$ is a one-dimensional interval. Note that the second part in (3.3.3) is just the right hand side of the membrane balance equation (3.3.1) in the model without spatial dimension. Thus, the cable equation is an example of a reaction-diffusion equation with reaction term given by the right hand side of the membrane balance equation. The aim of the equation is to describe the propagation of an action potential along the axon, which, in dynamical terms, is a *travelling wave solution* to the coupled system (3.3.3), (3.3.2).

A straightforward generalisation of the original Hodgkin-Huxley model yields *general conductance based models* which incorporate an arbitrary number $m \in \mathbb{N}$ different current sources through a membrane. To this end we consider a general membrane balance equation

$$\dot{u} = \sum_{i=1}^{m} g_i(p)\,(E_i - u) + I(t)\,, \quad (3.3.4)$$

where the non-negative conductances $g_i : \mathbb{R} \to \mathbb{R}_+$ are functions depending on time-dependent gating variables $p \in \mathbb{R}^m$. Naturally the generalised spatial version on a domain $D \subseteq \mathbb{R}^d$ of this equation is given by

$$\dot{u} = Au + \sum_{i=1}^{m} g_i(x,p)\,(E_i - u) + I(t,x) \quad (3.3.5)$$

where $I(t,x)$ is a space-time dependent externally applied current and A is a second order elliptic partial differential operator (C.0.2). As before p are space-time dependent gating variables and the conductances $g_i : \overline{D} \times \mathbb{R} \to \mathbb{R}_+$ may also be space dependent. There is no modelling reason at this point to consider operators A that include first and zeroth order terms, i.e., b_i, $c \neq 0$ in (C.0.2). However, there is also no mathematical reason not to include these operators, as the theory of partial differential equations of the type (3.3.5) is generally developed for arbitrary second order elliptic operators.

For the system defining the gating variables there are usually two possibilities. A first one is the definition of the gating variables by solution of the uncoupled, inhomogeneous system

$$\dot{p}_i = a_i(u)(1 - p_i) - b_i(u)\, p_i\,, \tag{3.3.6}$$

where the rates a_i, b_i are $\mathbb{R} \to \mathbb{R}_+$ and in general unbounded and locally Lipschitz. In particular, this is the form of the system for the gating variables as originally introduced by Hodgkin and Huxley, cf. (3.3.2), and encountered almost exclusively when deterministic models are discussed in the literature. However, a second possible choice is a coupled linear system derived from a reaction rate approximation, e.g., the van Kampen system size expansion [118], of the dynamics of the channels. That is, we obtain the coupled, linear system

$$\dot{p}_i = \sum_{j=1}^{m} p_j\, q_{ji}(u) \tag{3.3.7}$$

where each $q_{ji} : \mathbb{R} \to \mathbb{R}$ is continuous, unbounded and locally Lipschitz. Further, $Q(u) = (q_{ji}(u))_{j,i=1,\dots,m}$ is a matrix with diagonal entries which are non-positive and off-diagonal elements which are non-negative for all $u \in \mathbb{R}$. That is, in vector form we can write the system (3.3.7) as $\dot{p} = Q(u)^T p$. Further, the matrix Q satisfies $Q(u)e = 0$ for all $u \in \mathbb{R}$, i.e., $q_{jj} = -\sum_{i \neq j} q_{ji}$. This implies that $e^T p(t) = const.$ over time and – usually, as it is physiologically meaningful – it holds that $e^T p(t) \in \mathbb{N}$. This integer corresponds to the number the different families of ion channels in the model. Then each entry of p_j, corresponding to one state of a channel from a particular family, is the proportion of channels out of all channels in this family which are in this particular state. The value q_{ji} corresponds to the rate of a channel switching from state j to state i. We note that for analytical purposes it is sufficient to consider gating variables that satisfy $e^T p = 1$, that is, we do not distinguish between different channel families.[3] The property $Q(u)e = 0$ then states that the total mass, i.e., the sum of all proportions being 1, is conserved over time.[4] This stems from the interpretation

[3] When gating systems of the form (3.3.7) are considered then the conductances g_i are usually linear in p. Hence a simple transformation of variables always yield an equivalent system such that $e^T p = 1$.

[4] In a probabilistic interpretation $Q(u)$ is, for fixed u, the intensity matrix of the continuous time

of a set of individual channels as a chemical reaction system. The different channel states are different 'molecules' and a 'reaction' is a state change of a channel. It is clear that in this system a 'reaction' does not destroy or generate a 'molecule'. Thus the total number is conserved.

Further, for both cases of the gating systems the variables p_i are interpreted as fractions of channels in a particular state. Therefore, the values of the variables should remain bounded within $[0,1]$. Assume that the gating variables are continuously differentiable with derivatives specified by the right hand sides of (3.3.6) or (3.3.7), respectively. Then it is easy to see that $p_i(t) \in [0,1]$ for all $t \geq 0$ as long as $p_i(0) \in [0,1]$, as the derivative evaluated at $p_i = 1$ is negative and positive for $p_i = 0$. Additionally, normally the transmembrane potential variable u remains within a bounded interval $[\overline{u}_-, \overline{u}_+]$ for initial conditions inside $[\overline{u}_-, \overline{u}_+]$. This property was first proven in [42] for the original Hodgkin-Huxley axon model (without current input). After the proof of well-posedness of excitable media equations in the next section we also present a proposition that makes these arguments precise.

We conclude this section by a brief comment on the connection between the two types of equations for the gating system. The first form of the gating system (3.3.6) is usually obtained from a system of the form (3.3.7) by an appropriate transformation of variables. The transformation then yields a state space reduction corresponding to an equilibrium distribution of channels over their respective states within one family. Technically this means that for appropriate initial condition the solution of the gating system (3.3.7) remains on a manifold which can be described by a lower dimensional equation. For example, in the Hodgkin-Huxley model this reduction technique reduces the full 13-dimensional model to a one-dimensional equation for the K-channel and a two-dimensional uncoupled system for the Na-channel, cf. (3.3.2). This is well known and exemplified, e.g., in [96]. Therefore we refer to the system (3.3.6) as the *reduced gating system* and to the system as the (3.3.7) the *full gating system*. In terms of channel modelling the dimension of the reduced gating system corresponds to the number of different, independent subunits in the channel models. Then the fraction of open channels is given by the product of fractions of open subunits.
As normally a substantial simplification of the model is achieved it is clear that the reduced form of the gating system is preferred in the deterministic literature on excitable media equations. However when dealing with stochastic equations the full form is important as connections of hybrid stochastic to deterministic models can only be derived using the full gating system, cf., [7, 50, 96]. We note that it is an important point to distinguish between the two gating system, particularly, when

Markov chain which models a single channel. The vector $p(t)$ contains the point probabilities of the channel being in a particular state at time t and the system $\dot{p} = Q(u)^T p$ is the corresponding *Kolmogorov forward equation*.

a top-down approach to channel noise is applied. Adding any type of noise to the reduced gating system *cannot be justified* by a Langevin, linear noise or moment equation approximation. These bottom-up methods, see Section 1.2.2, always lead to an SDE version of the full gating system, see [96] in the context of stochastic hybrid systems and, e.g., [89] for a discussion of general Langevin approximation to the chemical master equation.

3.3.1 Well-posedness of deterministic equations for excitable media

To conclude this brief introduction into deterministic axon equations we consider the well-posedness of the general cable equation (3.3.5) as an initial-boundary value problem. Theorem 3.3.1 extends earlier results derived in [42, 81, 87, 102]. Therein the authors consider excitable media systems on the whole real line [42], bounded one-dimensional domains [81, 87] and bounded domains of up to dimension three [102]. However, all studies are restricted to the case of the linear operator A being the Laplacian and for the nonlinear term in the membrane equation and the gating system either a Lipschitz assumption is introduced [42, 102] or the particular choice for the Hodgkin-Huxley model is considered [81, 87]. The result we present in Theorem 3.3.1 avoids these restrictions and we aim for sufficient generality including most general possibilities of excitable media system. On a more technical note the authors in [42, 81, 87] approach existence and uniqueness within the semigroup approach to PDEs whereas we, as always in this thesis, deal with weak solutions to PDEs. We note that Hodgkin-Huxley type systems are also considered in the monographs [111, 116]. Therein well-posedness and invariant regions, cf. the discussion of the pointwise bounds on the solutions in the preceding section, are proved. However, the particular equations discussed are different to the system (3.3.3), (3.3.2), as diffusion terms are included in the gating variables.
In the following, D denotes, as always, an open, bounded domain in $\mathbb{R}^d$ where due to physical relevance it is sufficient to restrict the spatial dimension to $d \leq 3$. The most interesting cases are a one-dimensional cable $d = 1$ for axon equations or a two-dimensional surface $d = 2$ for models of cardiac tissue and in calcium dynamics. In order for the general cable equation to be well-posed we introduce boundary conditions that govern the behaviour of the potential variable u on the boundary ∂D of the domain. In particular, to be consistent with the presentation in the preceding section we consider *Dirichlet boundary conditions*, i.e.,

$$u(t,x) = 0 \quad \forall\, x \in \partial D,\, t \geq 0\,. \tag{3.3.8}$$

This type of boundary conditions refers to an experimental setup where the nerve end is cut open and thus the membrane potential is equal to the surrounding environment (*killed end boundary condition*), hence constant with value zero. We note that usually in excitable media equations $u(t,x) \equiv 0$ is a steady state for the membrane balance equation. Before we state a general existence and uniqueness theorem for excitable media systems, we first briefly collect a list of technical assumptions on the coefficients of the general excitable media equation. These are normally satisfied for specific examples of excitable media systems such as the Hodgkin-Huxley system. The proof of Theorem 3.3.1 is deferred to Section 3.4.2.

(EC1) Let A be an *elliptic second order differential operator* (C.0.2) with *time-independent coefficients* $a_{ij}, b_i, c \in L^\infty(D)$ for $i,j \leq d$, which, in case of $d = 2,3$, are additionally *twice continuously differentiable* on $\overline{D}$. Further, when $d = 1$ the spatial domain D is a bounded interval and when $d = 2,3$ it is a bounded C^3–domain.

(EC2) The conductances $g_i : \overline{D} \times \mathbb{R}^m : (x,p) \mapsto g_i(x,p)$ are for all $i = 1,\dots,m$ *globally bounded for* $x \in D$ and *locally bounded and locally Lipschitz continuous in* p. Further, they are *twice partially differentiable with respect to* p having *derivatives locally bounded in* p *and globally bounded for* $x \in D$.[5]

(EC3) The rate functions $q_{ij} : \mathbb{R} \to \mathbb{R}_+$, $i \neq j$, for the full gating system (3.3.7) or $a_i, b_i : \mathbb{R} \to \mathbb{R}_+$, $i,j = 1,\dots,m$, for the reduced gating system (3.3.6) are *locally bounded*, *bounded away from zero*, *locally Lipschitz continuous* and satisfy a *polynomial growth condition*, i.e., there exists a $\gamma \in \mathbb{N}$ and $C > 1$ such that

$$|q_{ij}(y)| + |a_i(y)| + |b_i(y)| < C(1 + |y|^\gamma) \qquad \forall\, y \in \mathbb{R},\ i,j = 1,\dots,m\,. \tag{3.3.9}$$

Moreover, they are *twice continuously differentiable* with *derivatives being locally bounded and locally Lipschitz continuous.*[6]

(EC4) The initial condition p_0 for the gating variable satisfies $p_0 \in [0,1]^m$. If the full gating system (3.3.7) is considered then the initial conditions satisfy additionally $e^T p = 1$.

[5]For functions arising in models of excitable media these conditions are usually satisfied. In these models using the reduced gating system these functions are polynomials, thus obviously differentiable. For example, in the Hodgkin-Huxley model we have $p = (m,h,n)$ and $g_{\mathrm{Na}} = \overline{g}_{\mathrm{Na}} m^3 h$ and $g_{\mathrm{K}} = \overline{g}_{\mathrm{K}} n^4$. Hence, the gradients are given by $\nabla g_{\mathrm{Na}} = (3\overline{g}_{\mathrm{Na}} m^2 h, \overline{g}_{\mathrm{Na}} m^3)$ and $\nabla g_{\mathrm{K}} = 4\overline{g}_{\mathrm{K}} n^3$ which have the bounds $\nabla g_{\mathrm{Na}} \cdot (m,h,n) \leq 4\overline{g}_{\mathrm{Na}}$ and $\nabla g_{\mathrm{Na}} \cdot (m,h,n) \leq 4\overline{g}_{\mathrm{K}}$ for $(m,h,n) \in [0,1]^3$. For the full gating system g_{Na} and g_{K} are even linear in p and so the conditions are also satisfied.

[6]We note that the rate functions defined on $\mathbb{R}$ may not be bounded away from zero or satisfy a polynomial growth condition, cf., e.g., the rate functions of the Hodgkin-Huxley model given in Appendix A which are exponentially increasing and decrease to zero for either $u \to \infty$ or $u \to -\infty$. This, however, does not make the existence theorem inappropriate. Usually, the values of the trans-membrane potential variable are bounded pointwise within a finite interval, cf. Proposition 3.3.1. Hence, changing the rate functions outside this interval does not change the solution. In this way we easily find that rate functions are bounded away from zero and satisfy a polynomial growth condition. Moreover, they could even be chosen bounded.

Theorem 3.3.1. *Assume that conditions* (EC1)–(EC4) *are satisfied and that the system* (3.3.5) *coupled to either the full or reduced gating systems, that is, to* (3.3.7) *or to* (3.3.6)*, is equipped with Dirichlet boundary conditions* (3.3.8)*.*

(a) *Let $d = 1$, then for every $T > 0$, every set of initial conditions $u_0 \in H^1_0(D)$ and $p_i(0) \in H^1(D)$, $i = 1, \dots, m$, satisfying the conditions connected with the respective type of gating system, and any input current $I \in L^2((0,T), L^2(D))$, the generalised excitable media equation possesses a unique weak solution (u, p) satisfying*

$$u \in L^2((0,T), H^2(D)) \cap H^1((0,T), L^2(D)) \subset C^0([0,T], H^1_0(D))$$

$$p_i \in C^1([0,T], H^1(D)) \quad \forall\, i = 1, \dots, m.$$

(b) *Let $d = 2, 3$, then for every $T > 0$, for every set of initial conditions $u_0 \in H^1_0(D) \cap H^2(D)$ and $p_i(0) \in H^2(D)$, $i = 1, \dots, m$, satisfying the conditions connected with the respective typo of gating system, and any input current $I \in H^1((0,T), L^2(D)) \cap L^2((0,T), H^1_0(D))$, the generalised excitable media equation possesses a unique weak solution (u, p) satisfying*

$$u \in L^2((0,T), H^3(D)) \cap H^1((0,T), H^1(D)) \subset C^0([0,T], H^2(D)) \cap C^1([0,T], L^2(D))$$

$$p_i \in C^1([0,T], H^2(D)) \quad \forall\, i = 1, \dots, m.$$

In addition, for both (a) *and* (b) *the solution (u, p) depends continuously on the initial data u_0, p_0 and the current input I.*

Finally, as has been already heuristically argued we can derive a-priori pointwise bounds on the solutions of the general membrane equation. Therefore we obtain an invariant rectangle for solutions of the excitable media system. We propose bounds $\overline{u}_-$, $\overline{u}_+ \in \mathbb{R}$ for the solution satisfying the following relations

$$\begin{aligned} c(x)\overline{u}_- + \sum_{i=1}^m g_i(x,p)(E_i - \overline{u}_-) + I(t,x) \geq 0 \quad \forall\, x \in D,\, t \in [0,T],\, p \in [0,1]^m, \\ c(x)\overline{u}_+ + \sum_{i=1}^m g_i(x,p)(E_i - \overline{u}_+) + I(t,x) \leq 0 \quad \forall\, x \in D,\, t \in [0,T],\, p \in [0,1]^m, \end{aligned} \tag{3.3.10}$$

where c is the zeroth order coefficient of the operator A. Clearly, the existence of such bounds is guaranteed if c is non-positive and together with the input current I pointwise bounded. Moreover, for classical conductance-based models the operator A is a pure diffusion operator (in particular, $c \equiv 0$) and in the absence of any current input we can choose the bounds $\overline{u}_- = \min E_i$ and $\overline{u}_+ = \max E_i$ as were first proved for membrane equations in [42].

Proposition 3.3.1. *Assume that* $\overline{u}_-$, $\overline{u}_+$ *satisfy conditions* (3.3.10). *Then, if the initial condition satisfy* $u_0(x) \in [\overline{u}_-, \overline{u}_+]$ *and* $p_0(x) \in [0,1]^m$ *for all* $x \in \overline{D}$, *the solutions in the sense of* Theorem 3.3.1 *remain within this rectangle, i.e.,*

$$u(t,x) \in [\overline{u}_-, \overline{u}_+],\ p(t,x) \in [0,1]^m \qquad \forall\,(t,x) \in [0,T] \times \overline{D}\,.$$

Here each $u(t)$, $p(t)$ *is identified with the unique continuous element of its equivalence class.*

Proof. The result for the gating variables is already a by-product of Theorem 3.3.1, see Lemma 3.4.1, and the result for the variable u follows from an application of Theorem 3.4.2 which proves such an assertion in a more general way. □

Remark 3.3.1. Firstly, the regularity obtained for the solutions in Theorem 3.3.1 is physically meaningful. The Sobolev Embedding Theorems guarantees that for $d = 1$ it holds that $H^1(D) \hookrightarrow C^0(\overline{D})$ and for $d \leq 3$ it holds that $H^2(D) \hookrightarrow C^0(\overline{D})$ and these embeddings are optimal. Thus the solution can be identified with the unique element in the equivalence class that is continuous over time as a Banach space valued function with values $u(t)$ that are continuous functions over the spatial domain. As the variable u denotes a difference in an electrical potential or, as in the case of calcium dynamics, the concentration of some chemical substance, it is physically reasonable to require u to be continuous and bounded. Further, employing standard methods of deriving higher spatial and temporal regularity of solutions, see, e.g., [43], we could proceed demanding higher regularity on the coefficients and current input function as well as the initial condition. However, we do not need them in this thesis so we do not pursue this path any further.
Secondly, we have considered the membrane balance equation equipped with Dirichlet boundary conditions (3.3.8). Certainly, other types of boundary conditions also make perfect sense for excitable media equations, e.g., *Neumann* or *Robin boundary conditions*. Neumann boundary conditions that set the outward normal at the boundary to zero are referred to as *sealed end* boundary conditions for neuron models. These refer to models of neurons such that the end of the fibre is covered with neuronal membrane. Its resistance is so large that it can be assumed infinite. Finally, arbitrary Neumann or Robin boundary conditions give the leak current through the boundary and describe more accurately the physical situation. For a further discussion of boundary condition with respect to the (linear cable equation) we refer to [74, Chap. 2]. We note that in [102] the well-posedness for a particular excitable media system $-A = \Delta$, $m = 1$, and the nonlinear reaction term in the membrane equation and the right hand side

in the gating system are globally Lipschitz continuous – with boundary conditions on the outward normal given by $\frac{\mathrm{d}u}{\mathrm{d}n} = a(x)u$ is discussed for spatial dimensions $d \le 3$. The result is analogous to the above Theorem 3.3.1 (b) and the author states that it can be extended in a straightforward way to more general systems.
Thirdly, Theorem 3.3.1 only covers a one-dimensional macroscopic variable, i.e., $u(t,x) \in \mathbb{R}$, which is the setup for equations of excitable media modelling usual neuronal membranes or cardiac tissue. In models which additionally take also calcium dynamics into account it may be necessary to have multi-dimensional macroscopic variables, e.g., one for the transmembrane potential and a second for the calcium concentration. We expect the existence and regularity results of Theorem 3.3.1 to extend to these cases.

3.4 Section Appendix

3.4.1 The Austin model

In this section we briefly discuss the particular type of membrane balance equation considered in [7]. This type of model is restricted to one-dimensional spatial domains only, i.e., in this section D is an open, non-empty interval in $\mathbb{R}$. The initial-boundary value problem which models the evolution of the membrane potential in between state changes of single ion channels is given by

$$\begin{aligned} \dot{u} &= Av + \sum_{i=1}^{m} g_i \cdot (E_i - u) \cdot \delta_{x_i}, \\ u(t,x) &= 0 \qquad \forall\, x \in \partial D,\ t \ge 0, \\ u(0) &= u_0\,, \end{aligned} \tag{3.4.1}$$

where $A : H_0^1(D) \to H^{-1}(D)$ is a coercive operator, g_i are smooth, non-negative functions and $E_i \in \mathbb{R}$ are constants. The model in [7] is a particular case where $A = \Delta$ and $g_i(x) \equiv \overline{g}_i$ are constants. This type of models corresponds to using (3.1.3) as a model for the conductances and thus δ_{x_i} in (3.4.1) is the Dirac delta function with mass at the point $x_i \in D$. In order to prove well-posedness we employ Theorem C.0.2 and thus consider the equation (3.4.1) as an abstract evolution equation

$$\dot{u} = \widehat{A}u + f\,,$$

where the linear operator A and the inhomogeneous term f are given by

$$\widehat{A}u = Au - \Big(\sum_{i=1}^{n} g_i \cdot \delta_{x_i}\Big) \cdot u, \qquad f = \sum_{i=1}^{n} E_i g_i(x) \delta_{x_i}\,. \tag{3.4.2}$$

As usual the evolution triple is given by $H_0^1(D) \subset L^2(D) \subset H^{-1}(D)$. Before proving well-posedness in Proposition 3.4.2 below, we briefly comment on the use of the Dirac delta function. Being a generalised function it is well defined as a linear functional on the space of test functions $C_c^\infty(D)$. However, an application of the Dirac delta function to a function v makes sense for all continuous functions $v \in C^0(\overline{D})$ and thus, particularly, to elements of Sobolev spaces $H^s(D)$ where $s > d/2$. Moreover, due to continuous embedding of $H^s(D)$ into $C^0(\overline{D})$ it follows that $\delta \in H^{-s}(D)$. Further, a pointwise multiplication of generalised functions is well-defined for smooth functions, hence f as defined in (3.4.2) is in $H^{-1}(D)$. However, for the definition of the operator $\widehat{A}$ in (3.4.2) this assumption needs to be relaxed.

Proposition 3.4.1. *Let $D \subset \mathbb{R}$ be a bounded interval. Then, for $u \in C^0(\overline{D})$ the linear functional $u \cdot \delta_x$ defined as*

$$\langle u \cdot \delta_x, \varphi \rangle = u(x)\varphi(x)$$

for test functions φ on D is a distribution. Moreover, $u \cdot \delta_x$ is a distribution for all $u \in H^1(D)$ and $u \cdot \delta_x \in H^{-1}(D)$.

Remark 3.4.1. The same remains true for bounded domains $D \subset \mathbb{R}^d$ with sufficiently regular boundary and Sobolev spaces $H^s(D)$ with $s > d/2$.

Proof. Firstly, if $\varphi_n \to \varphi$ then $(u \cdot \delta_x, \varphi_n) \to (u \cdot \delta_x, \varphi)$, hence $u \cdot \delta_x$ is a distribution. Secondly, due to the Sobolev Embedding Theorem $H^1(D) \hookrightarrow C^0(\overline{D})$ and hence $u \cdot \delta_x$ is a distribution for all $u \in H^1(D)$. It remains to show that $u \cdot \delta_x$ belongs in fact to H^{-1}, thus is a linear, bounded functional on $H^1(D)$. Linearity is clear and boundedness follows as for every $v \in H^1(D)$

$$|\langle u \cdot \delta_x, v \rangle| = |u(x)v(x)| \leq \|uv\|_\infty \leq C^2 \, \|u\|_{H^1} \|v\|_{H^1} \,,$$

where C is the constant from the continuous embedding $H^1(D) \hookrightarrow C^0(\overline{D})$. □

The next proposition gives well-posedness of the initial-boundary value problem (3.4.1) the proof of which we have deferred after the discussion of the result compared to [7].

Proposition 3.4.2. (a) *The operator $\widehat{A}$ satisfies $\widehat{A} \in L(H_0^1(D), H^{-1}(D))$ and the associated quadratic form is coercive.*

(b) *The equation* (3.4.1) *is well posed, i.e., for every initial condition $u \in L^2(D)$ it possesses a unique weak solution $u \in L^2((0,T), H_0^1(D)) \cap H^1((0,T), H^{-1}(D)) \subset C^0([0,T], L^2(D))$, which depends continuously on the initial condition in $L^2(D)$.*

Remark 3.4.2. This existence result extends the hybrid approach of using Dirac delta functions as current inputs to a much broader class of one-dimensional models. Further, the result obviously remains valid when adding an input current

$I \in L^2((0,T), L^2(D))$ to the right hand side of the equation (3.4.1). However, we note that Proposition 3.4.2 is not the same existence result as postulated in Austin [7] wherein there the existence of a (mild) solution $u \in C^0([0,T], H_0^1(D))$ is proved. This additional regularity, which implies $u(t) \in C^0(\overline{D})$ for very $t \in [0,T]$, is indispensable for the definition of the model with choices of jump rates and Markov kernels given by (3.1.3). Unfortunately, we cannot use the usual methods of improving the regularity of weak solutions based on the general Hilbert space theory of abstract evolution equations to infer $u \in C^0([0,T], H_0^1(D))$. These require $\dot{u} \in L^2((0,T), L^2(D))$, which by definition of the problem (3.4.1) does not hold. In order to prove the additional regularity Austin employs a completely different method of proof. His approach depends on the specific form of the membrane equation (3.1.2), in particular the choice of $A = \Delta$ and the explicit knowledge of the semigroup generated by the Laplacian with Dirichlet boundary conditions. The essential step are estimates of the norm of the semigroup applied to the Dirac delta function in the variational formula of the solution. If this can be proved for more general semigroups then his model may be extendable to general equations and multi-dimensional domains. For the sake of completeness we cite his result.

Proposition 3.4.3. [7, Prop. 3.4] *The initial-boundary value problem* (3.4.1) *with* $A = \Delta$ *and* g_i *being constant has a unique mild solution* $u \in C^0([0,T], H_0^1(D))$ *for every initial condition* $u \in H_0^1(D)$.

Further, it is an immediate consequence of the definition of a mild solution that it depends continuously on the initial condition. Therefore, we can use the result in Proposition 3.4.3 for the definition of inter-jump dynamics for a PDMP.
As stated it remains open how the approach by Austin can be used in order to generalise the model to membrane equations of the type (3.4.1). That is, e.g., equations on two-dimensional domains, different types of boundary conditions and more general diffusion operators A. A further investigation in this direction is not within the scope of this thesis which provides a general modelling framework. In any case, any results will heavily depend on the knowledge of properties of the semigroup that is generated by the linear part of the equation and such detailed investigations should be strongly motivated by a particular application that demands this particular model.

Proof of Proposition 3.4.2. First note that given part (a) the statement in part (b) is a consequence of Theorem C.0.2. Thus it remains to prove (a), i.e., the operator $\widehat{A}$ is linear, bounded and coercive, i.e., the associated quadratic form

$$a(u,v) := -\langle \widehat{A}u, v\rangle_{H^1} = -\langle Au, v\rangle_{H^1} + \sum_{i=1}^{n} \langle g_i \cdot \delta_{x_i} \cdot u, v\rangle_{H^1} \quad \forall\, u,v \in H_0^1(D), \quad (3.4.3)$$

satisfies for all $u \in H_0^1(D)$

$$a(u,u) \geq C_1 \|u\|_{H^1}^2 - C_2 \|u\|_{L^2}^2$$

for some constants $C_1, C_2 > 0$.

Firstly, due to Proposition 3.4.1 the product $g_i \cdot \delta_x \cdot u$ is in $H^{-1}(D)$ for any $u \in H_0^1(D)$. Hence, the operator $\widehat{A}$ does map into $H^{-1}(D)$. Next, the linearity of $\widehat{A}$ is obvious, thus it remains to consider boundedness which for linear operators between Banach spaces is equivalent to continuity. Let u_n be a sequence converging to u strongly in $H_0^1(D)$, then

$$\begin{aligned} \|\widehat{A}u_n - \widehat{A}u\|_{H_0^{-1}} &= \left\| A(u_n - u) - \sum_{i=1}^{n} g_i \cdot \delta_{x_i} \cdot (u_n - u) \right\|_{H_0^{-1}} \\ &\leq \|A(u_n - u)\|_{H^{-1}} + \sum_{i=1}^{n} \|g_i \cdot \delta_{x_i} \cdot (u_n - u)\|_{H^{-1}}. \end{aligned}$$

The estimate in the right hand side converges to zero for $\|u_n - u\|_{H^1} \to 0$ as, on the one hand, the operator A is bounded by assumption, and, on the other hand, for each summand in the second term in the right hand side we have the estimates

$$\begin{aligned} \|g_i \cdot \delta_{x_i} \cdot (u_n - u)\|_{H^{-1}} &= \sup_{\substack{v \in H_0^1(D) \\ \|v\|_{H^1} = 1}} |\langle g_i \cdot \delta_{x_i} \cdot (u_n - u), v \rangle_{H_0^1}| \\ &= \sup_{\substack{v \in H_0^1(D) \\ \|v\|_{H^1} = 1}} |g_i(x_i)(u_n(x_i) - u(x_i)) v(x_i)| \\ &\leq C \left| u_n(x_i) - u(x_i) \right| \left| g_i(x) \right|. \end{aligned}$$

Here C is the constant arising from the Sobolev Embedding Theorem which yields $|v(x_i)| \leq \|v\|_0 \leq C\|v\|_{H^1}$ for all $x_i \in D$ due to the continuous embedding of $H_0^1(D)$ into $C^0(\overline{D})$. Finally, $|g_i(x_i)|$ is bounded due to the smoothness of the functions g_i and $|u_n(x) - u(x)| \to 0$ for $u_n \to u$ in $H_0^1(D)$ again due the Sobolev Embedding Theorem.

Therefore, it is left to prove the coercivity condition for the quadratic form a. Due to the coercivity of the operator A there exist constants $C_1, C_2 >$ such that for all

$u \in H_0^1(D)$ we obtain the estimate

$$\begin{aligned} a(u,u) &= -\langle Au, u\rangle_{H^1} + C\sum_{i=1}^{n}\langle g_i \cdot \delta_{x_i} \cdot u, u\rangle_{H_0^1} \\ &\geq C_1\|u\|_{H^1}^2 - C_2\|u\|_{L^2}^2 + \sum_{i=1}^{n} g_i(x_i)u^2(x_i) \\ &\geq \frac{D}{c}\|u\|_{H^1}^2 \end{aligned}$$

as $g_i(x_i)u^2(x_i) \geq 0$ for all $i = 1, \ldots, n$. Hence the coercivity condition is satisfied and the proof is completed. □

3.4.2 Proof of Theorem 3.3.1 (Well-posedness of det. equations)

In this section we prove Theorem 3.3.1 stating the well-posedness of the coupled system of equations for deterministic models of excitable media. The proof employs a fixed point argument following the general outline used in [81]. In contrast to the approaches by [42, 87, 102] this approach directly addresses the problem of having non-globally Lipschitz reaction term as is frequently the case in models of excitable media.

We present the proof of parts (a), the case of $d = 1$, and (b), the case of $d = 2, 3$, of the theorem simultaneously, as well as simultaneously consider the full and the reduced gating system. Essentially the techniques employed are analogous for the different spatial dimension and the different types of gating system. If we need to distinguish between the two cases then the different points are marked by (a) and (b). The proof is split into the following steps and we present intermediate results in individual lemmata. First we present in (A) preliminary considerations and initial definitions. Then, in (B) we consider for a given solution of the cable equation the well-posedness of the gating systems (3.3.7) and (3.3.6), respectively. This part of the proof contains a major difference in technique to the approach of [81] on which we comment at the appropriate place. Next, in (C) we show for given solutions of the gating system that equation (3.3.5) is well-posed and the map allocating a solution to the membrane equation to every solution of the gating system is a contraction. Finally, in part (D) first a local existence result is proved using a fixed point argument. Then the local existence result is used to prove global existence. The proof is concluded sketching the proof of the continuous dependence of the solution on the initial data. The additional effort that arises in parts (C) and (D) concerns estimation procedures which for multi-dimensional domains have to be extended from the space $L^2((0,T), L^2(D))$, which is sufficient for one-dimensional domains, to Sobolev spaces of higher order in the temporal as well as the spatial domain.

Before we start the actual proof we recall the conditions imposed in Theorem 3.3.1 and straight-forward consequences. For ease of notation we omit space domains when dealing with function spaces, i.e., we abbreviate, e.g., $L^2 := L^2(D)$, $H_0^1 := H_0^1(D)$ and $C^0 := C^0(\overline{D})$. First of all, the conditions on the initial data are

(a) $u_0 \in H_0^1$, $p_1(0), \dots, p_m(0) \in H^1$ and $I \in L^2((0,T), L^2)$;

(b) $u_0 \in H_0^1 \cap H^2$, $p_1(0), \dots, p_m(0) \in H^2$ and $I \in H^1((0,T), L^2) \cap L^2((0,T), H^1)$.

For simplicity we use for the reaction term in the membrane equation (3.3.5) the notation

$$f_1(x,p)u + f_2(x,p) := \sum_{i=1}^m g_i(x,p)\,(E_i - u)\,, \tag{3.4.4}$$

where $f_1, f_2 : \overline{D} \times \mathbb{R}^m \to \mathbb{R}$. Given the conditions (EC2) in Section 3.3.1 these functions satisfy the Carathéodory condition, that is $x \mapsto f_i(x,p)$ is measurable for all $p \in \mathbb{R}^m$ and $p \mapsto f(x,p)$ is continuous for almost all $x \in D$. This implies that the composition $(t,x) \mapsto f_i(x, p(t,x))$ is jointly measurable for any $p(t,x)$ being jointly measurable. Further f_1, f_2 satisfy all the conditions assumed for the conductances g_i in (EC2). Clearly, the same applies to the rate functions of the gating system. Particularly, the compositions $q_{ij}(u(t,x))$, $a_i(u(t,x))$ and $b_i(u(t,x))$ are jointly measurable functions for any jointly measurable u.

Further, throughout the proof we use $\overline{q}(\epsilon)$ and $L(\epsilon)$ to denote a local bound and a local Lipschitz condition on particular functions which satisfy such a condition, i.e., the rate functions q_{ij}, a_i, b_i with respect to u, the functions f_1, f_2 and their derivatives as specified in (EC2) and (EC3) with respect to p. We use the same constants for all of these functions. The argument ϵ is used to indicate the dependence on local bounds of the arguments of the functions which should be self-explanatory in the context. For example, in part (A) we introduce for every $\epsilon > 0$ the pointwise bounds $K + \epsilon K_1$ on the membrane variable for fixed K, K_1 and a dependence on this bound is denoted only by ϵ, i.e.,

$$|q_{ij}(u)| + |q_{ij}^{(1)}(u)| \le \overline{q}(\epsilon) \quad \forall\, |u| \le K + K_1\epsilon\,. \tag{3.4.5}$$

or

$$|q_{ij}(u) - q_{ij}(v)| + |q_{ij}^{(1)}(u) - q_{ij}^{(1)}(v)| \le L(\epsilon)\,|u - v| \quad \forall\, |u|, |v| \le K + K_1\epsilon\,. \tag{3.4.6}$$

When dealing with bounds or Lipschitz conditions on f_i and its derivatives with respect to the variable p we omit the dependence of the bound and the Lipschitz constant on an argument ϵ as it will always be the case that the arguments p are bounded by one, i.e., for example

$$|f_i(x,p)| + |\partial_{p_k} f_i(x,p)| \le \overline{q} \qquad \forall\, p \in [0,1]^m,\ x \in D$$

and

$$|f_i(x,p) - f_i(x,q)| \leq L \sum_{i=1}^{m} |p_i - q_i| \qquad \forall\, p,q \in [0,1]^m,\; x \in D\,.$$

Finally, we briefly comment on the *chain and product rule* in connection with weak derivatives. In general we distinguish between weak derivatives of real-valued functions on D with respect to the spatial variables and L^2-valued functions on a time-interval $[0,T]$. In the first case the chain rule holds in the sense that for $f \in C^1(\mathbb{R})$ such that $f' \in L^\infty(\mathbb{R})$ and $u \in H^1(D)$ it holds that $f(u) \in H^1(D)$ and $D_{x_i}(f(u)) = f'(u) D_{x_i} u$. Obviously, when u is bounded, then it is sufficient that f' is bounded over the values u attains. The product rule holds in the sense that for $u, v \in H^1(D)$ such that $uv \in L^2(D)$ and $u'v$, $uv' \in L^2(D)$ it follows that $uv \in H^1(D)$ and $(uv)' = u'v + uv'$.
Finally, we need the chain rule for L^2-valued functions. Let $f \in C^1(\mathbb{R})$ such that $f' \in L^\infty(\mathbb{R}))$ and $u \in H^1((0,T), L^2)$ then

$$\frac{\mathrm{d}}{\mathrm{d}t} f(u(t)) = f'(u(t))\dot{u}(t) \in L^2$$

in the sense that $(\frac{\mathrm{d}}{\mathrm{d}t} f(u(t)))(x) = f'(u(t,x))\,\dot{u}(t,x)$ almost everywhere on D. Obviously in case $u(t,x)$ is pointwise bounded it is sufficient that f' is bounded over the range of values p attains. The product rule holds analogously, i.e., if $u, v \in H^1((0,T), L^2)$ and uv, $\dot{u}v$, $u\dot{v} \in L^2((0,T), L^2)$ then also $uv \in H^1((0,T), L^2)$ and the weak derivative is given by $\dot{u}v + u\dot{v}$.

(A) The passive membrane equation

Initially we consider the solution of the membrane balance equation uncoupled to the gating system, i.e.,

$$\begin{aligned} \dot{w} &= Aw + I(t), \\ w(t,x) &= 0 \qquad \forall\,(t,x) \in [0,T] \times \partial D, \\ w(0) &= u_0\,, \end{aligned} \tag{3.4.7}$$

which is obviously a special case of the full system (3.3.5) when $\bar{g}_i = 0$. The system (3.4.7) is the equation for a passive membrane. By assumption the linear operator A is uniformly elliptic and the inhomogeneous term I is in $L^2((0,T), L^2)$, therefore due to Theorem C.0.2 there exists a unique weak solution $w \in L^2((0,T), H_0^1) \cap C^0([0,T], L^2)$ for every initial condition $u_0 \in L^2$. In addition, under the conditions as presented in Theorem 3.3.1 (a) and (b) we obtain the improved regularity due to Theorem C.0.3 and Remark C.0.1:

(a) For an initial condition $u_0 \in H_0^1$ the solution satisfies $w \in C^0((0,T), H_0^1)$ and in particular $w(t) \in H_0^1$ for all $t \in [0,T]$. As the Sobolev Embedding Theorem implies for $d = 1$ that $H_0^1 \hookrightarrow C^0$, there exists an element $\widehat{w}$ in the equivalence class of the solution w such that $\widehat{w}(t) \in C^0$ and $\sup_{x\in\overline{D}} |\widehat{w}(t,x)| \leq K$ for all $t \in [0,T]$ and a suitable constant $K < \infty$. We always identify the solution with this unique element $\widehat{w}$ of its equivalence class.

(b) For an initial condition $u_0 \in H_0^1 \cap H^2$ and an inhomogeneous term $I \in H^1((0,T), L^2) \cap L^2((0,T), H^1)$ the solution satisfies $w \in C^0([0,T], H^2)$. Again the Sobolev Embedding Theorem then implies for $d \leq 3$ that $H^2 \hookrightarrow C^0$. Hence there exists an element $\widehat{w}$ in the equivalence class of the solution such that $\widehat{w}(t) \in C^0$ and $\sup_{x\in\overline{D}} |w(t,x)| \leq K$ for all $t \in [0,T]$. We always identify the solution with this unique element of its equivalence class.

Note that by identifying the solution with the particular element of its equivalence class the solution is uniquely defined pointwise. Moreover, the regularity of the solution guarantees that there exists a constant $K > 0$ which bounds the solution pointwise, i.e., for all $t \in [0,T]$ and all $x \in \overline{D}$

$$|w(t,x)| \leq K\,.$$

Given the unique solution w of (3.4.7) we define for the cases (a) and (b) the sets

$$\begin{aligned} B^{(a)}(\epsilon,T) &:= \big\{v \in L^2((0,T),L^2)\,:\, \|w-v\|_{L^\infty((0,T),H_0^1)} \leq \epsilon\big\}\,,\\ B^{(b)}(\epsilon,T) &:= \big\{v \in L^2((0,T),L^2)\,:\, \|w-v\|_{L^\infty((0,T),H^2)} \leq \epsilon\big\}\,. \end{aligned} \tag{3.4.8}$$

That is, these sets are the subset of functions in $L^2((0,T),L^2)$ which are inside a closed ϵ-ball in the spaces $L^\infty((0,T),H_0^1)$ or $L^\infty((0,T),H^2)$, respectively, around the solution w of the passive membrane equation (3.4.7). Due to the continuous embedding of these spaces into $L^2((0,T),L^2)$ the sets $B^{(a)}(\epsilon,T)$ and $B^{(b)}(\epsilon,T)$ are closed in $L^2((0,T),L^2)$. Moreover, being a closed subset of a complete metric space they are a complete metric space itself. For simplicity we use the notation $B^\bullet(\epsilon,T)$ when referring to either $B^{(a)}(\epsilon,T)$ or $B^{(b)}(\epsilon,T)$.

Finally, we obtain a pointwise bound almost everywhere for elements $v \in B^\bullet(\epsilon,T)$. Note that the Sobolev Embedding Theorem implies $H^1 \hookrightarrow C^0$ when $d = 1$ and $H^2 \hookrightarrow C^0$ when $d \leq 3$. Hence, $B^\bullet(\epsilon,T) \subset L^\infty((0,T)\times L^\infty)$ and it holds for all $v \in B^\bullet(\epsilon,T)$

$$\|v\|_{L^\infty((0,T)\times D)} \leq \|w\|_{L^\infty((0,T)\times D)} + \|v-w\|_{L^\infty((0,T)\times D)} \leq K + K_1\epsilon$$

for some constant K_1 resulting from the Sobolev Embedding Theorem. Moreover, on the one hand, it holds that for almost all $t \in (0, T)$

$$\|v(t)\|_{L^\infty} \leq \|w(t)\|_{L^\infty} + \|v(t) - w(t)\|_{L^\infty} \leq K + K_1\epsilon\,.$$

On the other hand, it holds that for almost all $x \in D$ the map $t \mapsto v(t,x)$ is bounded by $K + K_1\epsilon$ for almost all $t \in (0, T)$.

(B) The gating system

Next we consider for a given transmembrane potential trajectory $v \in B^\bullet(\epsilon, T)$ solutions to the non-autonomous initial value problem resulting from the system of gating variables. That is, with respect to the two types of gating system, we consider either the linear, non-autonomous system

$$\begin{aligned} \dot{p} &= Q(v(t))^T\, p, \\ p(0) &= p_0 \end{aligned} \tag{3.4.9}$$

corresponding to the full, coupled gating system (3.3.7) or for every $i = 1, \ldots, m$ the inhomogeneous, non-autonomous equation

$$\begin{aligned} \dot{p}_i &= a_i(v(t))\,(1 - p_i) - b_i(v(t))\,p_i, \\ p_i(0) &= p_i^0 \end{aligned} \tag{3.4.10}$$

corresponding to the reduced, uncoupled gating system (3.3.6). The following lemma provides the existence of a unique solution to these systems and their continuous dependence on $v \in B^\bullet(\epsilon, T)$. We note that in the following lemma the distinguished statements in the parts (a) and (b) relate to parts (a) and (b) of Theorem 3.3.1. This is understood in the sense that for proving part (a) of the theorem we need part (a) of the following lemma, analogously for parts (b), and each part only assumes the conditions as in the corresponding parts of the Theorem 3.3.1. This convention is also employed in the further lemmata and theorems in the remainder of the proof.

Lemma 3.4.1. (a) *For every $v \in B^{(a)}(\epsilon, T)$ and every set of initial conditions $p_i^0 \in H^1$, $i = 1, \ldots, m$, a unique solution p to the gating system exists, such that the components satisfy*

$$p_i \in H^1((0,T), H^1) \subset C^0([0,T], H^1) \quad \forall\, i = 1, \ldots, m\,.$$

If in addition $v \in C^0([0,T], H^1)$ then it holds that $p_i \in C^1((0,T), H^1)$ for every $i = 1, \ldots, m$.

(b) *For every $v \in B^{(b)}(\epsilon, T)$ and every set of initial conditions $p_i^0 \in H^2$, $i = 1, \ldots, m$,* *a unique solution p to the gating system exists, such that the components satisfy*

$$p_i \in H^1((0,T),H^2) \subset C^0([0,T],H^2) \quad \forall\, i=1,\dots,m\,.$$

If in addition $v \in C^0([0,T],H^2)$ then it holds that $p_i \in C^1((0,T),H^2)$ for every $i=1,\dots,m$.

Further, for $p_0(x) \in [0,1]^m$ for all $x \in D$ it holds that $p(t,x) \in [0,1]^m$ for all $x \in D$ and every $t \in [0,T]$. In addition, if in the case of the full gating system $\sum_{i=1}^m p_i(0,x) = 1$ for all $x \in D$, then it holds $\sum_{i=1}^m p_i(0,x) = 1$ for all $x \in D$ and every $t \in [0,T]$. Moreover, the solutions in (a) *and* (b) *satisfy for two trajectories v, $u \in B^\bullet(\epsilon,T)$ and for two sets of initial condition $p_i^v(0)$, $p_i^u(0)$ that it holds for all $t \in [0,T]$*

$$\sum_{i=1}^m \|p_i^v(t) - p_i^u(t)\|_{L^p} \leq C\left(\sum_{i=1}^m \|p_i^v(0) - p_i^u(0)\|_{L^p} + \int_0^t \|v(s) - u(s)\|_{L_p}\,\mathrm{d}s\right), \quad (3.4.11)$$

where $1 \leq p \leq \infty$. Further, for all t, $h > 0$ such that $t+h \leq T$ the Lipschitz-type inequality

$$\sum_{i=1}^m \|p_i^v(t+h) - p_i^v(t)\|_{L^p} \leq C\left(h + \int_0^t \|v(s+h) - v(s)\|_{L_p}\,\mathrm{d}s\right) \quad (3.4.12)$$

holds, where the constant $C < \infty$ depends only on ϵ and T. In case (b)*, if in addition $v \in C^0([0,T],H^1)$, then it holds that*

$$\sum_{i=1}^m \|p_i^v(t+h) - p_i^v(t)\|_{H^1}^2 \leq C\Big(h^2 + \int_0^t \|v(s+h) - v(s)\|_{H^2}^2\,\mathrm{d}s\Big)\,, \quad (3.4.13)$$

where the constant $C < \infty$ depends on ϵ, T, $\|v\|_{C^0([0,T],H^1)}$ and $\sum_{i=1}^m \|p_i(0)\|_{H^1}$. In particular, the constant C in each of (3.4.11) – (3.4.13) *is finite for finite values of the quantities it depends on and the dependence is monotonic.*

Proof of Lemma 3.4.1. We consider the gating system as an equivalent system of integral equations, which is coupled in the full form and decouples in the reduced form. We only discuss the full gating system in detail. The results for the reduced gating system follow by fully analogous arguments. Moreover, as the reduced gating system usually arises as a transformation of the full gating system, existence and uniqueness of the reduced system can be inferred from the existence and uniqueness of a corresponding full system. The central differences of the present setup and the setup [81] is that we consider a coupled system instead of scalar equations. Further, the proposed regularity of the solution to the excitable media equation in spatial dimensions $d \leq 3$ necessitates that we derive higher spatial regularity results for the gating system, i.e., solutions with components taking values in H^1 in the case of

(a) and H^2 in the case of (b), compared to [81] wherein only solutions in L^2 are derived. In the subsequent proof we employ the Banach Fixed Point Theorem to establish existence and uniqueness of solutions and the proposed regularity. This is a completely different approach than employed in [81]. For the same reason that allows lesser spatial regularity in [81], that is, a restriction to one-dimensional domains, the Lipschitz-type inequalities (3.4.11) and (3.4.12) are only considered for the space L^2 therein. For the excitable media system discussed here the general inequalities (3.4.11) and (3.4.12) in L^p are necessary. Moreover, additionally we have to derive the inequality (3.4.13) in the case of (b).

Successively, we prove in the following that (i) in case (a) and (b) there exists a unique solution using the Banach Fixed Point Theorem and show (ii) that this solution satisfies the proposed pointwise bounds. The dependence on the initial data and the Lipschitz-type inequalities (3.4.11) and(3.4.12) are shown in part (iii) and the additional inequality (3.4.12) in case of (b) is proven in (iv). For simplicity of notation we omit the superscript v denoting a fixed element $v \in B^\bullet(\epsilon, T)$ whenever there is no risk of confusion.

(i) We prove existence and uniqueness of a solution via the Banach Fixed Point Theorem in the space $C^0([0,T], H^1)$ in case of (a) and in the space $C^0([0,T], H^2)$ in case of (b) for all $T > 0$. Here the gating system (3.4.9) is considered an integral equation in the Hilbert space $(H^1)^m$ or $(H^2)^m$, respectively. To this end we first show (i.1) the necessary prerequisite that $v \in B^\bullet(\epsilon, T)$ implies $q_{ij}(v) \in L^\infty((0,T), H^1)$ or $q_{ij}(v) \in L^\infty((0,T), H^2)$, respectively. This guarantees that the right hand side of (3.4.9) takes values in the respective Sobolev space. Then part (i.2) presents the existence result and (i.3) provides the improved regularity for continuous paths v.

(i.1) First of all, as q_{ij} is locally bounded it holds that $\|q_{ij}(v(t))\|_{L^2} \leq \overline{q}(\epsilon)$. Secondly, for the first derivatives we get that

$$\begin{aligned} \left\| D_{x_k} q_{ij}(v(t)) \right\|_{L^2}^2 &= \left\| q_{ij}^{(1)}(v(t)) D_{x_k} v(t) \right\|_{L^2}^2 \\ &\leq \overline{q}_{ij}(\epsilon)^2 \left\| D_{x_k} v(t) \right\|_{L^2}^2 . \end{aligned}$$

Overall, these estimates imply that there exists a constant C depending on ϵ such that

$$\|q_{ij}(v(t))\|_{H^1}^2 \leq C(\epsilon)(1 + \|v(t)\|_{H^1}^2) \qquad \text{for a.e. } t \in (0,T).$$

That is, for case (a) and (b) it holds that $\|q_{ij}(v)\|_{L^\infty((0,T),H^1)}$ is bounded by a bound depending only on ϵ and the norm $\|v\|_{L^\infty((0,T),H^1)}$. Finally, for the stronger result in case (b) it remains to additionally consider the second derivatives of $q_{ij}(v(t))$:

$$D_{x_k x_l}\big(q_{ij}(v(t,x))\big) \;=\; q_{ij}^{(2)}(v(t,x))D_{x_l}v(t,x)D_{x_k}v(t,x) + q_{ij}^{(1)}(v(t,x))D_{x_k x_j}v(t,x)\,.$$

Employing the uniform bounds and Cauchy-Schwarz inequality we obtain that

$$|D_{x_k x_l}q_{ij}(v(t,x))|^2 \;\leq\; \overline{q}(\epsilon)^2\left(|D_{x_l}v(t,x)|^4 + |D_{x_k}v(t,x)|^4\right) + \overline{q}(\epsilon)^2|D_{x_k x_j}v(t,x)|^2$$

almost everywhere in D. Taking the integral over D we find

$$\|D_{x_k x_l}q_{ij}(v(t))\|_{L^2}^2 \;\leq\; C(\epsilon)\left(\|D_{x_l}v(t)\|_{L^4}^4 + \|D_{x_k}v(t)\|_{L^4}^4 + \|D_{x_k x_j}v(t)\|_{L^2}^2\right)$$

for a suitable constant C depending only on ϵ. As the Sobolev Embedding Theorem implies that $H^1 \hookrightarrow L^4$ for $d \leq 3$ we find that

$$\|D_{x_k x_l}q_{ij}(v(t))\|_{L^2}^2 \;\leq\; C(\epsilon)\left(\|D_{x_l}v(t)\|_{H^1}^4 + \|D_{x_k}v(t)\|_{H^1}^4 + \|v(t)\|_{H^2}^2\right)$$

for a suitable constant depending only on ϵ. This implies $q_{ij}(v(t)) \in H^2$ for $v(t) \in H^2$. Furthermore, it follows $\|q_{ij}(v(t))\|_{H^2}$ is bounded depending only on ϵ and $\|v(t)\|_{H^2}$. This in turn implies that $q_{ij}(v(t)) \in L^\infty((0,T),H^2)$ with norm depending only on ϵ and $\|v\|_{L^\infty((0,T),H^2)}$.

(i.2) Next we proceed to prove existence of a solution to the gating system. In the following we use X to denote either the space H^1 or H^2. We define for every initial condition $y \in X^m$ and every $v \in B^\bullet(\epsilon,T)$ an operator $V^{y,v}$ by

$$V^{y,v} : p \in C^0([0,T],X^m) \mapsto Vp$$

where for all $i = 1,\ldots,m$ and $t \in [0,T]$

$$(V^{y,v}p)_i(t) := y_i + \int_0^t \sum_{j=1}^m p_j(s)q_{ji}(v(s))\,\mathrm{d}s\,. \tag{3.4.14}$$

We first show that the integral in the right hand side is well defined in the sense of Bochner. It then follows that $V^{y,v}$ maps the Banach space $C^0([0,T],X^m)$ into itself. Obviously the integrand in (3.4.14) is measurable as a sum of a pointwise product of measurable functions. Furthermore, the pointwise multiplication of functions is a bilinear, continuous map for the Sobolev spaces in question, i.e., for H^1 in the case $d = 1$ and H^2 in the case $d \leq 3$. Thus it holds for a suitable constant $K < \infty$ that

$$\big\|p_j(s)q_{ji}(v(s))\big\|_X \leq K\,\|q_{ji}(v(s))\|_X\,\|p_j(s)\|_X \;\leq K\,\|q_{ji}(v)\|_{L^\infty((0,T),X)}\,\|p_j\|_{C^0([0,T],X)}\,. \tag{3.4.15}$$

Here the right hand side is an integrable majorant and we infer that the integrand in (3.4.14) is integrable.
Clearly, the map $t \mapsto (V^{y,v}p)(t)$ is continuous and hence $V^{y,v}p \in C^0([0,T],X^m)$. We next show that $V^{y,v}$ is even a Lipschitz continuous operator on $C^0([0,T],X^m)$. Let $p, \widehat{p} \in C^0([0,T],X^m)$ then

$$\begin{aligned}\|V^{y,v}p - V^{y,v}\widehat{p}\|_{C^0([0,T],X^m)} &= \sup_{t\in[0,T]} \sum_{i=1}^m \Big\| \int_0^t \sum_{j=1}^m \big(p_j(s) - \widehat{p}_j(s)\big)\, q_{ji}(v(s))\,\mathrm{d}s \Big\|_X \\ &\leq \sup_{t\in[0,T]} \sum_{i,j=1}^m \int_0^t \big\| \big(p_j(s) - \widehat{p}_j(s)\big)\, q_{ji}(v(s)) \big\|_X \,\mathrm{d}s\,.\end{aligned}$$

Due to (3.4.15) we further obtain

$$\begin{aligned}\|V^{y,v}p - V^{y,v}\widehat{p}\|_{C^0([0,T],X^m)} &\leq \Big(mT \max_{j,i} \|q_{ji}(v)\|_{L^\infty((0,T),X)} \Big) \sup_{t\in[0,T]} \sum_{i=1}^m \|p_i(t) - \widehat{p}_i(t)\|_X \\ &\leq \Big(mTK \max_{j,i} \|q_{ji}(v)\|_{L^\infty((0,T),X)} \Big) \|p - \widehat{p}\|_{C^0([0,T],X^m)}\,.\end{aligned}$$

Note that $\max_{j,i} \|q_{ji}(v)\|_{L^\infty((0,T),X)}$ is finite due to part (i.1). Furthermore for small enough T^*, i.e.,

$$T^* < \frac{1}{mK \max_{j,i} \|q_{ji}(v)\|_{L^\infty((0,T),X)}}\,,$$

the map $V^{y,v}$ is a contraction on the Banach space $C^0([0,T^*],X^m)$. In particular note that T^* is independent of the initial condition y. Now the Banach Fixed Point Theorem guarantees the existence of a unique fixed $p \in C^0([0,T^*],X^m)$ such that

$$p(t) = y + \int_0^t Q(v(s))^T p(s)\,\mathrm{d}s \text{ in } X^m \quad \forall\, t \in [0,T^*],$$

which also implies $p(0) = y$. It is straightforward to extend this local existence result to a global existence result as the time T^* does not depend on the initial condition: For $2T^* < T$ we obtain a solution on $[0,2T^*]$ by defining $t \to p(T^*+t)$ to be the unique fixed point of the operator

$$(V\widetilde{p})(t) = p(T^*) + \int_0^t Q(v(T^*+s))^T \widetilde{p}(s).$$

Repeating this procedure we obtain a unique solution $p \in C^0([0,T],X^m)$ for every $T > 0$.
Finally, as p is a solution of an integral equation we infer that it is absolutely contin-

uous and satisfies

$$\dot{p}_i(t) = \sum_{j=1}^{m} p_j(t) q_{ji}(v(t)) \quad \text{for a.e. } t \in [0,T],\, i = 1,\dots,m\,.$$

This gives $p_i \in H^1((0,T),X)$.

(i.3) We assume that in fact $v \in C^0([0,T],X)$. Then it holds that $p_i \in C^1([0,T],X)$: Estimating the term $\|q_{ji}(v(t)) - q_{ji}(v(s))\|_X$ analogously to part (i.1) we find that $v \in C^1([0,T],X)$ implies $q_{ji}(v) \in C^0([0,T],X)$. Then it is straightforward to infer that the pointwise product $p_j q_{ji}(v)$ is continuous in X and hence $\dot{p}_i$ is a continuous function.

(ii) We next prove the proposed pointwise bounds $p_i(t,x) \in [0,1]$ and the conservation property $\sum_i p_i(t,x) = 1$ which hold for almost all x and all t. Note that the Bochner integral can be evaluated pointwise as $p_i \in L^2((0,T),L^2)$ which is isomorphic to $L^2((0,T)\times D)$. That is for any $i = 1,\dots,m$ and all $t \in [0,T]$ it holds that

$$\Big(\int_0^t \sum_{j=1}^{m} p_j(s) q_{ji}(v(s))\,\mathrm{d}s\Big)(x) = \int_0^t \sum_{j=1}^{m} p_j(s,x) q_{ji}(v(s,x))\,\mathrm{d}s \qquad \text{for a.e. } x \in D\,.$$

We first show the conservation property. To this end we define a linear, bounded operator $\mathcal{T} : (L^2)^m \to L^2$ by $\mathcal{T}p = \sum_{i=1}^m p_i$. Then, as the Bochner integral allows for changing integration and the application of linear, bounded operators, it holds that

$$\begin{aligned}
\mathcal{T}p(t) &= \mathcal{T}p(0) + \int_t^0 \mathcal{T}\Big(Q(v(s)^T p(s))\Big)\,\mathrm{d}s \\
&= \sum_{i=1}^{m} p_i(0) + \int_0^t \sum_{j=1}^{m} p_j(s)\Big(\sum_{i=1} q_{ji}(v(s))\Big)\,\mathrm{d}x \\
&= 1 + 0\,.
\end{aligned}$$

Thus $\mathcal{T}p(t)(x) = 1$ for almost all $x \in D$ which in fact holds for all x as due to the regularity of p and the Sobolev Embedding Theorem it holds that $\mathcal{T}p(t) \in C^0(\overline{D})$.

We next show that the solution remains pointwise within the interval $[0,1]$. By definition $p_i(0,x) \in (0,1)$. Then let $t_0 > 0$ be the infimum of times $t > 0$ such that for some $i = 1,\dots,m$ either $p_i(t,x) \geq 1$ or $p_i(t,x) \leq 0$ on a set $D' \subset D$ of non-zero Lebesgue measure. If t_0 finite then we always take $T > t_0$ in the following.

First, we assume that t_0 is finite and the former holds, i.e., it holds that $p_i(t_n,x) \geq 0$ on a set D' of non-zero Lebesgue measure for a sequence of times $t_n \geq t_0$ decreasing to t_0. Next, the continuity $p_i \in C^0([0,T],L^2)$ implies that the set of all x such that $p_i(t_0,x) > 1$ is a set of measure zero: We assume otherwise which implies that on this set $p_i(t_0) - 1$ is positive. Then, the assumption that for all $t < t_0$ it holds that $p_i(t,x) \leq 1$ for almost all x is a contradiction to $p_i \in C^0([0,T], L^2)$. Analogously, also $p_i(t_0) = 1$ on a set of measure zero is a contradiction to $p_i \in C^0([0,T], L^2)$. Hence, we infer that t_0 is the first time such that $p_i(t_0, x) = 1$ on a set $D' \subset D$ of non-zero Lebesgue measure. This characterisation of t_0 implies that there exists a $\delta > 0$ such that

$$0 < \sum_{j=1} p_j(t,x) q_{ji}(v(t,x)) = p_i(t,x) q_{ii}(v(t,x)) + \sum_{j \neq i} p_j(t,x) q_{ji}(v(t,x)) \qquad (3.4.16)$$

for almost all $t \in (t_0 - \delta, t_0)$ and almost everywhere on D'. As the rate functions q_{ji}, $j \neq i$, are bounded and bounded away from zero, say larger than $\overline{q}_{\min}$, and as $q_{ii} = -\sum_{j \neq i} q_{ij}$ it follows that

$$p_i(t,x) q_{ii}(v(t,x)) + \sum_{j \neq i} p_j(t,x) q_{ji}(v(t,x)) < -(m-1)\, \overline{q}_{\min}\, p_i(t,x) + \overline{q} \sum_{j \neq i} p_j(t,x)\,.$$

Now as $p_j \in C^0([0,T], L^2)$ for all $j = 1, \ldots, m$ and $\sum_{j=1}^m p_j = 1$ it follows that the upper bound in the right hand side becomes negative for t close enough to t_0 which is a contradiction to (3.4.16). Hence, t_0 cannot be finite. An analogous argument yields that t_0 cannot be finite also in case it is the infimum of times such that $p_i(t,x) \leq 0$ for some $i = 1, \ldots, m$. Overall, we have proved that for all $t \geq 0$ it holds that $p_i(t,x) \in [0,1]$ for almost all $x \in D$. Moreover, as the Sobolev Embedding Theorem implies that $p_i(t) \in C^0(\overline{D})$ the bounds hold everywhere on D.

(iii) Next we show the continuous dependence of the solution p on v, i.e., (3.4.11), and the Lipschitz-type inequality (3.4.12). We reintroduce the upper indices for the gating variables to indicate the given $v \in B^\bullet(\epsilon, T)$. Thus, let p^u, p^v denote the unique solutions of the gating system with respect to two functions $u, v \in B^\bullet(\epsilon, T)$ and note that embedding theorems imply that $p^u, p^v \in C^0([0,T], L^\infty)$ and $u, v \in L^1((0,T), L^p)$ for all $p \leq \infty$. Thus we obtain for all $i = 1, \ldots, m$ the estimate

$$\|p_i^v(t) - p_i^u(t)\|_{L^p} \leq \|p_i^v(0) - p_i^u(0)\|_{L^p} + \int_0^t \sum_{j=1}^m \left\| p_j^v(s) q_{ji}(v(s)) - p_j^u(s) q_{ji}(u(s)) \right\|_{L^p} \mathrm{d}s\,. \qquad (3.4.17)$$

It remains to estimate the summands in the integrand in the right hand side. Using

$$\begin{aligned}\big\|p_j^v(s)q_{ji}(v(s)) - p_j^u(s)q_{ji}(u(s))\big\|_{L^p} &\le \big\|\big(p_j^v(s) - p_j^u(s)\big)q_{ji}(v(s))\big\|_{L^p}\\ &\quad + \big\|p_j^u(s)\big(q_{ji}(v(s)) - q_{ji}(u(s))\big)\big\|_{L^p}\\ &\le \|q_{ji}(v(s))\|_{L^\infty}\,\big\|p_j^v(s) - p_j^u(s)\big\|_{L^p}\\ &\quad + \|p_j^u(s)\|_{L^\infty}\,\big\|q_{ji}(u(s)) - q_{ji}(u(s))\big\|_{L^p}\end{aligned}$$

we obtain due the pointwise bounds almost everywhere for $q_{ji}(v)$, cf. (3.4.5), and p_j^v and the Lipschitz continuity of the rate functions q_{ji}, cf. (3.4.6), that

$$\big\|p_j^v(s)q_{ji}(v(s)) - p_j^u(s)q_{ji}(u(s))\big\|_{L^p} \le \overline{q}(\epsilon)\,\big\|p_j^v(s) - p_j^u(s)\big\|_{L^p} + L(\epsilon)\,\big\|v(s) - u(s)\big\|_{L^p}\,. \tag{3.4.18}$$

We use this estimate for the integrands in the right hand side of (3.4.17) and sum the inequalities (3.4.17) over all $i = 1, \ldots, m$ to obtain

$$\begin{aligned}\sum_{i=1}^m \|p_i^v(t) - p_i^u(t)\|_{L^p} &\le \sum_{i=1}^m \|p_i^v(0) - p_i^u(0)\|_{L^p}^2 + m^2 L(\epsilon)\int_0^t \|v(s) - u(s)\|_{L^p}\,\mathrm{d}s\\ &\quad + m\overline{q}(\epsilon)\int_0^t \sum_{i=1}^m \|p_i^v(s) - p_i^u(s)\|_{L^p}\,\mathrm{d}s\,.\end{aligned}$$

Then an application of Gronwall's inequality yields

$$\sum_{i=1}^m \|p_i^v(t) - p_i^u(t)\|_{L^p} \le \Big(\sum_{i=1}^m \|p_i^v(0) - p_i^u(0)\|_{L^p} + m^2 L(\epsilon)\int_0^t \|v(s) - u(s)\|_{L^p}\,\mathrm{d}s\Big)\,\mathrm{e}^{m\overline{q}(\epsilon)t}\,.$$

This proves the estimate (3.4.11) with $C = \max\{1, m^2 L(\epsilon)\}\,\mathrm{e}^{m\overline{q}(\epsilon)T}$.

In order to prove the Lipschitz-type inequality (3.4.12) we set $u(t) := v(t+h)$ and $p^u(0) := p^v(h)$ then $p^u(t) = p^v(t+h)$ and we obtain from the last estimate that

$$\sum_{i=1}^m \|p_i^v(t+h) - p_i^v(t)\|_{L^p} \le C\Big(\sum_{i=1}^m \|p_i^v(h) - p_i^v(0)\|_{L^p} + \int_0^t \|v(s+h) - v(s)\|_{L^p}\,\mathrm{d}s\Big).$$

Obviously we can estimate

$$\|p_i^v(h) - p_i^v(0)\|_{L^p} \le \int_0^h \sum_{j=1}^m \|p_j(s)q_{ji}(u(s))\|_{L^p}\,\mathrm{d}s \le m\,\overline{q}(\epsilon)|D|^{1/p}\,h,$$

and hence (3.4.12) holds with a different C as before, however, still depending only on ϵ and T.

(iv) We now consider only case (b), i.e., $v \in L^\infty((0,T),H^2)$, and the additional assumption $v \in C^0([0,T],H^1)$ is satisfied. Under these assumptions it remains to show the Lipschitz-type estimate (3.4.13). To this end we require two preliminary estimates which we derive in the parts (iv.1) and (iv.2) below. In part (iv.3) these are used to establish (3.4.13).

(iv.1) We first establish a bound on the norm

$$\|p_i^v(t)\|_{H^1} = \Big(\|p_i^v(t)\|_{L^2}^2 + \sum_{k=1}^d \|D_{x_k} p_i^v(t)\|_{L^2}^2\Big)^{1/2}.$$

The differential operator D_{x_k}, $k \le d$, is a bounded linear operator from H^2 to H^1 and we can change its application and integration in the sense of Bochner. Thus it holds for $k \le d$ that

$$D_{x_k} p_i^v(t) = D_{x_k} p_i^v(0) + \int_0^t D_{x_k}\Big(\sum_{j=1}^m p_j^v(s) q_{ji}(v(s))\Big)\,\mathrm{d}s\,.$$

Clearly, the product and chain rule hold such that

$$D_{x_k} p_j^v(t) q_{ji}(v(t)) = p_j^v(t) q_{ji}^{(1)}(v(t)) D_{x_k} v(s) + q_{ji}(v(t)) D_{x_k} p_j^v(t)$$

and thus we obtain the estimate

$$\Big\|D_{x_k}\Big(\sum_{j=1}^m p_j^v(s) q_{ji}(v(s))\Big)\Big\|_{L^2}^2 \le 2m^2 \overline{q}(\epsilon)^2 \|D_{x_k} v(s)\|_{L^2}^2 + 2m\overline{q}(\epsilon)^2 \sum_{j=1}^m \|D_{x_k} p_j^v(s)\|_{L^2}^2\,. \tag{3.4.19}$$

A summation over all $k \le d$ and $j = 1,\dots,m$ yields

$$\begin{aligned}\sum_{i=1}^m \Big(\|p_i^v(t)\|_{L^2}^2 + \sum_{k=1}^d \|D_{x_k} p_i^v(t)\|_{L^2}^2\Big) \;\le\; & \sum_{i=1}^m \|p_i^v(0)\|_{H^1}^2 + C\int_0^t \|v(s)\|_{H^1}^2\,\mathrm{d}s \\ & + C\int_0^t \sum_{i=1}^m \Big(\|p_i^v(t)\|_{L^2}^2 + \sum_{k=1}^d \|D_{x_k} p_i^v(t)\|_{L^2}^2\Big)\,\mathrm{d}s\,.\end{aligned}$$

where the constant C depends only on T and a bound to the norm $\|v\|_{L^\infty((0,T)\times D)}$. Here we used the estimates (3.4.19) and that it holds for all $j = 1,\dots,m$ that

$$\|p_i^v(t)\|_{L^2} \le \|p_i^v(0)\|_{L^2} + \overline{q}(\epsilon)\int_0^t \sum_{j=1}^m \|p_j^v(s)\|_{L^2}\,\mathrm{d}s.$$

These estimates yield the constant C and its dependencies; recall that ϵ refers to a bound on the norm $\|v\|_{L^\infty((0,T)\times D)}$. Next we apply Gronwall's Lemma and obtain

$$\sum_{i=1}^m \|p_i^v(t)\|_{H^1}^2 \leq C\Big(\sum_{i=1}^m \|p_i^v(0)\|_{H^1}^2 + \|v\|^2_{L^2((0,t),H^1)}\Big)\,\mathrm{e}^{Ct} \quad \forall\, t\leq T\,. \tag{3.4.20}$$

(iv.2) Secondly, let u denote another function with the same regularity as v. Then we estimate the norm

$$\begin{aligned}
&\Big\|D_{x_k}\sum_{j=1}^m\Big(p_j^u(t)q_{ji}(u(t)) - p_j^v(t)q_{ji}(v(t))\Big)\Big\|_{L^2}\\
&\leq\ \sum_{j=1}^m\Big(\overline{q}(\epsilon)\,\|p_j^u(t)-p_j^v(t)\|_{L^\infty}\,\|D_{x_k}v(s))\|_{L^2}\\
&\qquad + \|q_{ji}^{(1)}(u(t)) - q_{ji}^{(1)}(v(t))\|_{L^\infty}\|D_{x_k}v(s)\|_{L^2} + \overline{q}(\epsilon)\,\|D_{x_k}u(s)-D_{x_k}v(s)\|_{L^2}\\
&\qquad + \|q_{ji}(u(t)) - q_{ji}(v(t))\|_{L^\infty}\|D_{x_k}p_j^v(t)\|_{L^2} + \overline{q}(\epsilon)\|D_{x_k}p_j^u(t) - D_{x_k}p_j^v(t)\|_{L^2}\Big)\,.
\end{aligned}$$

Moreover, due to the continuous embedding of $H^2 \hookrightarrow C^0$ and the local Lipschitz continuity of q_{ji} and $q_{ji}^{(1)}$, cf. (3.4.6), there exists a constant $L(\epsilon)$ such that

$$\|q_{ji}(u(t)) - q_{ji}(v(t))\|_{L^\infty} + \|q_{ji}^{(1)}(u(t)) - q_{ji}^{(1)}(v(t))\|_{L^\infty} \leq L(\epsilon)\,\|u(t)-v(t)\|_{H^2}\,.$$

Therefore, using (3.4.11) to estimate $\|p_j^u(t) - p_j^v(t)\|_{L^\infty}$ and (3.4.20) to estimate $\|D_{x_k}p_j^u(t)\|_{L^2}$, we overall obtain that

$$\begin{aligned}
&\Big\|D_{x_k}\sum_{j=1}^m\Big(p_j^u(t)q_{ji}(u(t)) - p_j^v(t)q_{ji}(v(t))\Big)\Big\|_{L^2}\\
&\qquad\leq\ C\Big(\|p_j^u(0)-p_j^v(0)\|_{L^\infty} + \|u(t)-v(t)\|_{H^2} + \|D_{x_k}u(t) - D_{x_k}v(t)\|_{L^2}\\
&\qquad\qquad + \int_0^t \|u(s)-v(s)\|_{H^2}\,\mathrm{d}s + \sum_{j=1}^m \|D_{x_k}p_j^v(t) - D_{x_k}p_j^v(t)\|_{L^2}\Big),
\end{aligned} \tag{3.4.21}$$

where the constant C depends only on ϵ, T, $\|v\|_{C^0([0,T],H^1)}$ and $\sum_{i=1}^m \|p_i^v(0)\|_{H^1}$.

(iv.3) Now we proceed to establish the Lipschitz-type estimate (3.4.13). We apply the inequality (3.4.21) derived in (iv.2) to estimate the derivatives

$$D_{x_k}(p_i^u(t)-p_i^v(t)) = D_{x_k}(p_i^u(0)-p_i^v(0)) + \int_0^t D_{x_k}\Big(\sum_{j=1}^m p_j^u(s)q_{ji}(u(s)) - p_j^v(s)q_{ji}(v(s))\Big)\,\mathrm{d}s$$

and also (3.4.11) to obtain

$$\sum_{i=1}^{m}\Big(\|p_i^u(t)-p_i^v(t)\|_{L^2}^2+\sum_{k=1}^{d}\|D_{x_k}(p_i^u(t)-p_i^v(t))\|_{L^2}^2\Big)$$

$$\leq\; C\sum_{i=1}^{m}\Big(\|p_i^u(0)-p_i^v(0)\|_{H^1}^2+\|p_i^u(0)-p_i^v(0)\|_{L^\infty}^2\Big)+\int_0^t\|u(s)-v(s)\|_{H^2}^2\,\mathrm{d}s$$

$$+\,C\int_0^t\sum_{i=1}^{m}\Big(\|p_i^u(s)-p_i^v(s)\|_{L^2}^2+\sum_{k=1}^{d}\|D_{x_k}(p_i^u(s)-p_i^v(s))\|_{L^2}^2\Big)\,\mathrm{d}s\,,$$

where the constant C depends only on ϵ, T, $\|v\|_{C^0([0,T],H^1)}$ and $\sum_{i=1}^m\|p_i^v(0)\|_{H^1}$. Once again we apply Gronwall's Lemma and obtain for all $t\leq T$ the estimate

$$\sum_{i=1}^{m}\|p_i^u(t)-p_i^v(t)\|_{H^1}^2 \tag{3.4.22}$$

$$\leq\; C\Big(\sum_{i=1}^{m}(\|p_i^u(0)-p_i^v(0)\|_{H^1}^2+\|p_i^u(0)-p_i^v(0)\|_{L^\infty}^2)+\int_0^t\|u(s)-v(s)\|_{H^2}^2\,\mathrm{d}s\Big)\mathrm{e}^{Ct}$$

for a suitable constant C depending on ϵ, T, $\|v\|_{C^0([0,T],H^1)}$ and $\sum_{i=1}^m\|p_i^v(0)\|_{H^1}$.

To infer (3.4.13) from the estimate (3.4.22) we set $u(t):=v(t+h)$ and $p^u(0):=p^v(h)$. We estimate the terms $\|p_i^v(h)-p_i^v(0)\|_{L^2}$ and $\|p_i^v(h)-p_i^v(0)\|_{L^\infty}$ using (3.4.12). Finally, it remains only to consider the terms $\|D_{x_k}(p_i^v(h)-p_i^v(0))\|_{L^2}$. To estimate these we use (3.4.19) in combination with (3.4.20) and obtain

$$\|D_{x_k}(p_i^v(h)-p_i^v(0))\|_{L^2}^2\;\leq\;\widehat{C}h^2 \tag{3.4.23}$$

for a constant $\widehat{C}$ which depends on ϵ, T, $\|v\|_{C^0([0,T],H^1)}$ and $\sum_{i=1}^m\|p_i^v(0)\|_{H^1}$. Thus overall we find that

$$\|p_i^v(h)-p_i^v(0)\|_{H^1}^2+\|p_i^v(h)-p_i^v(0)\|_{L^\infty}^2\leq C\Big(h^2+\int_0^t\|v(s+h)-v(s)\|_{H^2}^2\,\mathrm{d}s\Big)$$

for a suitable constant C. Hence, we have proven (3.4.13).

□

(C) The active membrane equation

For a given $v\in B^\bullet(\epsilon,T)$ let p^v denote the unique solution of the gating system given by

Lemma 3.4.1. We then consider the non-autonomous initial boundary value problem

$$
\begin{aligned}
\dot{u}^v &= \Big(A - f_1(p^v(s))\Big)u^v + f_2(p^v(t)) + I(t), \\
u^v(t,x) &= 0 \qquad \forall\,(t,x) \in [0,T] \times \partial D, \\
u^v(0) &= u_0\,.
\end{aligned}
\tag{3.4.24}
$$

Due to Lemma 3.4.1 it holds that p_i^v satisfies $p_i^v \in H^1((0,T),L^2) \cap C^0([0,T],L^2)$. Moreover, $\|p^v\|_{L^\infty((0,T)\times D)} \leq 1$ yields that $\|f_1(p^v)\|_{L^\infty((0,T)\times D)}$ is bounded independently of v. Finally, $f_1(p^v) \in H^1((0,T),L^2)$ and due to the chain rule we obtain for almost all $t \in [0,T]$ that

$$
\frac{\mathrm{d}}{\mathrm{d}t} f_1(p^v(t)) = \sum_{i=1}^m \partial_{p_i} f_1(p^v(t))\, \dot{p}_i^v(t)\,,
$$

where each individual term, i.e., p^v, $\partial_{p_i} f_1$ and $\dot{p}_i^v(t)$, is almost everywhere bounded. Thus, the norm $\|\frac{\mathrm{d}}{\mathrm{d}t} f_1(p^v)\|_{L^\infty((0,T)\times D)}$ can be bounded independently of $v \in B^\bullet(\epsilon,T)$. Obviously, analogous results hold for $f_2(p^v(s)) \in H^1((0,T),L^2)$.

(a) Therefore, it follows that under the assumptions of part (a) of Theorem 3.3.1, that is, $u_0 \in H_0^1$ and $I \in L^2((0,T),L^2)$, there exists for every $v \in B^{(a)}(\epsilon,T)$ a unique solution u^v to (3.4.24) such that

$$
u^v \in L^2((0,T),H^2) \cap L^\infty((0,T),H_0^1)\,.
$$

(b) The assumptions of part (b) in Theorem 3.3.1, that is, $u_0 \in H_0^1 \cap H^2$ and $I \in H^1((0,T),L^2)$, imply that for every $v \in B^{(b)}(\epsilon,T)$ there exists a unique solution u^v to (3.4.24) such that

$$
u^v \in L^\infty((0,T),H^2)\,.
$$

Obviously the solution u depends on the choice of $v \in B^\bullet(\epsilon,T)$. Thus we can define a solution map P which assigns to every v the corresponding unique solution u^v of the active membrane equation (3.4.24), that is,

$$
P : B^\bullet(\epsilon,T) \subset L^2((0,T),L^2) \to L^2((0,T),L^2) : v \mapsto P(v) := u^v.
$$

In order to prove the existence of a solution to the coupled system of the membrane equation and the gating variables we aim to apply a fixed point theorem on the solution map P. To this end we have to prove the following properties: P maps into $B^\bullet(\epsilon,T)$ and it is a contraction thereon. The statement is presented in the following lemma.

Lemma 3.4.2. *For every $\epsilon > 0$ there exists a time $T(\epsilon) > 0$ such that for all $T < T(\epsilon)$ the solution map P maps $B^\bullet(\epsilon, T)$ into itself and it is a contraction in $L^2((0,T), L^2)$, i.e., there exists a $\gamma < 1$ such that for all $v, u \in B^\bullet(\epsilon, T)$ it holds that*

$$\|P(v) - P(u)\|_{L^2((0,T),L^2)} \leq \gamma \, \|v - u\|_{L^2((0,T),L^2)} . \tag{3.4.25}$$

Proof. The proof is split into two parts. We show (i) that for small enough T the solution map P maps $B^\bullet(\epsilon, T)$ into itself and (ii) that it is a contraction.

(i) Firstly, we want to find a $T(\epsilon)$ such that P maps $B^\bullet(\epsilon, T)$ into itself for all $T < T(\epsilon)$. To this end we have to consider the difference of a solution $P(v)$ of (3.4.24) for a given $v \in B^\bullet(\epsilon, T)$ to the solution w of the passive membrane equation (3.4.7). Define $z := P(v) - w$ and observe that z solves the initial-boundary value problem

$$\begin{aligned} \dot{z} &= \Big(A - f_1(p^v(s))\Big) z - f_1\big(p^v(s)\big) w(s) + f_2\big(p^v(s)\big), \\ z(t,x) &= 0 \qquad \forall\, (t,x) \in [0,T] \times \partial D, \\ z(0) &= 0 . \end{aligned}$$

Thus z satisfies due to Theorem C.0.3 in case (a) the estimate

$$\|P(v) - w\|_{L^\infty([0,T],H^1)} \leq C_1 \left\| f_1(p^v) w + f_2(p^v) \right\|_{L^2((0,T),L^2)} ,$$

or in case (b) the estimate

$$\|P(v) - w\|_{L^\infty([0,T],H^2)} \leq C_1 \left\| f_1(p^v) w + f_2(p^v) \right\|_{H^1((0,T),L^2)}$$

for an appropriate constant $C_1 < \infty$ depending only on T, the domain D, the coefficients of the operator A and the $L^\infty((0,T) \times D)$–norm of $f_1(p^v(t))$ and its first derivative with respect to t, cf. Remark C.0.1. This yields that the constant C_1 can be chosen independently of $v \in B^\bullet(\epsilon, T)$. We next consider the bounds in the above right hand sides which depend on v and derive estimates independent of $v \in B^\bullet(\epsilon, T)$. Note that the dependence on w is irrelevant as w is fixed for a fixed initial condition u_0 and current input I.

(i.1) We start considering the norm in $L^2((0,T), L^2)$. Then the pointwise bounds almost everywhere yield the estimate

$$\int_0^T \|f_1(p^v(s)) w(s)\|_{L^2}^2 \, \mathrm{d}t \leq |D| \int_0^T \|f_1(p^v(s))\|_{L^\infty}^2 \|w(s)\|_{L^\infty}^2 \, \mathrm{d}t \leq \overline{q}^2 K^2 |D| \, T ,$$

where $K < \infty$ is the pointwise bound on w, cf. part (A). Analogously we obtain $\|f_2(p^v)\|^2_{L^2((0,T),L^2)} \leq \overline{q}^2|D|\,T$.

(i.2) Next, we consider the norm in $H^1((0,T),L^2)$. First of all note that w, $f_1(p^v)$ and $f_2(p^v)$ are in $H^1((0,T),L^2)$. Furthermore for the product $f_1(p^v)w$ the chain rule applies as both factors as well as the first temporal derivative of $f_1(p^v)$ are almost everywhere bounded and thus $f_1(p^v)w$ and $w\frac{\mathrm{d}}{\mathrm{d}t}f_1(p^v) + f_1(p^v)\dot{w}$ are in $L^2((0,T),L^2)$. Furthermore, the regularity of w implies that $\dot{w}$ is bounded in $L^2((0,T),L^2)$. Hence, in order to estimate the norm of $f_1(p^v)w$ in $H^1((0,T),L^2)$ it remains to estimate $\|\frac{\mathrm{d}}{\mathrm{d}t}f_1(p^v)\|_{L^2((0,T),L^2)}$. For this term the chain rule gives, in the case of the full gating system,

$$\tfrac{\mathrm{d}}{\mathrm{d}t}f_1(p^v(t)) = \sum_{i=1}^m \partial_{p_i}f_1(p^v(s))\sum_{j=1}^m p_j^v(s)q_{ji}(v(s))\,.$$

Here each individual term, i.e., $\partial_{p_i}f_1$, q_{ij} and p_i^v, is almost everywhere bounded by assumption or the preceding considerations. Particularly, the bounds are independent of $v \in B^\bullet(\epsilon,T)$. Therefore the derivative $\frac{\mathrm{d}}{\mathrm{d}t}f_1(p^v(t))$ is bounded almost everywhere independently of $v \in B^\bullet(\epsilon,T)$. Analogously we obtain a bound for $\frac{\mathrm{d}}{\mathrm{d}t}f_2(p^v)$ independently of $v \in B^\bullet(\epsilon,T)$. Obviously, analogous reasoning yields the same result for the reduced gating system.

Hence, we infer that there exists a constant $C(\epsilon,T) < \infty$ depending on ϵ and T such that

$$\text{(a)}\quad \|P(v) - w\|_{L^\infty((0,T),H_0^1)} \;\leq\; C(\epsilon,T)\sqrt{T}\,,$$

$$\text{(b)}\quad \|P(v) - w\|_{L^\infty((0,T),H^2)} \;\leq\; C(\epsilon,T)\sqrt{T}\,.$$

In each case we set $T(\epsilon) := (\epsilon/C(\epsilon,T))^2$ which yields that for all $T^* \leq T(\epsilon)$ the solution $P(v)$ is in $B^\bullet(\epsilon,T^*)$.

(ii) We proceed to prove that the solution map forms a contraction. Let $T < T(\epsilon)$ as derived in (i) and v,u be in $B^\bullet(\epsilon,T)$ which implies that $P(v)$, $P(u) \in B^\bullet(\epsilon,T)$. Next, define $z := P(v) - P(u)$ and observe that z is the solution to the initial-boundary value problem

$$\begin{aligned} \dot{z} &= \Big(A - f_1(p^v(t))\Big)\,z + F(t), \\ z(t,x) &= 0 \qquad \forall\,(t,x) \in [0,T]\times\partial D, \\ z(0) &= 0\,, \end{aligned}$$

where the inhomogeneous term F is given by

$$F(t) := \big(f_1(p^u(t)) - f_1(p^v(t))\big)\,P(u)(t) + f_2(p^v(t)) - f_2(p^u(t)). \tag{3.4.26}$$

Due to preceding regularity results we infer that $F \in L^2((0,T),L^2)$. Thus taking the duality pairing of $\dot{z}(t)$ with $z(t)$ in H_0^1 we obtain

$$\langle \dot{z}(t), z(t)\rangle_{H^1} = \langle Az(t), z(t)\rangle_{H^1} - (f^1(v)z(t), z(t))_{L^2} + (F(t), z(t))_{L^2}\,.$$

We next estimate this equality considering each term in the right hand side separately. To estimate the first we employ the coercivity of the linear operator A, i.e., there exist ξ_1, $\xi_2 > 0$ such that $-\langle Az(t), z(t)\rangle_{H^1} \geq \xi_1 \|u\|_{H^1}^2 - \xi_2 \|z(t)\|_{L^2}^2$. For the second term we employ the fact that $(f_1(v)z(t), z(t))_{L^2}$ is non-negative as f_1 is non-negative. Finally the third terms is estimated using the Cauchy-Schwarz in combination with Young's inequality. Then we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\|z(t)\|_{L^2}^2 \leq (2\xi_2 + 1)\,\|z(t)\|_{L^2}^2 + \|F(t)\|_{L^2}^2\,.$$

Integrating both sides of this inequality over $[0,t]$ and setting $\widehat{C} := 2\xi_2 + 1$ yields

$$\|z(t)\|_{L^2}^2 \leq \widehat{C}\int_0^t \|z(s)\|_{L^2}^2\,\mathrm{d}s + \int_0^t \|F(s)\|_{L^2}^2\,\mathrm{d}s\,.$$

We apply Gronwall's Lemma to this inequality and obtain for all $t \leq T$

$$\|z(t)\|_{L^2}^2 \leq \mathrm{e}^{\widehat{C}t}\,\|F\|_{L^2((0,T),L^2)}^2\,.$$

Again, integrating both sides of this inequality over $[0,T]$ we obtain

$$\|z\|_{L^2((0,T),L^2)}^2 \leq \widehat{C}^{-1}(\mathrm{e}^{CT} - 1)\,\|F\|_{L^2((0,T),L^2)}^2\,. \tag{3.4.27}$$

Finally, we estimate the norm of the inhomogeneous term F by

$$\begin{aligned}
\|F\|_{L^2((0,T),L^2)}^2 &= \int_0^T \big\|P(u)(t)\big(f_1(p^u(t)) - f_1(p^v(t))\big) + f_2(p^v(t)) - f_2(p^u(t))\big\|_{L^2}^2\,\mathrm{d}t\\
&\leq 2\|P(u)\|_{L^\infty((0,T)\times D)}^2 \int_0^T \|f_1(p^v(t)) - f_1(p^u(t))\|_{L^2}^2\,\mathrm{d}t\\
&\quad + 2\int_0^T \|f_2(p^v(t)) - f_2(p^u(t))\|_{L^2}^2\,\mathrm{d}t\,.
\end{aligned}$$

Due to the local Lipschitz continuity of the functions f_1, f_2 and the estimate (3.4.11) – note that we have the same initial conditions for the two solutions of the gating system – we obtain from this last inequality that

$$\begin{aligned} \|F\|^2_{L^2((0,T),L^2)} &\leq 4\big(LTC\big)^2 \left(1+\|P(u)\|^2_{L^\infty((0,T)\times D)}\right) \int_0^T \|v(s)-w(s)\|^2_{L_2}\,\mathrm{d}s \\ &\leq 4\big(LTC\big)^2 (1+K+K_1\epsilon)^2 \|v-w\|^2_{L^2((0,T),L^2)}\,, \end{aligned}$$

where the constant C depends only on ϵ and T. Here the second inequality is valid as $K+K_1\epsilon$ is a pointwise bound almost everywhere for all elements $v \in B^\bullet(\epsilon,T)$, cf. part (A). A combination of this last estimate with (3.4.27) yields

$$\|P(v)-P(u)\|^2_{L^2((0,T),L^2)} \leq 4\big(LTC\big)^2 (1+K+K_1\epsilon)^2\, \widehat{C}^{-1}\, (\mathrm{e}^{\widehat{C}T}-1)\, \|v-u\|^2_{L^2((0,T),L^2)}.$$

Thus contractivity follows for the map P for $T(\epsilon)$ sufficiently small such that

$$\gamma^2 := 4\big(LTC\big)^2 (1+K+K_1\epsilon)^2 \widehat{C}^{-1} (\mathrm{e}^{\widehat{C}T}-1) < 1 \qquad \forall\, T < T(\epsilon)\,.$$

We note that it is guaranteed that $\gamma^2 \to 0$ for $T \to 0$ as the specific structure of the constant C arising from an application of (3.4.11) is decreasing for $T \to 0$.

□

(D) Local and global existence of solutions

To finish the proof we first present a local existence result, which is afterwards extended to a global result.

Lemma 3.4.3. (a) *Let $d=1$, then for every set of initial conditions $u_0 \in H_0^1$ and $p_i(0) \in H^1$, $i=1,\dots,m$, and any input current $I \in L^2((0,T),L^2)$, there exists a $T>0$ such that the generalised excitable media system* (3.3.5)*, equipped with Dirichlet boundary conditions and coupled to* (3.3.7) *or* (3.3.6)*, possesses a unique weak solution (u,p) on $(0,T)$ satisfying*

$$u \in L^2((0,T),H^2) \cap H^1((0,T),L^2) \subset C^0([0,T],H_0^1),$$

$$p_i \in C^1([0,T],H^1) \quad \forall\, i=1,\dots,m\,.$$

(b) *Let $d \leq 3$, then for every set of initial conditions $u_0 \in H_0^1 \cap H^2$ and $p_i(0) \in H^2$, $i=1,\dots,m$, and any input current $I \in H^1((0,T),L^2) \cap L^2((0,T),H_0^1)$, there exists a $T>0$ such that the generalised excitable media system* (3.3.5)*, equipped with Dirichlet boundary conditions and coupled to* (3.3.7) *or* (3.3.6)*, possesses a* *unique weak solution* (u,p) *on* $(0,T)$ *satisfying in addition to* (a)

$$u \in L^2((0,T),H^3) \cap H^1((0,T),H^1) \subset C^0([0,T],H^2) \cap C^1([0,T],L^2),$$

$$p_i \in C^1([0,T],H^2) \quad \forall\, i = 1,\dots,m\,.$$

Proof. We have split the proof into two parts. First, (i) the Banach Fixed Point Theorem is used to establish existence of a solution, then in (ii) uniqueness of this solution is shown.

(i) Due to Lemma 3.4.2 and under the conditions of the above lemma the solution map $v \mapsto P(v)$ is a contraction on the complete metric space $B^\bullet(\epsilon,T)$ for small enough $T>0$. Hence, due to the Banach Fixed Point Theorem there exists a unique fixed point $u \in B^\bullet(\epsilon,T)$ of the solution map, i.e., $u = P(u)$. Thus, u solves the initial-boundary value problem

$$\begin{aligned} \dot u &= Au - f_1(p^u(t)u + f_2(p^u(t)) + I(t), \\ u(t,x) &= 0 \qquad \forall\,(t,x) \in [0,T]\times\partial D, \\ u(0) &= u_0\,. \end{aligned}$$

Here (a) $p^u \in H^1((0,T),H^1)$ or (b) $p^u \in H^1((0,T),H^2)$, respectively, is the unique solution to the gating system such that u is the input function out of $B^\bullet(\epsilon,T)$ in the sense of Lemma 3.4.1. Hence (u,p^u) is a solution to the general excitable media system (3.3.5), equipped with Dirichlet boundary conditions and coupled to (3.3.7) or (3.3.6), satisfying, in particular,

$$\text{(a)}\quad u \in C^0([0,T],H^1)\,, \qquad \text{(b)}\quad u \in C^0([0,T],H^2)\,. \tag{3.4.28}$$

Statement (a) in (3.4.28) follows immediately from Theorem C.0.3 (a) and Remark C.0.1. In order to employ these results we reason in the following way: We have already established that u is a weak solution to (3.3.5), almost everywhere bounded and satisfying certain regularity conditions weaker than (3.4.28). To proceed we interpret u as a solution to a linear parabolic PDE where the zeroth order term $\widetilde{c}$ of the linear operator is given by $\widetilde{c}(t) := c + f_1(p^u(t))$ and the inhomogeneous term by $f_2(p^u(t)) + I(t)$. To this linear PDE we can now apply the results in Appendix C. Similarly as (a) we infer (b) in (3.4.28) as stated in Remark C.0.1. To this end we have to check for the conditions that arise on $f_1(p^u)$ and $f_1(p^u)$: Firstly, note that $p^u \in L^\infty((0,T),H^2)$ implies due to the chain rule that $\frac{\mathrm{d}}{\mathrm{d}t}f_1(p^u) \in L^4((0,T),L^4)$ and $f_2(p^u) \in H^1((0,T),L^2)$. Secondly, the chain rule and the fact that $p^u \in C^0([0,T],H^2)$ further allows to infer that $f_1(p^u) \in L^\infty((0,T),H^1) \cap L^2((0,T),H^2)$. Thirdly, analo-

gously we obtain $f_2(p^u) \in L^2((0,T),H^2) \cap H^1((0,T),L^2)$. This completes the set of conditions necessary to infer (b) in (3.4.28). Overall, we have obtained the proposed regularity of the solution component u as proposed in Lemma 3.4.3.
Finally note that due to Lemma 3.4.1 the regularity (3.4.28) of u implies the improved regularity of the gating variables, i.e., (a) $p_i^u \in C^1([0,T],H^1)$ or (b) $C^1([0,T],H^2)$, respectively. Hence, the existence part of Lemma 3.4.3 is proved.

(ii) It remains to show uniqueness of the solution u in $L^2((0,T),L^2)$. This is accomplished using the same argument as in [81] which we repeat for the sake of completeness. Assume that $\widetilde{u}$ is another solution, then it has to hold $|\widetilde{u}(t,x)-w(t,x)| \leq K+K_1\epsilon$ for small enough t. Further assume that t^* is the largest time $t \leq T$ such that $\widetilde{u}$ is in the set $B^\bullet(\epsilon,t)$. Due to the uniqueness of the fixed point in the Banach Fixed Point Theorem it follows that $\widetilde{u} = u$ in $L^2((0,t^*),L^2)$. Then, on the one hand, if $t^* = T$ we have obtained uniqueness of the solution on $[0,T]$. On the other hand if $t^* < T$ we apply the same argument again for the local solution started at the new initial point $u_0^* = \widetilde{u}(t^*) = u(t^*)$ and obtain a contradiction to the maximality of t^*.

□

The proof of well-posedness of the general excitable media system is completed by the following result extending local existence and uniqueness of a solution to a global existence result.

Theorem 3.4.1. (a) *Let $d = 1$, then for every $T > 0$, every set of initial conditions $u_0 \in H_0^1$, $p_i(0) \in H^1$, $i = 1,\dots,m$, and any input current $I \in L^2((0,T),L^2)$, the generalised excitable media system* (3.3.5)*, equipped with Dirichlet boundary conditions and coupled to* (3.3.7) *or* (3.3.6)*, possesses a unique weak solution (u,p) on $(0,T)$ satisfying*

$$u \in L^2((0,T),H^2) \cap H^1((0,T),L^2) \subset C^0([0,T],H_0^1)$$

$$p_i \in C^1([0,T],H^1) \quad \forall i = 1,\dots,m.$$

(b) *Let $d \leq 3$, then for every $T > 0$, every set of initial conditions $u_0 \in H_0^1 \cap H^2$, $p_i(0) \in H^2$, $i = 1,\dots,m$, and any input current $I \in H^1((0,T),L^2) \cap L^2((0,T),H_0^1)$, the generalised excitable system,* (3.3.5)*, equipped with Dirichlet boundary conditions and coupled to* (3.3.7) *or* (3.3.6)*, possesses a unique weak solution (u,p)on $(0,T)$ satisfying in addition to* (a)

$$u \in L^2((0,T),H^3) \cap H^1((0,T),H^1) \subset C^0([0,T],H^2) \cap C^1([0,T],L^2)$$

$$p_i \in C^1([0,T],H^2) \quad \forall i = 1,\dots,m.$$

Moreover, in both cases the solution depends continuously on the initial data u_0, p_0 and the current input I.

Proof. The proof is again split into two parts. Firstly, (i) we prove the existence and uniqueness of a global solution and secondly, (ii) we show the continuous dependence on the initial data u_0 and I. We note that throughout the subsequent proof we regularly use the same symbol C to denote different finite constants. The dependence of constants C is always explained in the context.

(i) In order to prove global existence of a solution, i.e., existence for for ever $T > 0$, we employ the local existence result in Lemma 3.4.3 and piece together local solutions. That is we use the endpoint of the first solution as the new initial value of the following and thus obtain existence on a larger interval. The difficulty here is that the time T^* for which a local solution exists depends on the initial condition. Hence, with varying initial conditions it is possible that the times of existence of subsequent local solution pieced together become smaller and smaller and sum up approaching a finite value $T_0 < \infty$. Then, a global solution does not exist. In the following we prove by contradiction that this scenario cannot be the case motivated by the method of proof employed in [81].

Let T_0 be the maximal time in $[0, \infty]$ such that a solution (u, p) to the excitable media system exists for all $T \in [0, T_0)$ and assume $T_0 < \infty$. Then we define

$$\begin{aligned} \|u\|_{C^0([0,T_0],L^2)} &:= \lim_{T \to T_0} \|u\|_{C^0([0,T],L^2)}\,, \\ \|u\|_{C^0([0,T_0],H^1)} &:= \lim_{T \to T_0} \|u\|_{C^0([0,T],H^1)}\,. \end{aligned} \tag{3.4.29}$$

It holds that these two limits are finite for $T_0 < \infty$ due to the estimates in Theorem C.0.3 (a) and Remark C.0.1, cf. the explanation of the applicability of these results in the present situation following (3.4.28). In case $d = 1$, i.e., under the assumptions of part (a), the finiteness of the limits (3.4.29) now implies that there exists a constant $\kappa < \infty$ such that it holds $|u(t, x)| \le \kappa$ almost everywhere on $(0, T) \times D$ for all $T < T_0$.

Under the assumptions in part (b) we define the limits

$$\begin{aligned} \|u\|_{L^\infty([0,T_0],H^2)} &:= \lim_{T \to T_0} \|u\|_{L^\infty([0,T],H^2)}\,, \\ \|\dot u\|_{C^0([0,T_0],L^2)} &:= \lim_{T \to T_0} \|\dot u\|_{C^0([0,T],L^2)}\,. \end{aligned} \tag{3.4.30}$$

Again these limits are finite due to Theorem C.0.3(b) and Remark C.0.1 as the norms $\|\frac{\mathrm{d}}{\mathrm{d}t} f_1(p)\|_{L^4((0,T),L^4)}$ and $\|f_2(p)\|_{H^1((0,T),L^2)}$ can be bounded independently of $T < T_0$: We start showing the former. Considering, for example, the full gating system the chain rule yields the derivatives

$$\tfrac{\mathrm{d}}{\mathrm{d}t} f_1(p(t)) = \sum_{i=1}^{m} \partial_{p_i} f_1(p(t)) \sum_{j=1}^{m} p_j(t) q_{ji}(u(t))\,.$$

The terms $\partial_{p_i} f_1(p(t))$ and $p_j(t)$ are almost everywhere bounded independently of $T < T_0$. Thus in order to estimate the norm $\|\tfrac{\mathrm{d}}{\mathrm{d}t} f_1(p)\|_{L^4((0,T),L^4)}$ it suffices to estimate the norm $\|q_{ji}(u)\|_{L^4((0,T),L^4)}$. Due to the polynomial growth condition (EC3) on q_{ji} we obtain for a $\gamma \in \mathbb{N}$

$$\|q_{ji}(u)\|^4_{L^4((0,T),L^4)} \le C \int_0^T \int_D (1 + |u(t,x)|^\gamma)^4 \,\mathrm{d}x\,\mathrm{d}t \le 6CT + 6C \sum_{l=1}^{4} \int_0^T \|u(t)\|^{l\gamma}_{L^{l\gamma}}\,\mathrm{d}t$$

for a suitable constant C. The Sobolev Embedding Theorem states that $H^2 \hookrightarrow L^p$ for all $p \le \infty$ and thus , in order to bound the above norm, it suffices to derive estimates of the form

$$\|u\|_{L^{l\gamma}((0,T),H^2)} \le \widetilde{M} \qquad \forall\, T < T_0,\ l = 1,\dots,4 \tag{3.4.31}$$

for some finite constant $\widetilde{M}$. An obvious modification of the estimation procedure in the proof of [43, Chap. 7,Thm. 5(a)][7] shows that

$$\int_0^T \|u(t)\|^{l\gamma}_{H^2}\,\mathrm{d}t \le C\Big(\int_0^T \|f_2(p(t)) + I(t)\|^{l\gamma}_{L^2}\,\mathrm{d}t + \|u_0\|^{l\gamma}_{H^1}\Big)\,,$$

where the constant C depends only on the coefficients of the linear operator A. Now the statement (3.4.31) follows as, on the one hand, the term $f_2(p(t))$ is almost everywhere bounded independently of $t < T_0$ and, on the other hand, $I \in H^1((0,T),L^2)$ implies $I \in C^0([0,T],L^2) \subset L^{l\gamma}((0,T),L^2)$. Overall, we have obtained that $\|q_{ij}(u)\|_{L^4((0,T),L^4)}$ is bounded independently of $T < T_0$. This in turn implies that $\|\tfrac{\mathrm{d}}{\mathrm{d}t} f_1(p)\|_{L^4((0,T),L^4)}$ is bounded independently of $T < T_0$.
Analogously we obtain that $\|\tfrac{\mathrm{d}}{\mathrm{d}t} f_2(p)\|_{L^2((0,T),L^2)}$ is bounded independently of $T < T_0$ and hence we infer that $\|f_2(p)\|_{H^1((0,T),L^2)}$ is bounded independently of $T < T_0$.
Finally, we remark that completely analogous considerations apply to the reduced gating system. Overall, these considerations show that the limits (3.4.30) exist.

Finally, the finiteness of $\|u\|_{L^\infty([0,T_0],H^2)}$ implies that also in case $d \le 3$ and under the assumptions of (b) there exists a constant $\kappa < \infty$ such that $|u(t,x)| \le \kappa$ holds almost everywhere on $(0,T) \times D$ for all $T < T_0$. As the pointwise bound κ depends only on

[7] This result in [43] proves, in the notation of the present situation, that

$$\int_0^T \|u(t)\|^2_{H^2}\,\mathrm{d}t \le C\Big(\int_0^T \|f_2(p(t)) + I(t)\|^2_{L^2}\,\mathrm{d}t + \|u_0\|^2_{H^1}\Big)\,.$$

This should give an idea what kind of modifications in the estimation procedure are needed.

T_0 we subsequently subsume a dependence on κ and T_0 as a dependence on T_0 only.

We emphasise that the 'norms' $\|\cdot\|_{C^0([0,T_0],\bullet)}$ in (3.4.29) and (3.4.30) are just a notation for the limit of the respective norms for $T \uparrow T_0$ because a priori the existence of $u(T_0)$ is not given. The latter is precisely what we prove in the following. That is, we prove that the limit

$$u(T_0) := \lim_{t\to T_0} u(t)$$

exist and $u(T_0) \in X$ where X denotes either (a) H_0^1 or (b) H^2. Given the existence of $u(T_0)$, which implies that $u \in C^0([0,T_0],X)$, we immediately obtain that the limits $p_j(T_0)$ of the gating variables exist in (a) H^1 and (b) H^2, respectively, due to Lemma (3.4.1). Hence we infer that there exists a solution to the excitable media system on the interval $[0,T_0]$. Then, due to the local existence result starting at time T_0 with the initial condition $(u(T_0), p(T_0))$ we obtain that there exist a solution on $[0, T_0+T^*)$ for some $T^* > 0$. This is a contradiction to T_0 being the maximal finite time such that a solution exist. Hence a global solution exists.

Therefore it remains to prove in the remainder of part (i) that the limit $u(T_0)$ exists. To this end it suffices to prove that the function u is uniformly continuous, that is,

$$\forall\,\epsilon > 0\ \exists\,\delta > 0\ :\ \|u(t+h) - u(t)\|_X < \epsilon \quad \forall h < \delta,\ t, t+h < T_0\,, \tag{3.4.32}$$

as every uniformly continuous function on a metric space has a unique continuous extension to the completion of its domain of definition.

In the following we use $p^u := p$ to denote the gating component of the local solution (u,p) to the excitable media system to emphasise the dependence on the component u. We set $v(t) := u(t+h)$, then $v(t)$ solves the initial-boundary value problem

$$\begin{aligned}
\dot v &= \Big(A - f_1(p^u(t+h))\Big)v + f_2(p^u(t+h)) + I(t+h),\\
v(t,x) &= 0 \qquad \forall\,(t,x) \in [0,T]\times\partial D,\\
v(0) &= u(h)\,.
\end{aligned}$$

Here we can identify $p^v(t) = p^u(t+h)$ with $p^v(0) = p^u(h)$. Next we define $u_h(t) := v(t) - u(t)$ and observe that u_h solves the initial-boundary value problem

$$\begin{aligned}
\dot u_h &= \Big(A - f_1(p^u(t+h))\Big)u_h + F_h(t),\\
u_h(t,x) &= 0 \qquad \forall\,(t,x) \in [0,T]\times\partial D,\\
u_h(0) &= u(h) - u_0\,,
\end{aligned} \tag{3.4.33}$$

where

$$F_h(t) := -\Big(f_1(p^u(t+h))-f_1(p^u(t))\Big)u(t)+\Big(f_2(p^u(t+h))-f_2(p^u(t))\Big)+\Big(I(t+h)-I(t)\Big)\,.$$

Due to Theorem C.0.3 and Remark C.0.1 the solution u_h satisfies for all $t \leq T$ such that $T + h < T_0$ the estimate (a)

$$\|u(t+h) - u(t)\|^2_{H^1} \;\leq\; C_{\mathrm{a}}\Big(\|u(h) - u_0\|^2_{H^1} + \|F_h(t)\|^2_{L^2((0,T),L^2)}\Big), \tag{3.4.34}$$

and (b)

$$\|u(t+h) - u(t)\|^2_{H^2} \;\leq\; C_{\mathrm{b}}\Big(\|u(h) - u_0\|^2_{H^2} + \|F_h(t)\|^2_{H^1((0,T),L^2)} + \|F_h(t)\|^2_{L^2((0,T),H^1)}\Big). \tag{3.4.35}$$

Most importantly the constants C_{a}, C_{b} can be chosen independently of $T \in [0, T_0)$: Initially, as provided by Theorem C.0.3 (a) a constant C_{a} satisfying (3.4.34) depends on T, D, the operator A and $\|f_1(p^u(h+\cdot))\|_{L^\infty((0,T)\times D)}$. However, this last norm can be bounded independently of $T \in [0, T_0)$. Therefore the constant C_{a} can be chosen independently of $T \in [0, T_0)$.
In addition to the dependencies of C_{a} the constant C_{b} in (3.4.35) depends also on $\|\frac{\mathrm{d}}{\mathrm{d}t} f_1(p^u(h+\cdot))\|_{L^\infty((0,T)\times D)}$ and $\|f_1(p^u(h+\cdot))\|_{L^\infty((0,T),H^1)}$. As the solution u is almost everywhere bounded by κ independently of $T \in [0, T_0)$ it follows that the first of these norms is bounded almost everywhere independently of $T \in [0, T_0)$. That also the second can be bounded independently of $T \in [0, T_0)$ holds due to (3.4.20) as $\|u\|_{L^\infty((0,T)\times D)}$ and $\|u\|_{C^0([0,T],H^1)}$ are bounded independently of $T \in [0, T_0)$ due to the existence of the limits (3.4.29) and (3.4.30).

The next step in the proof is to derive estimates for the norms of the term F_h in the right hand sides of (3.4.34) and (3.4.35) such that $F_h \to 0$ for $h \to 0$. To this end we split $F_h(t) = f_h(t) + \big(I(t+h) - I(t)\big)$ where

$$f_h(t) \;:=\; -\Big(f_1(p^v(t)) - f_1(p^u(t))\Big)\,u(t) + \Big(f_2(p^v(t)) - f_2(p^u(t))\Big)$$

and consider the two terms separately starting with f_h. That is we first estimate (i.1) the norm $\|f_h\|_{L^2((0,T),L^2)}$ and in the case of (b) also the norms (i.2) $\|f_h\|_{H^1((0,T),L^2)}$ and (i.3) $\|f_h\|_{L^2((0,T),H^1)}$. Then in (i.4) the same norms are considered for the term $I(t+h) - I(t)$. Finally, in part (i.5) these estimates are combined and applied to (3.4.34) and (3.4.35) and we infer uniform continuity of u.

(i.1) Note that the term $f_h(t)$ has the same structure as the term F, see (3.4.26), in the proof of Lemma 3.4.2. Hence by analogous considerations and using the inequality

(3.4.11) we obtain the estimate

$$\|f_h\|^2_{L^2((0,T),L^2)} \le 4L^2(m+T^2)\,(1+\kappa)^2\,C^2\Big(\sum_{i=1}^m \|p_i^v(0)-p_i^u(0)\|^2_{L^2}+\int_0^T \|v(s)-u(s)\|^2_{L^2}\,\mathrm{d}s\Big),$$

where the constant C depends only on κ and T_0. As $p^v(0)=p^u(h)$, an application of the Lipschitz-type estimate (3.4.12) yields overall the estimate

$$\|f_h\|^2_{L^2((0,T),L^2)} \le C\left(h^2+\int_0^T \|v(s)-u(s)\|^2_{L^2}\,\mathrm{d}s\right), \tag{3.4.36}$$

where the constant $C<\infty$ can be chosen independently of $T\in[0,T_0)$.

(i.2) We next estimate the norm $\|\frac{\mathrm{d}}{\mathrm{d}t}f_h\|_{L^2((0,T),L^2)}$. In earlier considerations we have already established that under the additional assumptions of (b) the chain and product rule applies. Thus it holds for almost all $t\in(0,T_0)$ that

$$\begin{aligned}\|\tfrac{\mathrm{d}}{\mathrm{d}t}f_h(t)\|^2_{L^2} &= (2m+1)\Big(\big\|\big(f_1(p^v(t))-f_1(p^u(t))\big)\,\dot u(t)\big\|^2_{L^2}\\ &\quad+\sum_{i=1}^m \big\|\big(\dot p_i^v(t)\,\partial_{p_i}f_1(p^v(t))-\dot p_i^u(t)\partial_{p_i}f_1(p^u(t))\big)\,u(t)\big\|^2_{L^2} \\ &\quad+\sum_{i=1}^m \big\|\dot p_i^v(t)\,\partial_{p_i}f_2(p^v(t))-\dot p_i^u(t)\,\partial_{p_i}f_2(p^u(t))\big\|^2_{L^2}\Big).\end{aligned} \tag{3.4.37}$$

We estimate each term in the right hand side if (3.4.37) separately. Firstly, due to the local Lipschitz condition on f_1 we obtain

$$\begin{aligned}\big\|\big(f_1(p^v(t))-f_1(p^u(t))\big)\,\dot u(t)\big\|^2_{L^2} &\le \|f_1(p^v(t))-f_1(p^u(t))\|^2_{L^\infty}\|\dot u(t)\|^2_{L^2}\\ &\le \overline{q}(\kappa)^2\,\|\dot u\|^2_{C^0([0,T_0],L^2)}\sum_{j=1}^m \|p_j^v(t)-p_j^u(t)\|^2_{L^\infty},\end{aligned}$$

where we have employed that $\|\dot u(t)\|^2_{L^2}\le\|\dot u\|_{C^0([0,T_0],L^2)}$ for all $t<T_0$. Furthermore, as $v(t)=u(t+h)$ we obtain due to the Lipschitz-type inequality (3.4.12)

$$\sum_{j=1}^m \|p_j^v(t)-p_j^u(t)\|^2_{L^\infty}\le C\Big(h^2+\int_0^t \|v(t)-u(t)\|^2_{L^\infty}\,\mathrm{d}s\Big),$$

where the constant C depends on κ and T_0. Finally, using the continuous embedding $H^2\hookrightarrow C^0$ yields overall the estimate

$$\big\|(f_1(p^v(t))-f_1(p^u(t)))\dot u(t)\big\|^2_{L^2}\le C\Big(h^2+\int_0^t \|v(t)-u(t)\|^2_{H^2}\,\mathrm{d}s\Big) \tag{3.4.38}$$

for a suitable constant $C < \infty$ independent of $T \in [0, T_0)$.

Next we extend the individual terms in the sums in the second and third row in the right hand side of (3.4.37) by $\pm(\dot{p}_i^u\, \partial_{p_i} f_1(p^v))u$ and $\pm\dot{p}_i^u\, \partial_{p_i} f_2(p^v)$, respectively, and estimate analogously as above. That is, we use the fact that $|u(t,x)| \leq \kappa$ almost everywhere, the local Lipschitz and boundedness assumptions on the functions f_1, f_2, their first derivatives and the rate functions q_{ij} and the Lipschitz-type estimate (3.4.12) to obtain

$$\big\|\big(\dot{p}_i^v(t)\, \partial_p f_1(p^v(t)) - \dot{p}_i^u(t)\partial_{p_i} f_1(p^u(t))\big)\, u(t)\big\|_{L^2}^2$$

$$\leq \; C\Big(h^2 + \|v-u\|_{L^2}^2 + \int_0^t \|v(s)-u(s)\|_{L^2}^2\, \mathrm{d}s\Big)\,, \quad (3.4.39)$$

and

$$\|\dot{p}_i^v(t)\, \partial_{p_i} f_2(p^v(t)) - \dot{p}_i^u(t)\, \partial_{p_i} f_2(p^u(t))\|_{L^2}^2$$

$$\leq \; C\Big(h^2 + \|v-u\|_{L^2}^2 + \int_0^t \|v(s)-u(s)\|_{L^2}^2\, \mathrm{d}s\Big) \quad (3.4.40)$$

for suitable constants $C < \infty$ independent of $T \in [0, T_0)$.

Then, a combination of the inequalities (3.4.37) – (3.4.40) and employing the continuous embedding $H^2 \hookrightarrow L^2$ yields that overall it holds

$$\|\tfrac{\mathrm{d}}{\mathrm{d}t} f_h(t)\|_{L^2}^2 \;\leq\; C\Big(h^2 + \|v(t)-u(t)\|_{H^2}^2 + \int_0^t \|v(t)-u(t)\|_{H^2}^2\, \mathrm{d}s\Big)$$

for a suitable constant $C < \infty$ independent of $T \in [0, T_0)$. Integrating both sides of this inequality over $(0,T)$ we obtain the estimate

$$\|\tfrac{\mathrm{d}}{\mathrm{d}t} f_h\|_{L^2((0,T),L^2)}^2 \leq C\,\Big(h^2 + \int_0^T \|v(t)-u(t)\|_{H^2}^2\, \mathrm{d}s\Big) \quad (3.4.41)$$

for a suitable constant $C < \infty$ independent of $T \in [0, T_0)$. We emphasise at this point that the constant C, however, does depend on T_0.

(i.3) Finally, we estimate $\|f_h\|_{L^2((0,T),H^1)}$. As we have already estimated the norm $\|f_h\|_{L^2((0,T),L^2)}$ in part (i.1) it remains to consider the norms $\|D_{x_i} f_h\|_{L^2((0,T),L^2)}$ for

$i \leq d$. For almost all t we get the sequence of estimates

$$\begin{aligned}
&\big\| D_{x_i}\big(f_2(p^v(t)) - f_2(p^u(t))\big)\big\|_{L^2}^2 \\
&\leq\ 2m \sum_{j=1}^m \Big(\|\big(D_{x_i}p_j^v(t) - D_{x_i}p_j^u(t)\big)\, \partial_{p_j} f_2(p^v(t))\|_{L^2}^2 \\
&\qquad + \|D_{x_i}p_j^u(t)\, \big(\partial_{p_j} f_2(p^u(v)) - \partial_{p_j} f_2(p^u(t))\big)\|_{L^2}^2 \Big) \\
&\leq\ 2m \sum_{j=1}^m \Big(\overline{q}(\kappa)^2\, \|D_{x_i}p_j^u(t+h) - D_{x_i}p_j^u(t)\|_{L^2}^2 + \\
&\qquad + \|D_{x_i}p_j^u(t)\|_{L^2}^2\, \|\partial_{p_j} f_2(p^u(t+h)) - \partial_{p_j} f_2(p^u(t))\|_{L^\infty}^2 \Big) \\
&\leq\ \sum_{j=1}^m \Big(\overline{q}(\kappa)^2\, \|D_{x_i}p^v(t) - D_{x_i}p^u(t)\|_{L^2}^2 + \\
&\qquad + C_{p_j}(T)^2\, L^2 \sum_{k=1}^m \|p_k^v(t)) - p_k^u(t))\|_{L^\infty}^2 \Big),
\end{aligned}$$

where L is an appropriate local Lipschitz constant for $p \mapsto \partial_{p_j} f_2(p)$ and $C_{p_j}(T) := \|p_j^u\|_{L^\infty((0,T),H^1)}$ for all $j = 1,\ldots,m$. Moreover, as T_0 an $\|u\|_{C^0([0,T_0],H^1)}$ are finite, (3.4.20) implies that there exists a common upper bound to the constants $C_{p_j}(T)$ for all $j = 1,\ldots,m$, independent of $T \in [0,T_0)$.

Hence, after employing the Lipschitz-type estimate (3.4.12) a summation over all $i \leq d$ and adding an analogous Lipschitz-type estimate for $\big\|f_2(p^u(t+h)) - f_2(p^u(t))\big\|_{L^2}^2$ we obtain that

$$\begin{aligned}
&\big\| f_2(p^v(t)) - f_2(p^u(t))\big\|_{H^1}^2 \\
&\leq\ C \sum_{j=1}^m \Big(\|p_j^v(t) + p_j^u(t)\|_{H^1}^2 + h^2 + \int_0^t \|v(s) - u(s)\|_{H^2}^2\,\mathrm{d}s \Big)
\end{aligned}$$

for a suitable constant $C < \infty$ independent of $T \in [0,T_0)$. Finally, employing the Lipschitz-type estimate (3.4.13) we obtain

$$\big\| f_2(p^v(t)) - f_2(p^u(t))\big\|_{H^1}^2 \leq C\Big(h^2 + \int_0^t \|v(s) - u(s)\|_{H^2}^2\,\mathrm{d}s \Big) \tag{3.4.42}$$

for a suitable constant $C < \infty$ independent of $T \in [0,T_0)$.

Ultimately, for the only term left to consider we obtain the estimate

$$\begin{aligned}
&\big\|D_{x_i}\big((f_1(p^v(t)) - f_1(p^u(t)))\,u(t)\big)\big\|_{L^2}^2 \\
&\quad\le \big\|D_{x_i}\big(f_1(p^v(t)) - f_1(p^u(t))\big)\big\|_{L^2}^2\,\|u(t)\|_{L^\infty}^2 + \|f_1(p^v(t)) - f_1(p^u(t))\|_{L^\infty}^2\,\|D_{x_i}u(t)\|_{L^2}^2 \\
&\quad\le \kappa\,\|D_{x_i}\big(f_1(p^v(t)) - f_1(p^u(t))\big)\big\|_{L^2}^2 + \|u(t)\|_{H^1}^2 L_f^2 \sum_{j=1}^m \|p_j^v(t) - p_j^u(t)\|_{L^\infty}^2\,.
\end{aligned}$$

Now, $\|u(t)\|_{H^1}^2 \le \|u\|_{C^0([0,T_0],H^1)}^2$ and the first difference in the last right hand side can be estimated just as the analogous difference for f_2 yielding (3.4.42). We apply the Lipschitz-type inequality (3.4.12) to the difference in the gating variables and overall obtain

$$\big\|f_1(p^v(t)) - f_1(p^u(t))\,u\big\|_{H^1}^2 \le C\Big(h^2 + \int_0^t \|v(s) - u(s)\|_{H^2}^2\,\mathrm{d}s\Big)$$

for a suitable constant $C < \infty$ independent of $T \in [0, T_0)$ and hence

$$\|f_h(t)\|_{H^1}^2 \;\le\; C\Big(h^2 + \int_0^t \|v(s) - u(s)\|_{H^2}^2\,\mathrm{d}s\Big)\,.$$

Integrating both sides of this inequality over $(0, T)$ we obtain the estimate

$$\|f_h\|_{L^2((0,T),H^1)}^2 \le C\,\Big(h^2 + \int_0^T \|v(t) - u(t)\|_{H^2}^2\,\mathrm{d}s\Big) \tag{3.4.43}$$

for a suitable constant $C < \infty$ independent of $T \in [0, T_0)$.

(ii.4) Next, we estimate the input current term in F_h. To this end note that for all $t \le T_0$

$$\int_0^t \|I(t+h) - I(t)\|_{L^2}^2\,\mathrm{d}t \le \int_0^{T_0} \|I(t+h) - I(t)\|_{L^2}^2\,\mathrm{d}t \le o(h)\,, \tag{3.4.44}$$

as the shift operator is continuous in L^2, see [81]. In case (b) it further holds that $\dot I(\cdot + h) - \dot I \in L^2((0,T_0), L^2)$ and $I(\cdot + h) - I \in L^2((0,T_0), H^1)$. Hence, due to the continuity of the shift operator we obtain in this case also that

$$\|I(\cdot + h) - I\|_{H^1((0,T_0),L^2)}^2 = o(h), \qquad \|I(\cdot + h) - I\|_{L^2((0,T_0),H^1)}^2 = o(h)\,. \tag{3.4.45}$$

(ii.5) We next combine the estimates derived in (i.1)–(i.4) and use them to estimate the right hand sides (3.4.34) and (3.4.35). Thus in case (a) we infer from (3.4.34) using the estimate (3.4.36) and (3.4.44) that

$$\|u_h(T)\|_{H^1}^2 \leq C\Big(\|u_h(0)\|_{H^1}^2 + o(h) + \int_0^T \|u_h(s)\|_{L^2}^2 \,\mathrm{d}t\Big)$$

and in part (b) we infer from (3.4.35) using additionally the estimates (3.4.41), (3.4.43) and (3.4.45) that

$$\|u_h(T)\|_{H^2}^2 \leq C\Big(\|u_h(0)\|_{H^2}^2 + o(h) + \int_0^T \|u_h(s)\|_{H^2}^2 \,\mathrm{d}t\Big)$$

for all $T \in [0, T_0)$ such that $T + h \in [0, T_0)$ and the constant C depends only on T_0. Then an application of Gronwall's Lemma yields

$$\begin{aligned} &\text{(a)} \qquad \|u_h(t)\|_{H^1}^2 \leq C\mathrm{e}^{CT_0}\Big(o(h) + \|u_h(0)\|_{H^1}^2\Big)\,, \\ &\text{(b)} \qquad \|u_h(t)\|_{H^2}^2 \leq C\mathrm{e}^{CT_0}\Big(o(h) + \|u_h(0)\|_{H^2}^2\Big)\,. \end{aligned}$$

Finally, as (a) $u \in C^0([0,T], H^1)$ and (b) $u \in C^0([0,T], H^2)$ for all $T < T_0$ it follows that $\|u_h(0)\|_X = \|u(h) - u_0\|_X$ becomes arbitrarily small for $h \to 0$ in the respective norm $X = H^1$ or $X = H^2$. In particular, for every $\epsilon > 0$ there exists $\delta > 0$ such that $\|u_h(t)\|_X < \epsilon$ for all $h < \delta$ and any t such that $t + h < T_0$, i.e., uniform continuity (3.4.32) holds.
Overall we infer that the limit $u(T_0)$ exists in the space H_0^1 or H^2 respectively and the contradiction to the maximality and finiteness of T_0 follows. The existence part of the proof is completed. Finally, uniqueness follows from the uniqueness of the local solution.

(iii) It remains to show the continuous dependence on the initial data. Let (u, p^u) and (v, p^v) denote the solutions to two different set of initial conditions (u_0, p_0^u), (v_0, p_0^v) and two different input currents I_u and I_v. Then the difference $z := u - v$ between the two solutions satisfies the equation

$$\begin{aligned} \dot{z} &= \Big(A - f_1(p^u(t))\Big)z + F(t), \\ z(t,x) &= 0 \qquad \forall\,(t,x) \in [0,T] \times \partial D, \\ z(0) &= u_0 - v_0\,, \end{aligned}$$

where

$$F(t) := -\Big(f_1(p^v(t)) - f_1(p^u(t))\Big)v(t) + \Big(f_2(p^u(t)) - f_2(p^v(t))\Big) + \Big(I_u(t) - I_v(t)\Big)\,.$$

We note that this is exactly the same structure as the initial-boundary value problem

(3.4.33) we consider in part (ii). Therefore using the same arguments we obtain for all $t \geq 0$ the estimates (a)

$$\|u(t) - v(t)\|_{H^1}^2 \leq C_1\Big(\|u_0 - v_0\|_{H^1}^2 + \sum_{i=1}^{m} \|p_i^u(0) - p_i^v(0)\|_{H^1}^2 + \|I_u - I_v\|_{L^2((0,T),L^2)}^2\Big)$$
$$+ C_1 \int_0^t \|u(t) - v(t)\|_{H^1}^2 \mathrm{d}s$$

and (b)

$$\|u(t) - v(t)\|_{H^2}^2 \leq C_1\Big(\|u_0 - v_0\|_{H^2}^2 + \sum_{i=1}^{m} \|p_i^u(0) - p_i^v(0)\|_{H^2}^2 + \|I_u - I_v\|_{H^1((0,T),L^2)}^2$$
$$+ \|I_u - I_v\|_{L^2((0,T),H^1)}^2\Big) + C_1 \int_0^t \|u(t) - v(t)\|_{H^2}^2 \mathrm{d}s\,,$$

where the constants C_1, C_2 may depend on t. An application of Gronwall's inequality yields the proposed dependence of u on the initial data. Finally, the continuous dependence on the initial data of the gating variables can be established along the lines of establishing the Lipschitz-type estimates of Lemma 3.4.1. For the sake of brevity of presentation and avoiding repetition we omit the detailed calculations. The proof of well-posedness of the general excitable media system is completed.

□

3.4.3 Invariant regions equations of excitable media

One of the statements of Theorem 3.1.1 is that the membrane variable, i.e., solutions to the PDE problem (3.1.2) with coefficients (3.1.3) or (3.1.4), remains bounded, that is, it stays in the physiological domain, whenever the initial values are in this domain. The consequence of modelling the conductances as, e.g., point sources (3.1.3) or by (3.1.4), is that we have to deal with non-smooth coefficients. The usual way of treating this is to approximate the solutions of the problem by smooth functions satisfying the same PDE problem, prove the desired results in this case, and then ensure that the result carries over in the limit of the smooth approximations to the solution of the original problem. In particular, we prove a more general statement, which is formulated in Theorem 3.4.2 below, than the boundedness of solutions to the cable equation (3.1.2) in the sense that we consider general parabolic PDEs on arbitrary domains $D \subset \mathbb{R}^d$. Then this result can be readily applied to the various extensions of the models discussed in Section 3.1.1 as well as, with slight adaptations, for deterministic models of excitable media. This provides the proof of Proposition 3.3.1 as announced. The proofs we present in this section mainly rely on results from

functional analysis and we refer to [2, 122] for the theoretical background.
Next we first describe the setting generalising that of hybrid stochastic models considered in Theorem 3.1.1, and the approximating smoothed sequence. Then Theorem 3.4.2 is stated and proved. Finally, at the end of the section we discuss the various applications of the theorem in Sections 3.1 and 3.3.

Subsequently D denotes a bounded domain in $\mathbb{R}^d$ with sufficiently smooth boundary ∂D with $\overline{D} = D \cup \partial D$ denoting the closure of D. Function spaces always refer to spaces of real-valued functions over the domain D, e.g., $H_0^1 = H_0^1(D)$, or the closure $\overline{D}$ for spaces of continuous functions, respectively, e.g., $C^0 = C^0(\overline{D})$. Further, we note that the duality pairing $\langle \cdot, \cdot \rangle_{H^1}$ in H_0^1 is understood as the inner product on L^2 whenever the inner product is well defined.

Let $u \in L^2((0,T), H_0^1) \cap H^1((0,T), H^{-1})$ be the unique solution of the initial value problem

$$\begin{aligned} \dot{u} &= Au + \sum\nolimits_i \overline{g}_i\, d_i\, (E_i - u),\,, \\ u(t,x) &= 0 \qquad \forall\, (t,x) \in [0,T] \times \partial D\,, \\ u(0) &= u_0 \in H_0^1\,. \end{aligned} \tag{3.4.46}$$

The parameters $\overline{g}_i$ are non-negative, real weights and d_i are elements of H^{-1} such that each summand $\overline{g}_i\, d_i\, (v_i - u)$ in (3.4.46) is well defined as an element of H^{-1}. Here A is a strongly elliptic second order partial differential operator of the form

$$Au = \sum_{i,j=1}^{m} (a_{ij}(x) u_{x_i})_{x_j} + \sum_{i=1}^{m} b_i(x) u_{x_i} + c(x) u \tag{3.4.47}$$

with coefficients a_{ij}, b_i and c sufficiently smooth, i.e., they have to be at least in L^∞.

Further we consider a sequence of approximations $u_n \in C^2([0,T] \times \overline{D})$ for $n \in \mathbb{N}$ to the solution u of (3.4.46), which are themselves solutions to the initial value problems given by

$$\begin{aligned} \dot{u}_n &= A^n u_n + \sum\nolimits_i \overline{g}_i\, d_i^n(x)\, (E_i - u_n)\,, \\ u_n(t,x) &= 0 \qquad \forall\, (t,x) \in [0,T] \times \partial D\,, \\ u_n(0) &= u_0^n \in C^\infty\,. \end{aligned} \tag{3.4.48}$$

Here A^n are strongly elliptic second order partial differential operators of the form

$$A^n u = \sum_{i,j=1}^{m} (a_{ij}^n(x) u_{x_i})_{x_j} + \sum_{i=1}^{m} b^n(x) u_{x_i} + c^n(x) u \tag{3.4.49}$$

with coefficients a_{ij}^n, b_i^n and c^n in C^∞. Finally we state the pointwise bounds we aim to prove for the solution u. Let $\overline{u}_- \leq 0$ and $\overline{u} \geq 0$ be such that for all $n \geq 1$

$$\begin{aligned} c^n(x)\overline{u}_- + \sum_{i=1}^m \overline{g}_i\, d_i^n(x)\,(E_i - \overline{u}_-) \geq 0 \quad \forall\, x \in D, \\ c^n(x)\overline{u}_+ + \sum_{i=1}^m \overline{g}_i\, d_i^n(x)\,(E_i - \overline{u}_+) \leq 0 \quad \forall\, x \in D. \end{aligned} \tag{3.4.50}$$

In particular for models of excitable media we have that $c(x) \equiv 0$ and thus we may choose $\overline{u}_- = \min E_i$ and $\overline{u}_+ = \max E_i$. Also it is typically satisfied that then $\overline{u}_- < 0$ and $\overline{u}_+ > 0$. The existence of such bounds is guaranteed if c^n are non-positive and uniformly bounded.

To establish the convergence of u_n to u we assume the following conditions to hold.

(i) The sequence of initial conditions u_0^n converges to u_0 in L^2 which in particular implies that u_0^n is bounded in L^2.

(ii) The coefficients a_{ij}^n, b_i^n, c^n converge to a_{ij}, b_i and c such that $A^n \to A$ in the operator norm for linear operators from H_0^1 to H^{-1}.

(iii) The sequence of operators A^n satisfies the uniform energy estimate

$$\exists\; C_1 > 0 : |\langle A^n u, v\rangle_{H^1}| \leq C_1 \|u\|_{H_0^1} \|v\|_{H_0^1} \quad \forall\; u, v \in H_0^1,\; n \geq 1, \tag{3.4.51}$$

as well as a uniform coercivity condition, i.e.,

$$\exists\; C_2 > 0,\; C_3 \geq 0 : -\langle A^n u, u\rangle_{H^1} \geq C_2 \|u\|_{H_0^1}^2 - C_3 \|u\|_{L^2}^2 \quad \forall\; u \in H_0^1,\; n \geq 1. \tag{3.4.52}$$

We note at this point that the choice of Dirichlet boundary conditions for (3.4.46) and (3.4.48) only reflects the setup as discussed in Section 3.1. In general the statement of the theorem below is valid for other choices of boundary conditions as long as the uniform conditions (3.4.51) and (3.4.52) are satisfied. In particular this holds for the boundary conditions discussed in Section 3.3.1. For different boundary conditions we obviously have to consider the space H^1 and its dual instead of H_0^1.

(iv) The approximations $d_i^n \in C^\infty$ converge to d_i weakly in H^{-1} and satisfy the uniform bounds

$$\|d_i^n\|_{L^1} \leq K_1 \quad \text{and} \quad \|d_i^n v\|_{L^1} \leq K_2 \|v\|_{H_0^1} \quad \forall\, v \in H_0^1 \tag{3.4.53}$$

for constants K_1, K_2 independent of n. Finally, d_i^n is such that for sequences v_n which in $L^2((0,T), H_0^1)$ converge weakly to v with $\dot{v}_n$ converging weakly to $\dot{v}$ in

$L^2((0,T), H^{-1})$, it holds that for almost all $t \in [0,T]$

$$\langle d_i^n v_n(t), \phi\rangle_{H^1} \to \langle d_i v(t), \phi\rangle_{H^1} \quad \forall\, \phi \in H_0^1. \tag{3.4.54}$$

Theorem 3.4.2. *Under the assumptions* (i) - (iv) *and if the initial conditions are such that* $\overline{u}_- \le u(0,x)$, $u_n(0,x) \le \overline{u}_+$ *for all, respectively, almost all* $x \in \overline{D}$ *and all* $n \in \mathbb{N}$, *the solutions* u_n *and* u *of* (3.4.48) *and* (3.4.46), *respectively, satisfy*

$$\overline{u}_- \le u_n(t,x) \le \overline{u}_+ \quad \text{for all } (t,x) \in [0,T] \times \overline{D} \tag{3.4.55}$$

for all $n \in \mathbb{N}$, *and*

$$\overline{u}_- \le u(t,x) \le \overline{u}_+ \quad \text{for almost all } x \in D \text{ for all } t \in [0,T]\,. \tag{3.4.56}$$

If, in addition, $u(t) \in C^0$ *for all* $t \in [0,T]$ *then the bounds* $\overline{u}_-, \overline{u}_+$ *are valid everywhere, i.e.,* (3.4.56) *holds for all* $(t,x) \in [0,T] \times \overline{D}$.

Proof. We split the proof into four parts. In part (a) we show that the smooth solutions u_n satisfy the imposed bounds $[\overline{u}_-, \overline{u}_+]$. In part (b) we show that under the assumptions (i)–(iv) the sequence of problems (3.4.48) possesses a subsequence that converges weakly to a limit u^*. Then in part (c) we show that the limit satisfies $u^* = u$ and thus solves (3.4.46). We complete the proof in part (d) where we show that the weak convergence in this case implies that also u satisfies the bounds $[\overline{u}_-, \overline{u}_+]$ almost everywhere.

(a) We proof (3.4.55) by contradiction. Let t_0 be the infimum of the times t such that there exists an $x \in \overline{D}$ with either $u_n(t,x) \ge \overline{u}_+$ or $u_n(t,x) \le \overline{u}_-$. Assume that t_0 is finite. Then it holds due to the continuity of u_n that $\overline{u}_- \le u_n(t_0,x) \le \overline{u}_+$ for all $x \in \overline{D}$ and that there exists an $x_0 \in \overline{D}$ such that either $u_n(t_0,x_0) = \overline{u}_-$ or $u_n(t_0,x_0) = \overline{u}_+$.

We consider the case that $u_n(t_0,x_0) = \overline{u}_+$, the case that $u_n(t_0,x_0) = \overline{u}_-$ is then analogous. Thus the point x_0 is a maximum of the map $x \mapsto u_n(t_0,x)$ and thus its first derivatives vanish. Further, we apply the fact that the entrywise product of positive semidefinite matrices is positive semidefinite to the entrywise product of the Hessian matrix and the matrix of the coefficients a_{ij}^n of the elliptic operator A^n and obtain that

$$\sum_{i,j=1}^m a_{ij}^n(x_0)(u_n(t_0,x_0))_{x_i x_j} \le 0\,.$$

Next, as by definition of $\overline{u}_+$ it holds that

$$c(x)\overline{u}_+ + \sum_{i=1}^m \overline{g}_i d_i^n(x)(E_i - \overline{u}_+) \le 0,$$

we infer that $\dot{u}_n(t_0, x_0) \leq 0$. This is a contradiction to $u_n(t,x) < \overline{u}_+$ for all $t < t_0$. Hence t_0 cannot be finite. The property (3.4.55) is proven.

(b) In this part of the proof we show that the sequence u_n admits a limit in the appropriate sense, i.e., it possesses a subsequence which is weakly convergent in $L^2((0,T), H_0^1)$ and $H^1((0,T), H^{-1})$. To this end we have to show boundedness of the sequence u_n in these two spaces.

In order to obtain the boundedness result in the first space we apply the inner product in L^2 with $u_n(t)$ to both sides of equation (3.4.48) and arrive at

$$\big(\dot{u}_n(t), u_n(t)\big)_{L^2} - \big(A^n u_n(t), u_n(t)\big)_{L^2} = \sum_i \overline{g}_i \big(d_i^n\,(E_i - u_n(t)), u_n(t)\big)_{L^2}. \qquad (3.4.57)$$

Using the uniform coercivity condition (3.4.52) we get

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u_n(t)\|_{L^2}^2 + C_2\|u_n(t)\|_{H_0^1}^2 \leq C_3\|u_n(t)\|_{L^2}^2 + \sum_i \overline{g}_i\,\big(d_i^n\,(E_i - u_n(t)), u_n(t)\big)_{L^2}.$$

Thus an application of Gronwall's inequality yields for $\|u_n(t)\|_{L^2}^2$ the estimate

$$\|u_n(t)\|_{L^2}^2 \;\leq\; \mathrm{e}^{2C_3 t}\Big(\|u_n(0)\|_{L^2}^2 + 2\sum_i \overline{g}_i \int_0^t \big|\big(d_i^n\,(E_i - u_n(t)), u_n(t)\big)_{L^2}\big|\,\mathrm{d}t\Big).$$

Integrating both sides of (3.4.57) from 0 to T and using the just established estimate in the right hand side yields

$$\begin{aligned}
&\frac{1}{2}\big(\|u_n(T)\|_{L^2}^2 - \|u_n(0)\|_{L^2}^2\big) + C_2\int_0^T \|u_n(t)\|_{H_0^1}^2\,\mathrm{d}t \\
&\leq C_3\int_0^T \mathrm{e}^{2C_3 t}\,\mathrm{d}t\,\|u_0\|_{L^2}^2 + \Big(2C_3\int_0^T \mathrm{e}^{2C_3 t}\,\mathrm{d}t + 1\Big)\sum_i \overline{g}_i \int_0^T \big|\big(d_i^n\,(E_i - u_n(t)), u_n(t)\big)_{L^2}\big|\,\mathrm{d}t \\
&= \frac{1}{2}(\mathrm{e}^{2C_3 t} - 1)\,\|u_n(0)\|_{L^2}^2 + \mathrm{e}^{2C_3 t}\sum_i \overline{g}_i \int_0^T \big|\big(d_i^n\,(E_i - u_n(t)), u_n(t)\big)_{L^2}\big|\,\mathrm{d}t.
\end{aligned}$$

Hence it remains to bound the integrands in the right hand side. Due to the pointwise bounds on u_n obtained in part (a) we obtain for every integral using (3.4.53)

$$\begin{aligned}
\int_0^T \big|\big(d_i^n\,(E_i - u_n(t)), u_n(t)\big)_{L^2}\big|\,\mathrm{d}t \;&\leq\; \int_0^T\int_D d_i^n(x)\big|(E_i - u_n(t,x))\,u_n(t,x)\big|\mathrm{d}x\,\mathrm{d}t \\
&\leq\; \int_0^T\int_D d_i^n(x)\;2\overline{u}^2\;\mathrm{d}x\,\mathrm{d}t \\
&\leq\; 2TK_1\overline{u}^2,
\end{aligned}$$

where $\overline{u} = \max\{|\overline{u}_-|, |\overline{u}_+|\}$. Therefore, we have obtained the estimate

$$\frac{1}{2}\|u_n(T)\|_{L^2}^2 + C_2\|u_n\|^2_{L^2((0,T),H_0^1)} \leq \mathrm{e}^{2C_3T}\Big(\|u_n(0)\|^2_{L^2} + 2TK_1\overline{u}^2\sum_i \overline{g}_i\Big). \tag{3.4.58}$$

The boundedness of u_n in $L^2((0,T),H_0^1)$ now follows from the boundedness of the sequence of initial conditions, cf. condition (i).

Secondly, we show the boundedness of u_n in $H^1((0,T),H^{-1})$, i.e., we have to show that the sequences u_n and $\dot{u}_n$ are bounded in $L^2((0,T),H^{-1})$. Since H_0^1 is continuously embedded in H^{-1}, i.e., there exists $c > 0$ such that $\|v\|_{H^{-1}} \leq c\|v\|_{H_0^1}$ for all $v \in H_0^1$, the boundedness of the sequence u_n in $L^2((0,T),H^{-1})$ follows from the just established boundedness of u_n in $L^2((0,T),H_0^1)$.
Thus it remains to show that $\dot{u}_n$ is bounded in $L^2((0,T),H^{-1})$. To this end we fix some $v \in H_0^1$ such that $\|v\|_{H_0^1} \leq 1$. Then we obtain by applying the inner product in L^2 with v to both sides of equation (3.4.48)

$$\big(\dot{u}_n(t),v\big)_{L^2} = \big(A^n u_n(t),v\big)_{L^2} + \sum_i \overline{g}_i\big(d_i^n(E_i - u_n(t)),v\big)_{L^2}.$$

Employing the uniform energy estimate (3.4.51) yields

$$\big|\big(\dot{u}_n(t),v\big)_{L^2}\big| \leq C_1\|u_n(t)\|_{H_0^1} + \sum_i \overline{g}_i\big|\big(d_i^n\,(E_i - u_n(t)),v\big)_{L^2}\big|,$$

where each summand in the right hand side is uniformly bounded as

$$\begin{aligned}\big|\big(d_i^n\,(E_i - u_n(t)),v\big)_{L^2}\big| &\leq \int_D \big|d_i^n(x)\,(E_i - u_n(t,x))\,v(x)\big|\,\mathrm{d}x\\ &\leq \|E_i - u_n(t)\|_{L^\infty}\,\|d_i^n v\|_{L^1}\\ &\leq 2\overline{u}K_2\,.\end{aligned}$$

Here we have used Hölder's inequality for the second and the condition (3.4.53) for the third estimate. Since the norm on H^{-1} is defined as

$$\|w\|_{H^{-1}} = \sup_{\substack{v\in H_0^1:\\ \|v\|_{H_0^1}=1}} |\langle w,v\rangle_{H^1}|,$$

it follows by the identification of H_0^1 with its dual H^{-1}, where we employ that $\langle \dot{u}_n, v\rangle_{H^1} = (\dot{u}_n, v)_{L^2}$ for all $v \in H_0^1$ as $\dot{u}_n \in L^2$, that

$$\|\dot{u}_n(t)\|_{H^{-1}} \leq C_1\|u_n(t)\|_{H_0^1} + 2\overline{u}K_2\Big(\sum_i \overline{g}_i\Big).$$

Hence squaring both sides of this inequality and integrating over $(0,T)$ yields

$$\int_0^T \|\dot{u}_n(t)\|^2_{H^{-1}}\mathrm{d}t \leq 2C_1^2\|u_n\|^2_{L^2((0,T),H_0^1)} + 4TK_2^2\overline{u}^2\Big(\sum_i \overline{g}_i\Big)^2. \tag{3.4.59}$$

Finally, the boundedness of $\dot{u}_n$ in $L^2((0,T),H^{-1})$ follows by the already established boundedness of u_n in $L^2((0,T),H_0^1)$.
Overall we now infer that there exists a subsequence of u_n which converges weakly to a limit u^* in $L^2((0,T),H_0^1)$ as well as in $L^2((0,T),H^{-1})$. For the remainder we identify u_n with its convergent subsequence.

(c) We next show that $u^* = u$, that is, the weak limit solves equation (3.4.46). Recall that equations (3.4.46) and (3.4.48) are interpreted in the sense of H^{-1}–valued distributions, that is, e.g., u satisfies (3.4.46) if for all test functions φ taking values in H_0^1 it holds that

$$\int_0^T \langle\dot{u}(t),\varphi(t)\rangle_{H^1}\,\mathrm{d}t = \int_0^T \langle Au(t),\varphi(t)\rangle_{H^1} + \sum_i \overline{g}_i\langle d_i\,(E_i - u(t)),\varphi\rangle_{H^1}\,\mathrm{d}t \tag{3.4.60}$$

and $u(0) = u_0$ in L^2. We now consider for the convergent subsequence u_n for each term in the definition of its solution, i.e., the weak formulation of the problem (3.4.48), the convergence to the corresponding term in (3.4.60) separately.

As for all test functions φ the mapping $u \to \int_0^T \langle\dot{u},\varphi\rangle_{H^1}\,\mathrm{d}t$ defines a continuous, linear functional on $H^1((0,T),H^{-1})$, it follows from the weak convergence of $\dot{u}_n$ to $\dot{u}^*$ that for $n\to\infty$

$$\int_0^T \langle\dot{u}_n(t),\varphi(t)\rangle_{H^1}\,\mathrm{d}t \to \int_0^T \langle\dot{u}^*(t),\varphi(t)\rangle_{H^1}\,\mathrm{d}t. \tag{3.4.61}$$

Further, for every φ we can define continuous, linear functionals on $L^2((0,T),H_0^1)$ by

$$u \mapsto \int_0^T \langle A^n u(t),\varphi(t)\rangle_{H^1}\,\mathrm{d}t,\quad u\mapsto \int_0^T \langle Au(t),\varphi(t)\rangle\,\mathrm{d}t. \tag{3.4.62}$$

It is easy to see [122, Thm. 21.23] that for $n\to\infty$

$$\int_0^T \langle A^n u^n(t),\varphi(t)\rangle_{H^1}\,\mathrm{d}t \to \int_0^T \langle Au^*(t),\varphi(t)\rangle_{H^1}\,\mathrm{d}t$$

as A^n converges to A strongly in $L(H_0^1,H^{-1})$ and u_n is bounded in $L^2((0,T),H_0^1)$.
Next we consider the terms $d_i^n u_n$. The weak convergence of u_n to u^* in $L^2((0,T),H_0^1)$ implies the weak convergence of $u_n(t)$ in H_0^1 for almost every $t\in(0,T)$. Thus (3.4.54) yields that for almost all $t\in(0,T)$

$$\langle d_i^n u_n(t),\varphi(t)\rangle_{H^1} \to \langle d_i\,u^*(t),\varphi(t)\rangle_{H^1}\,.$$

Therefore by dominated convergence we obtain

$$\int_0^T \langle d_i^n u_n(t), \varphi(t)\rangle_{H^1}\,\mathrm{dt} \to \int_0^T \langle d_i\, u^*(t), \varphi(t)\rangle_{H^1}\,\mathrm{dt}\,.$$

Finally, as d_i^n converges weakly to d_i in H^{-1} for all i, it is valid by dominated convergence that

$$\int_0^T \langle d_i^n, \varphi(t)\rangle_{H^1}\,\mathrm{dt} \to \int_0^T \langle d_i, \varphi(t)\rangle_{H^1}\,\mathrm{dt}\,.$$

Hence overall we obtain

$$\int_0^T \langle A^n u_n(t), \varphi(t)\rangle_{H^1} + \sum_i \overline{g}_i \langle d_i^n\,(E_i - u_n(t)), \varphi(t)\rangle_{H^1}\,\mathrm{dt}$$

$$\to \quad \int_0^T \langle Au^*(t), \varphi(t)\rangle_{H^1} + \sum_i \overline{g}_i \langle d_i\,(E_i - u^*(t)), \varphi(t)\rangle_{H^1}\,\mathrm{dt}.$$

As in this formula the left hand side equals $\int_0^T \langle \dot{u}_n(t), \varphi(t)\rangle_{H^1}\,\mathrm{dt}$ we obtain by (3.4.61) that

$$\int_0^T \langle \dot{u}^*(t), \varphi(t)\rangle_{H^1}\,\mathrm{dt} = \int_0^T \langle Au^*(t), \varphi(t)\rangle_{H^1} + \sum_i \overline{g}_i \langle d_i\,(E_i - u^*(t)), \varphi(t)\rangle_{H^1}\,\mathrm{dt}, \tag{3.4.63}$$

that is, the weak limit u^* satisfies equation (3.4.60).

It just remains to show that u^* also satisfies the initial condition, that is $u^*(0) = u_0$ in L^2, which we obtain using similar argumentation as in [104, Ch. 7.4.4]. Choosing a test function φ that satisfies $\varphi(T) = 0$ and integration by parts yields for the approximating sequence

$$-\int_0^T \langle u_n(t), \dot{\varphi}(t)\rangle_{H^1}\,\mathrm{dt} - \langle u_0^n, \varphi(0)\rangle_{H^1} = \int_0^T \langle Au_n(t) + \sum_i \overline{g}_i\, d_i\,(E_i - u_n(t)), \varphi(t)\rangle_{H^1}\,\mathrm{dt}, \tag{3.4.64}$$

and for the limit in (3.4.63)

$$-\int_0^T \langle u^*(t), \dot{\varphi}(t)\rangle_{H^1}\,\mathrm{dt} - \langle u^*(0), \varphi(0)\rangle_{H^1} = \int_0^T \langle Au^*(t) + \sum_i \overline{g}_i\, d_i\,(E_i - u^*(t)), \varphi(t)\rangle_{H^1}\,\mathrm{dt}. \tag{3.4.65}$$

Note that u_0^n converges strongly to u_0 and we have already established the convergence of the integral terms in (3.4.64) to the corresponding terms in (3.4.65). Hence, it follows by comparison that $u^*(0) = u_0$ in L^2 since $\varphi(0)$ is arbitrary.

(d) So far we obtained that there exists a subsequence of u_n which converges weakly to u and satisfies $u_n(t,x) \in [\overline{u}_-, \overline{u}_+]$ pointwise for all n. To complete the proof we show that the bounds remain valid for the weak limit. To this end we employ the

following fact [61, p. 336]: Let K be a convex subset of a normed vector space V. Then, if $x^* \in V$ is the weak limit of a sequence in K, there exists a sequence in K such that x^* is its strong limit.
We take K to be the set

$$K = \{v \in C^\infty([0,T] \times \overline{D}) \mid \overline{u}_- \leq v(t,x) \leq \overline{u}_+ \ \forall\,(t,x)\}\,,$$

which is a convex subset of $L^2((0,T),L^2)$. As just established in part (c), u is the weak limit of a sequence in K, thus there exists a sequence $\widehat{u}_n$ in K such that

$$\lim_{n\to\infty} \int_0^T \|\widehat{u}_n(t) - u(t)\|_{L^2}^2 \,\mathrm{d}t = 0\,.$$

This implies that for almost every $t \in [0,T]$ the limit $u(t,x)$ satisfies the bounds $[\overline{u}_-, \overline{u}_+]$ for almost every $x \in D$. Next, recall that we always identify a weak solution with the unique element in its equivalence class which is in $C^0([0,T],L^2)$. Next assume that t^* is such that $u(t^*,x)$ does not satisfy the bounds $[\overline{u}_-, \overline{u}_+]$ for almost every $x \in D$. Then, however, there exists a sequence of t_n in $(0,T)$, $t_n \to t^*$ for $n \to \infty$, such that $u(t_n,x) \in [\overline{u}_-, \overline{u}_+]$ for almost all $x \in D$ for all $n \geq 0$. As the solution u is in $C^0([0,T],L^2)$ it follows that $u(t_n)$ converges strongly to $u(t^*)$ in L^2. This is a contradiction to $u(t^*,x) \notin [\overline{u}_-, \overline{u}_+]$ on a set of positive measure. Hence (3.4.56) is proved.

Finally, if in addition $u(t) \in C^0$ for all $t \in [0,T]$ the validity of the pointwise bounds for all $x \in \overline{D}$ follows immediately. Assume that for some t there exists a point x^* such that $u(t,x^*) \notin [\overline{u}_-, \overline{u}_+]$. Then there exists a neighbourhood of x^* in D such that $u(t,x) \notin [\overline{u}_-, \overline{u}_+]$ for all x in that neighbourhood. A neighbourhood has non-zero measure which yields a contradiction to (3.4.56). The proof is completed.

□

We conclude this section with a discussion of the various applications of Theorem 3.4.2 in this thesis.

Application 1: The first application is to hybrid stochastic models of axons as presented in Section 3.1. The membrane equations (3.1.2) for these models in between successive channel switchings together with a model for the conductances, i.e., (3.1.3) or (3.1.4), is a special cases of the setting presented above for $d = 1$, i.e, the domain D is an interval in $\mathbb{R}$. We first discuss the conductances (3.1.3) modelled as delta inputs and defer the discussion of other, more regular, conductance models, see Application 2 below. Then the equation (3.4.46) reduces to the specific problem (3.1.2) with conductances (3.1.3), the coefficients of A possessing the specific form $a_{11} \equiv r/2RC$ and $b_1 \equiv c \equiv 0$ and the weights $\overline{g}_i \equiv \overline{g}_{i,\theta}$ for each channel configuration θ. Regarding the approximating solutions u_n we remark that condition (i) is satisfied as for every initial condition $u_0 \in H_0^1$ an approximating sequence of smooth functions can always be found due to $C^\infty \cap H^1$ being dense in H^1 [2]. Further, conditions (ii) and (iii) are easily satisfied for the choice of coefficients $a_{11}^n = a_{11}$ for all $n \geq 1$.

It remains to consider condition (iv) for a one-dimensional domain and the choices of conductances (3.1.3). For the conductances $d_i = \delta(\cdot - x_i)$ being the Dirac delta function at x_i we consider as an approximation the standard mollifier μ_n, i.e.,

$$d_i^n(x) = \mu_n(x - x_i) = \begin{cases} n\, C_i^n \exp\left(\frac{1}{n^2|x-x_i|^2-1}\right) & n^2|x - x_i| < 1 \\ 0 & n^2|x - x_i| \geq 1, \end{cases}$$

where the constants C_i^n are selected such that $\int_D d_i^n(x)\,\mathrm{d}x = 1$. The Sobolev Embedding Theorem implies that $\delta(\cdot - x_i) \in H^{-1}$ and further d_i^n converges weakly to the delta function in H^{-1}. For the uniform bounds (3.4.53) we can obviously choose $K_1 = 1$, and Hölder's inequality together with the Sobolev Embedding Theorem yield $\|d_i^n v\|_{L^1} \leq \|d_i^n\|_{L^1}\|v\|_{L^\infty} \leq C\|v\|_{H_0^1}$ for a suitable constant $C < \infty$.
Finally, for a sequence v_n as in (iv) we apply [104, Thm. 8.1] for a choice of spaces $H_0^1 \overset{c}{\hookrightarrow} C^0 \hookrightarrow H^{-1}$ and obtain that there exists a subsequence of v_n which converges to a limit v strongly in $L^2((0,T), C^0)$. Here the compact embedding of H_0^1 into C^0 holds due to the Sobolev Embedding Theorem [2] as $d = 1$ and the continuous embedding of C^0 into H^{-1} follows from the continuous embedding of $C^0 \hookrightarrow L^2$ where the latter space in turn is continuously embedded in H^{-1}. Without loss of generality we can identify each sequence v_n with the strongly convergent subsequence. Next note that d_i^n and δ are elements of the dual of the Banach space C^0 and that in this space d_i^n converges weakly* to $\delta(\cdot - x_i)$. Moreover it holds that

$$\langle d_i^n v_n(t), \phi\rangle_{H^1} = \int_D d_i^n(x)\, v_n(t,x)\, \phi(x)\,\mathrm{d}x = \langle d_i^n, v_n(t)\phi\rangle_{C^0}$$

and due to the weak* convergence in the dual of C^0 and as $v_n(t)\phi$ converges strongly to $v(t)\phi$ in C^0 it follows by [122, Thm. 21(e)] that

$$\langle d_i^n v_n(t), \phi\rangle_{H^1} = \langle d_i^n, v_n(t)\phi\rangle_{C^0} \to \langle \delta(\cdot - x_i), v(t)\phi\rangle_{C^0} = \langle \delta(\cdot - x_i)v(t), \phi\rangle_{H^1}\,.$$

Therefore also (3.4.54) is satisfied.

Application 2: Next, we consider the same model setup as in Application 1 however with conductances as in (3.1.4) and we allow for the spatial domain all physically reasonable dimensions, i.e., $D \subset \mathbb{R}^d$ with $d \leq 3$ and sufficiently smooth boundary. For indicator functions $d_i = |C_i|^{-1}\mathbb{I}_{C_i}$, where $C_i \subset D$, possible choices for approximations are their mollifications which converge weakly to d_i, i.e., $d_i^n = \mu_n * d_i$, where $*$ denotes

the convolution of functions, cf. (2.2.10). Then Young's inequality for convolutions[8] yields that $\|\mu_n * d_i\|_{L_1} \leq \|\mu_n\|_{L^1}\|d_i\|_{L^1} = 1$ and thus $K_1 = 1$. For the second estimate we first consider Hölder's inequality which yields $\|\mu_n * d_i\, v\|_{L^1} \leq \|\mu_n * d^i\|_{L^2}\|v\|_{L^2}$ where $\|v\|_{L^2}$ can be estimated by $\|v\|_{H_0^1}$ due to the embedding $H^1 \hookrightarrow L^2$. Further, another application of Young's inequality yields that $\|\mu_n * d^i\|_{L^2} \leq \|\mu_n\|_{L^1}\|d^i\|_{L^2}$ where $\|\mu_n\|_{L^1} = 1$ and hence we can choose $K_2 = \max_i |C_i|^{-1/2}$. Hence conditions (3.4.53) are satisfied.
It remains to consider (3.4.54). To this end we apply [104, Thm. 8.1] for the choice of spaces $H_0^1 \overset{c}{\hookrightarrow} L^2 \hookrightarrow H^{-1}$ which yields that a sequence v_n as in (iv) possesses a subsequence that converges strongly to v in $L^2((0,T), L^2)$. Here the embedding of H^1 into L^2 is compact for $d \leq 3$ [116, p. 45]. Then by [122, Thm. 21(e)] it follows that

$$\langle d_i^n v_n(t), \phi\rangle_{H^1} = (d_i^n \phi, v_n(t))_{L^2} \to (d_i\phi, v(t))_{L^2} = \langle d_i\, v(t), \phi\rangle_{H^1}$$

as $d_i^n\phi$ converges strongly to $d_i\phi$ in L^2. Thus, also in this case condition (iv) is satisfied.

We note that the exact same line of argument shows that Theorem 3.4.2 is also applicable to compartmental type hybrid stochastic models, cf. Section 3.2.

Application 3: Finally, we can also employ this setup to prove an invariant rectangle for the membrane component of the deterministic axon equations (3.3.5), that is, to prove Proposition 3.3.1. We can assume there exists a sequence of operators A^n with smooth coefficients and also a sequence of smooth initial conditions such that conditions (i)–(iii) are satisfied. For most models, in particular for all models considered in this thesis, the operator A itself possesses smooth coefficients. Thus it remains to discuss condition (iv). For deterministic axon equations we set $d_i(x) = g_i(x, p(t,x))$ and $d_i^n = g^n(x, p_i^n(t,x))$ where $g^n(x,p)$ and $p^n(x,t)$ are smooth approximations converging pointwise to g_i and p, respectively. Moreover, we can assume that the approximations satisfy the same pointwise bounds, i.e., for all $i = 1,\dots,m$ and all $n \in \mathbb{N}$ it holds that $p_i^n(t,x) \in [0,1]$ for all $(t,x) \in [0,T] \times \overline{D}$ and $g_i^n(x,p) \leq \overline{g}$ for all $(x,p) \in \overline{D} \times [0,1]^m$. This choice implies that the weights $\overline{g}_i$ in (3.4.46) and (3.4.48) satisfy $\overline{g}_i = 1$.
Further, concerning the conditions in (vi) the uniform boundedness in L^1 is obvious as the gating variables are pointwise bounded by one and thus g_i^n is pointwise bounded, hence $K_1 = |D|\,\overline{g}$. The second condition in (3.4.53) with $K_2 = \sqrt{K_1}$ follows from the Cauchy-Schwarz inequality and the continuous embedding of H^1 into L^2. Finally d_i^n converges strongly to d_i in L^2 for almost all t, hence as an application of the inner product, i.e., considered as $(d_n^i, \cdot)_{L^2}$, it converges strongly in H^{-1}. Thus by [122,

[8]For $1 \leq p,q,r \leq \infty$ such that $1/p + 1/q = 1/r + 1$ it holds for all $f \in L^p$ and $g \in L^q$ that $\|f * g\|_{L^r} \leq \|f\|_{L^p}\|g\|_{L^q}$

Prop. 21.23(j)] condition (3.4.54) is satisfied.
However, one striking difference that necessitates minor technical adaptation in the proof of Theorem 3.4.2 is the additional time-dependence, on the one hand, in the functions d_i^n and, on the other hand, as the generalised membrane equation (3.3.5) possesses an additional time-dependent inhomogeneous term in the right hand side given by the input current. Such a term is not accounted for in (3.4.46). Firstly, we note that a time-dependence in the functions d_i^n and d_i does not alter the proof as the conditions in (iv) are satisfied with constants independent of t.
Secondly, however, the occurrence of the input current does demand additional attention. First of all, in accordance with the general setup we substitute $I \in L^2((0,T),L^2)$ by a smooth approximating sequence $I^n \in C^\infty([0,T]\times\overline{D})$ for the approximating sequence of equations (3.4.48). Due to the density of the latter space in the former such an approximation always exists. Further, we define the proposed pointwise bounds $[\overline{u}_-,\overline{u}_+]$ such that instead of (3.4.50) they satisfy (3.3.10), that is,

$$c^n(x)\overline{u}_- + \sum_{i=1}^m g^n(x,p)(E_i-\overline{u}_-) + I^n(t,x) \geq 0 \quad \forall\, x\in D,\, t\in[0,T],\, p\in[0,1]^m,$$
$$c^n(x)\overline{u}_+ + \sum_{i=1}^m g^n(x,p)(E_i-\overline{u}_+) + I^n(t,x) \leq 0 \quad \forall\, x\in D,\, t\in[0,T],\, p\in[0,1]^m.$$

The existence of such bounds is guaranteed if c^n are non-positive and are, together with the input current I^n, uniformly bounded.
Finally, in the proof of Theorem 3.4.2 the additional time-dependent inhomogeneous term I^n has to be considered in the estimation and convergence procedures. The necessary additions are the following: For the estimates in part (b) an additional term arises in the right hand side which can be estimated by

$$\int_0^T (I^n(t),v)_{L^2}\,\mathrm{d}t \leq \frac{1}{2}|D|T\overline{I} + \frac{1}{2}\int_0^T \|v\|_{L^2}^2\,\mathrm{d}t\,,$$

where $\overline{I}$ is a pointwise bound to the input current. The remaining integral in the right hand side of this inequality can be bounded by a-priori estimates as it holds that either $\|v\|_{L^2}\leq 1$ or $v=u_n$, in which case the pointwise bound $\overline{u}$ provides an a-priori uniform estimate. Finally, for part (c) we need to guarantee the additional convergence

$$\int_0^T \langle I^n(t),\varphi(t)\rangle_{H^1}\,\mathrm{d}t \to \int_0^T \langle I(t),\varphi(t)\rangle_{H^1}\,\mathrm{d}t$$

for all test functions φ. However this immediately follows from the convergence of I^n to I in $L^2((0,T),L^2)$.

Chapter 4

Limit Theorems for PDMPs

In the present chapter we establish limit theorems for sequences of PDMPs with values in a Hilbert space. In view of applications this part of the thesis is motivated by the interest in the derivation of a Langevin approximation to spatio-temporal stochastic hybrid models of excitable membranes. To this end we establish a law of large numbers and a martingale central limit theorem for certain martingales associated with the PDMPs. We further show how to represent the stochastic process arising as the limit in the central limit theorem as a solution of a stochastic partial differential equation (SPDE) which then yields a Langevin approximation for PDMPs by a system of SPDEs. The theoretical results of this chapter are of general interest to PDMP theory.

We briefly describe the general idea of our framework. We consider a family of fully coupled, Hilbert space-valued PDMPs in the sense of Section 3.0.1 indexed by n, $n \in \mathbb{N}$. Here fully coupled means that the rates of the jump part depend on the state of the full system and the continuous component depends on the state of the jump component. For the limit theorems we rely on two key assumptions. Firstly, we assume jumps with height decreasing to zero for $n \to \infty$ which occur at a high frequency roughly inversely proportional to the jump size. We are therefore in the *fluid limit* setting, cf. [79, 80]. Secondly, we assume that for each n the continuous dynamics in between jumps depends on the piecewise constant component only via *finitely* many (Hilbert space-valued) functions thereof, which we call *coordinate functions*. A precise definition of this notion follows in the next section. Typically for applications to biological excitable membranes the coordinates are the proportions of individual channels in a given state at given location of the axon. That is, for a simplified channel model when each channel can be either open or closed, the coordinate functions state the fractions of open and closed channels across the membrane. Further for a sequence of membrane models, taking the limit for $n \to \infty$ is understood as increasing the channel numbers to infinity (jump rate increases) and simultaneously reducing the single channel conductance (jump size decreases). We refer to Section 4.5 for a more detailed discussion of the interpretation of the limit theorems in the application to neuronal membranes.

The first limit theorem we present is a weak law of large numbers for PDMPs in infinite-dimensional Hilbert spaces where the deterministic limit is given by a solution of an abstract evolution equation. Next we proceed to the presentation of a central limit theorem for the martingales associated with a PDMP. This central limit theorem gives the basis for an approximation of PDMPs by diffusion processes which are solutions of stochastic partial differential equations. The new results presented extend previous results for PDMPs and pure jump processes in Euclidean space [80, 41, 96]. The main difficulties extending the fluid limit theorems in [79, 80, 96] to processes taking values in infinite-dimensional Hilbert spaces arise, on the one hand, in the appropriate treatment of Hilbert space-valued martingales. These arise by splitting a PDMP, being a semi-martingale, into a sum of a part with finite variation and a local martingale. As these considerations are essential we have devoted a full section, Section 4.1, to the discussion of the martingales. On the other hand, the more complicated existence theory of solutions to abstract evolution equations compared to solutions of ordinary differential equations in Euclidean space demands for additional technical rigour.

Next, we briefly comment on related work in the infinite-dimensional setting. For a model of linear chemical reactions by jump Markov processes a law of large numbers [6] and a central limit theorem [75] have been proven based on the original work of [79, 80] for finite-dimensional jump-processes. In these cases the deterministic limit is a reaction-diffusion partial differential equation and the central limit theorem yields diffusion processes given by stochastic partial differential equations. Limit theorems for variations of this model have been investigated in two series of studies, cf. [76, 77, 78] and [15, 16, 17, 18, 19]. A central difference between spatial models of excitable media to models of chemical reactions is that the latter exhibit diffusive motion of the reactants ($\sim$ channels) which is absent in the former. Additionally excitable media equations exhibit non-local interaction of channels as their dynamics are coupled via the membrane potential. The limit theorems we establish have to account for these differences. Further, there is also a difference on the technical side. The author in [75] and all subsequent work is based on the semigroup approach to stochastic / deterministic evolution equations. In contrast, we pursue in the present thesis the approach of a weak formulation. Finally, we also mention a central limit theorem for Hilbert-valued martingales in [91] and a diffusion approximation of SDEs on nuclear spaces driven by Poisson random measures in [71]. The methods of proof we employ for the theoretical results in this chapter are motivated by the two last references, but differ as the classes of stochastic processes considered therein and in the present thesis are different.

The remainder of the chapter is organised as follows. We first precisely state the assumptions for a sequence of PDMPs in Section 4.0.1 and then discuss in detail the associated martingale process in Section 4.1. Limit theorems and the diffusion approximation are presented in Sections 4.2 and 4.3. We have deferred the proofs of the main results to Section 4.4. Finally in Section 4.5 we discuss applications of these limit theorems to compartmental-type models of excitable membranes. The end of the chapter contains in Section 4.6 the proof of the technical Theorem 4.1.1 that guarantees the square-integrability of the associated Hilbert space valued martingales and establishes an appropriate Itô-isometry.

4.0.1 An appropriate sequence of PDMPs

Let $X \subset H \subset X^*$ be an evolution triple of separable, real Hilbert spaces and E be another separable, real Hilbert space. Further, for a certain $m \in \mathbb{N}$ (its significance is explained in the next paragraph) we denote by $\mathcal{H} = \times_{i=1}^m H$, $\mathcal{E} = \times_{i=1}^m E$ the direct products of the Hilbert spaces H and E which are Hilbert spaces themselves. Finally we set $\mathcal{E}^* = \times_{j=1}^m E^*$ which is the dual space to $\mathcal{E}$.

We now define the structure of the sequence of processes for which we derive the limit theorems. For all $n \in \mathbb{N}$ let $(\Omega^n, \mathcal{F}^n, (\mathcal{F}_t^n)_{t\geq 0}, \mathbb{P}^n)$ be a filtered probability space satisfying the usual conditions and the processes $(X_t^n)_{t\geq 0} = (Y_t^n, \theta_t^n)_{t\geq 0}$ defined thereon are regular standard PDMPs, cf. Definition 3.0.1, taking values in $H \times K_n$ with path properties as defined in Section 3.0.1. Correspondingly, the characteristics of the PDMPs are given by $(A^n, B^n, \Lambda^n, \mu^n)$, cf. the remark immediately following Definition 3.0.1 in Section 3.0.1. Note that the state space K_n for the piecewise constant component changes with varying index n whereas the state space H for the continuous component remains fixed. Therefore, in order for such a sequence of processes to allow for a limit we need to impose a special structure on the characteristics referring to the continuous component. To this end we assume there exists an $m \in \mathbb{N}$ introduced above, such that for each PDMP $(Y_t^n, \theta_t^n)_{t\geq 0}$ there exists a family of *coordinate functions* $z_i^n : K_n \to E$, $i = 1, \ldots, m$, such that the characteristics $A^n(\theta), B^n(\theta)$ depend on the piecewise constant component and on the index n only via the $\mathcal{E}$–valued *coordinate process* $z^n(\theta) = (z_1^n(\theta), \ldots, z_m^n(\theta))$. That is, there exists operators $A, B : \mathcal{E} \times X \to X^*$ such that for all $n \in \mathbb{N}$, all $u \in H$ and all $\theta \in K_n$

$$A^n(\theta)\, u \;=\; A(z^n(\theta))\, u, \quad B^n(\theta, u) \;=\; B(z^n(\theta), u). \tag{4.0.1}$$

The coordinates z^n can be interpreted as a 'sufficient statistic' of the piecewise constant component for the evolution of the continuous component. In statistics a sufficient statistic for a quantity of interest is a function of the observations that is sufficient to estimate this particular quantity. For example, the sample average of independently and identically distributed real random variables is a sufficient statistic for the mean of their distribution. In the present setting, this means that the coordinate functions contain all information about the vector θ that is needed to determine the continuous dynamics in between jumps. Further, the essence of the subsequent limit theorems is that the sequence of coordinate processes on the space $\mathcal{E}$ allows for a limit under certain conditions. As $\mathcal{E}$ is a Hilbert space itself we note that no generality would be lost if instead of the family of coordinate functions we assumed the existence of Hilbert space-valued functions z^n taking values in the same Hilbert space for each n. However, we decided to use this more detailed notation since in examples one usually encounters that it is a set of coordinate functions that encodes the information necessary for defining the dynamics of the continuous component.

Finally, we assume that the operators A, B satisfy a global one-sided Lipschitz-type condition in the sense that for all $T > 0$ there exists a constant $L_1 > 0$ such that for all $u, v \in L^2((0,T), X)$ and all $p, q \in L^2((0,T), \mathcal{E})$ it holds that

$$\int_0^T \langle A(q)\, v - A(p)\, u, v - u\rangle_X + \langle B(q,v) - B(p,u), v - u\rangle_X \,\mathrm{d}t \leq L_1 \int_0^T \|v-u\|_H^2 + \sum_{i=1}^m \|q_i - p_i\|_E^2 \,\mathrm{d}t\,. \quad (4.0.2)$$

Here we have omitted the argument t in the functions u, v, p and q.

Remark 4.0.3. In the proof of the law of large numbers, see Section 4.4.1, this Lipschitz condition is applied such that one pairing (v, q) refers to a path segment of the continuous component of a PDMP and the coordinate process and the second (u, p) to the deterministic limit functions. Thus for the application of (4.0.2) in the proof it is sufficient that it holds only for pairings (v, q) out of a set containing almost all paths of the sequence of PDMPs and (u, p) being the deterministic limit, i.e., one (!) distinguished pairing. This restriction of (4.0.2) to be satisfied only for particular pairings (v, q) and (u, p) out of the whole path space has a decisive advantage: We are able to incorporate additional qualitative results on the trajectories of the PDMPs and the deterministic limit in order establish (4.0.2). For example, in the case of hybrid neuron models such an additional qualitative result that allows to derive (4.0.2) in this application is that the components corresponding to u, v, p, q are pointwise bounded.

4.1 The associated martingale process

For the limit theorems we derive in this chapter, the central estimation concerns certain martingales associated with the PDMP. As these are of such central importance we discuss them in this separate section. The principle aim is, on the one hand, to derive conditions that imply the convergence in probability of the associated martingales as needed for the law of large numbers (cf. condition (4.2.4) in Theorem 4.2.1) and, on the other hand, we present some necessary structure for the central limit theorems. Therefore we define for all $j = 1, \dots, m$ the E-valued stochastic process M_j^n by

$$M_j^n(t) := z_j^n(\theta_t^n) - z_j^n(\theta_0^n) - \int_0^t \big[\mathcal{A}^n\langle\,\cdot\,, z_j^n(\cdot)\rangle_E\big](Y_s^n, \theta_s^n)\,\mathrm{d}s\,, \tag{4.1.1}$$

where the integrand in the right hand side is given by

$$\begin{aligned}\big[\mathcal{A}^n\langle\,\cdot\,, z_j^n(\cdot)\rangle_E\big](Y_s^n, \theta_s^n) &= \Lambda^n(Y_s^n, \theta_s^n)\int_{K_n}\Big(z_j^n(\xi) - z_j^n(\theta_s^n)\Big)\,\mu^n\big((Y_s^n, \theta_s^n), \mathrm{d}\xi\big)\\ &= \Lambda^n(Y_s^n, \theta_s^n)\sum_{\xi\in K_n}\Big(z_j^n(\xi) - z_j^n(\theta_s^n)\Big)\,\mu^n\big((Y_s^n, \theta_s^n), \{\xi\}\big)\,.\end{aligned}$$

Hence the integrand is a countable convex combination of elements in E with time-dependent coefficients. Moreover, in between jumps it depends continuously on s. Anticipating condition (4.1.3) below, which we generally assume to hold, we find that the integral in the right hand side of (4.1.1) almost surely exists in the sense of Bochner. For an application of a functional $\phi \in E^*$ to (4.1.1) we obtain

$$\langle\phi, M_j^n(t)\rangle_E = \langle\phi, z_j^n(\theta_t^n)\rangle_E - \langle\phi, z_j^n(\theta_0^n)\rangle_E - \int_0^t \big[\mathcal{A}^n\langle\phi, z_j^n(\cdot)\rangle_E\big](Y_s^n, \theta_s^n)\,\mathrm{d}s\,, \tag{4.1.2}$$

where the integrand is

$$\big[\mathcal{A}^n\langle\phi, z_j^n(\cdot)\rangle_E\big](Y_s^n, \theta_s^n) = \Lambda^n(Y_s^n, \theta_s^n)\int_{K_n}\langle\phi, z_j^n(\xi)\rangle_E - \langle\phi, z_j^n(\theta_s^n)\rangle_E\,\mu^n\big((Y_s^n, \theta_s^n), \mathrm{d}\xi\big)\,.$$

Thus the integral has the form of the extended generator, cf. Theorem 2.2.1, applied to the mapping $(u, \theta) \mapsto \langle\phi, z_j^n(\theta)\rangle_E$. This already suggests that the processes (4.1.2) are martingales under suitable boundedness conditions. In fact we are able to establish that the processes M_j^n are E–valued càdlàg martingales. We refer to [35, 100] for a brief discussion of martingales in infinite-dimensional spaces. The easiest way to validate the martingale property is due to the following result [100, Sec. 2.3]: If $\mathbb{E}^n\|M_j^n(t)\|_E < \infty$ for all $t \in [0, T]$, the Hilbert space-valued martingale property holds if and only if $\langle\phi, M_j^n(t)\rangle_E$ is a real-valued martingale for all $\phi \in E^*$. The following theorem gives a condition that guarantees that the processes (4.1.1) are square-integrable martingales and satisfy an Itô-isometry. The proof is rather technical and thus we have deferred it to the end of the chapter into Section 4.6.1.

Theorem 4.1.1. *Let $n \in \mathbb{N}$ be fixed and assume that for all $t > 0$ it holds that*

$$\mathbb{E} \int_0^t \Big[\Lambda^n(Y_s^n, \theta_s^n) \int_{K_n} \|z_j^n(\xi) - z_j^n(\theta_s^n)\|_E^2 \, \mu^n\big((Y_s^n, \theta_s^n), \mathrm{d}\xi\big)\Big] \, \mathrm{d}s \, < \infty \, . \tag{4.1.3}$$

Then the process M_j^n is a square-integrable martingale and satisfies the Itô-Isometry

$$\mathbb{E}^n \|M_j^n(t)\|_E^2 = \int_0^t \mathbb{E}^n \Big[\Lambda^n(Y_s^n, \theta_s^n) \int_{K_n} \|z_j^n(\xi) - z_j^n(\theta_s^n)\|_E^2 \, \mu^n\big((Y_s^n, \theta_s^n), \mathrm{d}\xi\big)\Big] \, \mathrm{d}s \, . \tag{4.1.4}$$

We continue the investigation of the processes M_j^n as Hilbert space valued martingales. From now on we always assume that assumption (4.1.3) holds. Note that the finiteness of the second moments of the jump sizes is a standard condition in related limit theorems [79, 96, 91]. We introduce a concept akin to the quadratic covariance operator in Euclidean spaces. This concept is important for the central limit theorems in, on the one hand, establishing weak convergence, and, on the other hand, characterising the limit. For further reference we refer to [90].

Definition 4.1.1. For the square-integrable, E–valued, càdlàg martingale M_j^n we denote by $(\ll M_j^n \gg_t)_{t \geq 0}$ its *quadratic variation process*, i.e., the unique (up to indistinguishability), predictable $L_1(E^*, E)$-valued[1] process which satisfies that for all $\phi, \psi \in E^*$ the real-valued process

$$t \mapsto \langle \phi, M_j^n(t) \rangle_E \, \langle \psi, M_j^n(t) \rangle_E - \langle \phi, \ll M_j^n \gg_t \psi \rangle_E \tag{4.1.5}$$

is a local martingale.

The aim now is to obtain an explicit formula for the quadratic variation process. To this end we define for $j = 1, \ldots, m$ the mappings

$$\psi \mapsto \; G_{jj}^n(u, \theta^n)\psi \; := \tag{4.1.6}$$

$$:= \; \Lambda^n(u, \theta^n) \int_{K_n} \langle \psi, z_j^n(\xi) - z_j^n(\theta^n) \rangle_E \left(z_j^n(\xi) - z_j^n(\theta^n) \right) \mu^n\big((u, \theta^n), \mathrm{d}\xi\big) \, .$$

These are linear, bounded operators mapping $E^* \to E$ and depend measurably on $(u, \theta^n) \in H \times K_n$. Each operator is *non-negative*, i.e., $\langle \phi, G_{jj}^n(u, \theta^n)\phi \rangle_E \geq 0$ for all $\phi \in E^*$, and *symmetric*, i.e., $\langle \psi, G_{jj}^n(u, \theta^n)\phi \rangle_E = \langle \phi, G_{jj}^n(u, \theta^n)\psi \rangle_E$ for all $\phi, \psi \in E^*$. Let $(\varphi_k)_{k \in \mathbb{N}}$ denote an orthonormal basis in E^*. Then we find due to the Riesz Rep-

[1] $L_1(E^*, E)$ denotes the space of trace class operators from the Hilbert space E^* into E.

resentation Theorem and Parseval's identity that the trace of each operator satisfies

$$\begin{aligned}\operatorname{Tr} G^n_{jj}(u,\theta^n) &= \Lambda^n(u,\theta^n)\int_{K_n}\sum_{k\in\mathbb{N}}\Big(\langle\varphi_k, z^n_j(\xi)-z^n_j(\theta^n)\rangle_E\Big)^2\mu^n\big((u,\theta^n),\mathrm{d}\xi\big)\\ &= \Lambda^n(u,\theta^n)\int_{K_n}\big\|z^n_j(\xi)-z^n_j(\theta^n)\big\|^2_E\,\mu^n\big((u,\theta^n),\mathrm{d}\xi\big)\,. \qquad (4.1.7)\end{aligned}$$

Proposition 4.1.1. *The quadratic variation of the martingale M^n_j satisfies*

$$\ll M^n_j\gg_t \;=\; \int_0^t G^n_{jj}(Y^n_s,\theta^n_s)\,\mathrm{d}s\,. \qquad (4.1.8)$$

Proof. First of all note that due to the characterisation of the trace (4.1.7) and condition (4.1.3) it holds that the process (4.1.8) takes values in $L_1(E^*,E)$ almost surely. Further, it holds that $\ll M^n_j\gg_t$ satisfies for all $\phi,\psi\in E$ that

$$\langle\phi,\ll M^n_j\gg_t\psi\rangle_E \;=$$

$$= \int_0^t \Lambda^n(Y^n_s,\theta^n_s)\int_{K_n}\langle\psi, z^n_j(\xi)-z^n_j(\theta^n_s)\rangle_E\,\langle\phi, z^n_j(\xi)-z^n_j(\theta^n_s)\rangle_E\,\mu^n\big((Y^n_s,\theta^n_s),\mathrm{d}\xi\big)\,\mathrm{d}s$$

as the right hand side is, due to Proposition B.2.5, the unique real-valued process such that the product $\langle\phi, M^n_j(t)\rangle_E\,\langle\psi, M^n_j(t)\rangle_E$ is a local martingale. Here $\langle\phi, M^n_j(t)\rangle_E$ and $\langle\psi, M^n_j(t)\rangle_E$ are understood as real-valued stochastic integrals with respect to the associated martingale measure of a PDMP. Thus we infer that for all $\phi,\psi\in E$ it holds

$$\langle\phi,\ll M^n_j\gg_t\psi\rangle_E \;=\; \int_0^t\langle\phi, G^n_{jj}(Y^n_s,\theta^n_s)\psi\rangle_E\,\mathrm{d}s\,.$$

Finally, the linearity of the Bochner integral (note that $L_1(E^*,E)$ is a Banach space) implies

$$\ll M^n_j\gg_t \;=\; \int_0^t G^n_{jj}(Y^n_s,\theta^n_s)\,\mathrm{d}s\,.$$

□

Further, a second property of the quadratic variation is that the process

$$t\mapsto \|M^n_j(t)\|^2_E - \operatorname{Tr}\ll M^n_j\gg_t$$

is a local martingale. We note that the *trace process* $t\mapsto \operatorname{Tr}\ll M^n_j\gg_t$ is the unique, predictable increasing process exhibiting this property. Using the characterisation (4.1.8) of the quadratic variation we thus obtain that the process

$$t\mapsto \|M^n_j(t)\|^2_E - \operatorname{Tr}\Big(\int_0^t G^n_{jj}(Y^n_s,\theta^n_s)\,\mathrm{d}s\Big) \;=\; \|M^n_j(t)\|^2_E - \int_0^t \operatorname{Tr} G^n_{jj}(Y^n_s,\theta^n_s)\,\mathrm{d}s \quad (4.1.9)$$

is a local martingale vanishing almost surely at $t = 0$.

We are now in a position to state a lemma which establishes the convergence in probability (4.1.11) of the processes $(M_j^n)_{t\geq 0}$ necessary for the law of large numbers, cf. condition (4.2.4) in Theorem 4.2.1.

Lemma 4.1.1. *Assume that for all $T > 0$*

$$\lim_{n\to\infty} \mathbb{E}^n \int_0^T \Big[\Lambda^n(Y_s^n, \theta_s^n) \int_{K_n} \|z_j^n(\xi) - z_j^n(\theta_s^n)\|_E^2 \, \mu^n\big((Y_s^n, \theta_s^n), \mathrm{d}\xi\big)\Big] \, \mathrm{d}s \, = \, 0 \, . \quad (4.1.10)$$

Then the process (4.1.9) *is a martingale and for all $T, \epsilon > 0$, it holds that*

$$\lim_{n\to\infty} \mathbb{P}^n \big[\sup{}_{t\in[0,T]} \|M_j^n(t)\|_E > \epsilon\big] = 0 \, . \quad (4.1.11)$$

Proof. As the process M_j^n is an E-valued càdlàg martingale, it holds that $\|M_j^n\|_E^2$ is a càdlàg submartingale. Thus an application of Markov's and Doob's inequalities yield the estimates

$$\mathbb{P}^n\Big[\sup{}_{t\in[0,T]} \, \|M_j^n(t)\|_E^2 > \epsilon\Big] \, \leq \, \frac{1}{\epsilon} \, \mathbb{E}^n \big[\sup{}_{t\in[0,T]} \, \|M_j^n(t)\|_E^2\big] \, \leq \, \frac{4}{\epsilon} \, \mathbb{E}^n \|M_j^n(T)\|_E^2 \, .$$

Now, the Itô-isometry (4.1.4) and condition (4.1.10) imply the convergence in probability (4.1.11).

It remains to show that the process (4.1.9) is a martingale. A sufficient condition[2] is that for all $T > 0$ it holds

$$\mathbb{E}^n \Big[\sup{}_{t\in[0,T]} \Big| \|M_j^n(t)\|_E^2 - \int_0^t \operatorname{Tr} G_{jj}^n(Y_s^n, \theta_s^n) \, \mathrm{d}s \Big|\Big] < \infty \, . \quad (4.1.12)$$

Estimating the term inside the expectation we obtain

$$\begin{aligned} &\sup_{t\leq T} \Big| \|M_j^n(t)\|_E^2 - \int_0^t \operatorname{Tr} G_{jj}^n(Y_s^n, \theta_s^n) \, \mathrm{d}s \Big| \\ &\quad \leq \, \sup_{t\leq T} \|M_j^n(t)\|_E^2 + \sup_{t\leq T} \int_0^t \Lambda^n(Y_s^n, \theta_s^n) \int_{K_n} \big\|z_j^n(\xi) - z_j^n(\theta_s^n)\big\|_E^2 \, \mu^n\big((Y_s^n, \theta_s^n), \mathrm{d}\xi\big) \, \mathrm{d}s \, . \end{aligned}$$

The expectation of the first supremum term in the right hand side is bounded due to Doob's inequality and the square-integrability of the martingale. The term inside the second supremum is increasing, thus its expectation is finite due to condition (4.1.10). ∎

To conclude this section we briefly discuss the vector-valued process of all martingales

[2] A sufficient condition for a real, càdlàg local martingale $(M_t)_{t\geq 0}$ to be a martingale is $\mathbb{E} \sup_{s\leq t} |M_s| < \infty$ for all $t \geq 0$, see, e.g., [64, p. 308].

M_i^n, i.e., the $\mathcal{E}$–valued process

$$t \mapsto M^n(t) = \big(M_1^n(t), \ldots, M_m^n(t)\big),$$

and, in particular, its quadratic variation. To this end we define analogously to (4.1.6) for all $i, j = 1, \ldots, m$, $i \neq j$, operators $G_{ij}^n(u, \theta^n) \in L(E^*, E)$ by

$$\psi \mapsto G_{ij}^n(u,\theta^n)\psi := \Lambda^n(u,\theta^n) \int_{K_n} \Big(\langle \psi, z_i^n(\xi) - z_i^n(\theta^n)\rangle_E\Big) \Big(z_j^n(\xi) - z_j^n(\theta^n)\Big)\, \mu^n\big((u,\theta^n), \mathrm{d}\xi\big)\,. \tag{4.1.13}$$

If the operators (4.1.6) are of trace class, then each of these operators is of trace class, as it follows from Young's inequality that

$$\mathrm{Tr}\, G_{ij}^n(u,\theta^n) \leq \frac{1}{2}\mathrm{Tr}\, G_{ii}^n(u,\theta^n) + \frac{1}{2}\mathrm{Tr}\, G_{jj}^n(u,\theta^n)\,.$$

Let $\Phi = (\phi_1, \ldots, \phi_m)$ and $\Psi = (\psi_1, \ldots, \psi_m)$ be elements of $\mathcal{E}^*$. Summing over all operators (4.1.6) and (4.1.13) applied to the components of Φ, Ψ as indicated by the indices, i.e.,

$$\langle \Phi, G^n(u,\theta^n)\, \Psi\rangle_{\mathcal{E}} := \sum_{i,j=1}^m \langle \phi_i, G_{ij}^n(u,\theta^n)\, \psi_j\rangle_E, \tag{4.1.14}$$

we obtain a linear, bounded operator $G^n(u,\theta^n)$ mapping $\mathcal{E}^*$ to $\mathcal{E}$. This operator is symmetric as the family of operators (4.1.13) satisfy $\langle \phi, G_{ij}^n(u,\theta^n)\psi\rangle_E = \langle \psi, G_{ji}^n(u,\theta^n)\phi\rangle_E$ for all i, j. Moreover, the operator $G^n(u,\theta^n)$ is non-negative as it holds that

$$\langle \Psi, G^n(u,\theta^n)\, \Psi\rangle_{\mathcal{E}} \;=\; \Lambda^n(u,\theta^n) \int_{K_n} \Big(\sum_{i=1}^m \big\langle \psi_i, z_i^n(\xi) - z_i^n(\theta^n)\big\rangle_E\Big)^2 \mu^n\big((u,\theta^n), \mathrm{d}\xi\big)\,.$$

Finally, the operator $G^n(u,\theta)$ is of trace class if the operators (4.1.6) are of trace class, as a simple estimate on its trace in terms of the trace of its diagonal elements (4.1.6) is given by

$$\mathrm{Tr}\, G^n(u,\theta^n) \leq m \sum_{i=1}^m \mathrm{Tr}\, G_{ii}^n(u,\theta^n),$$

and further, the trace satisfies

$$\mathrm{Tr}\, G^n(u,\theta^n) = \Lambda^n(u,\theta^n) \int_{K_n} \|z^n(\xi) - z^n(\theta)\|_{\mathcal{E}}^2\, \mu^n\big((u,\theta^n), \mathrm{d}\xi\big)\,. \tag{4.1.15}$$

4.2 A weak law of large numbers

In order to propose a deterministic limit for the sequence of PDMPs we consider functions $F_j : \mathcal{E} \times H \to E$ for all $j = 1, \ldots, m$. In combination with the operators A, B these functions are used to defined a coupled system of deterministic abstract evolution equations

$$\begin{aligned} \dot{u} &= A(p)\, u + B(p,u), \\ \dot{p}_j &= F_j(p,u), \qquad\qquad j = 1, \ldots, m\,. \end{aligned} \tag{4.2.1}$$

We assume that to suitable initial condition $(u_0, p_0) \in H \times \mathcal{E}$ there exists a unique weak solution $(u(t), p(t))_{t \geq 0}$ in $C^0(\mathbb{R}_+, H \times \mathcal{E})$ of (4.2.1). Particularly, we assume that for all $j = 1, \ldots, m$ the components p_j satisfy

$$\langle \phi, p_j(t) \rangle_E = \langle \phi, p_j(0) \rangle_E + \int_0^t \langle \phi, F_j(p(s), u(s)) \rangle_E \, \mathrm{d}s \qquad \forall\, t \in [0,T],\ \phi \in E^*\,. \tag{4.2.2}$$

Finally, we assume that the functions F_j, $j = 1, \ldots, m$, satisfy a global Lipschitz condition on $L^2((0,T), \mathcal{E} \times X)$, i.e., for every $T > 0$ there exists a constant L_2 such that for all $v, u \in L^2((0,T), X)$ and all $p, q \in L^2((0,T), \mathcal{E})$ it holds that

$$\Big(\int_0^T \| F_j(q,v) - F_j(p,u) \|_E \, \mathrm{d}t \Big)^2 \leq L_2 \int_0^T \| v - u \|_H^2 + \sum_{i=1}^m \| q_i - p_i \|_E^2 \, \mathrm{d}t \tag{4.2.3}$$

where we have omitted the arguments t of the functions u, v, p and q. Note it is only necessary for the Lipschitz condition to hold on a certain subset of the path space, cf. Remark 4.0.3.

We now present a weak law of large numbers in Theorem 4.2.1 below. The proof of the theorem follows the lines of previously published limit theorems considering processes in finite dimensions [79, 96]. The main difficulties arising in infinite-dimensional phase space concerns the bounds on the martingale part, cf. condition (C1), which is rarely a problem in finite dimensions. However, using the appropriate martingale theory in Hilbert spaces these can be kept to a minimum. Then the difficulties are mainly of a technical nature as martingale theory in connection with PDMPs in infinite-dimensional spaces gets more involved and is not covered by previous results, which are reviewed in Section B.2.4. We have established the necessary theory in the preceding section and addressed the question of the convergence of the martingale part (C1) within this framework. Most importantly, Lemma 4.1.1 states a sufficient condition for (C1) to be satisfied. In particular, the sufficient condition (4.1.10) we have proven is a natural extension of the condition employed in finite dimensions, cf. [79, 96].

A different approach to establishing condition (C1) which avoids using martingale theory in Hilbert spaces is exemplified in the law of large numbers proved in [7]. In infinite-dimensional space this approach encounters the problem of simultaneously controlling countably many real martingales compared to only finitely many in the case of its finite-dimensional counterpart. This problem can be overcome with an intricate compactness argument which relies on the assumption that the dual space E^* is compactly embedded in some additional normed space and all estimates – especially an estimate which also implies condition (4.1.10) – have to be derived in the norm of this additional space. Furthermore, the condition, that all martingales $(\langle\phi, M_j^n(t)\rangle_E)_{t\geq 0}$, $j = 1,\dots,M$ and $\phi\in E^*$, possess almost surely uniformly bounded paths, has to be introduced. We are of the opinion that our approach is more elegant, but, more importantly, it avoids the introduction of additional conditions.
Finally, consistently with the notation in Section 4.1 we use in the subsequent theorem and its proof the notation

$$\big[\mathcal{A}^n\langle\phi, z_j^n(\cdot)\rangle_E\big](Y_s^n,\theta_s^n) = \Lambda^n(Y_s^n,\theta_s^n)\int_{K_n}\langle\phi, z_j^n(\xi)\rangle_E - \langle\phi, z_j^n(\theta_s^n)\rangle_E\,\mu^n\big((Y_s^n,\theta_s^n),\mathrm{d}\xi\big).$$

Next, for given $(u,\theta)\in H\times K_n$ functionals $\big[\mathcal{A}^n\langle\,\cdot\,, z_j^n(\cdot)\rangle_E\big](u,\theta)$ on E^* are defined by the mappings $\phi\mapsto\big[\mathcal{A}^n\langle\phi, z_j^n(\cdot)\rangle_E\big](u,\theta)$. As usual we identify the bidual E^{**} with E and thus $\big[\mathcal{A}^n\langle\,\cdot\,, z_j^n(\cdot)\rangle_E\big](u,\theta)\in E$. Furthermore, $D([0,T],K_n)$ denotes the space of càdlàg functions on $[0,T]$ taking values in K_n. Then we characterise for all $n\geq 0$ and all $T>0$ a set $\mathcal{S}(n,T)\subset C^0([0,T],H)\times D([0,T],K_n)$ by the property that it satisfies $\mathbb{P}^n\big[(Y_t^n,\theta_t^n)_{t\in[0,T]}\in\mathcal{S}(n,T)\big]=1$. Clearly, all such sets differ only by a set of measure zero and the smallest is called the *support* of $(Y_t^n,\theta_t^n)_{t\in[0,T]}$.

Theorem 4.2.1. *We assume that the following conditions hold:*

(C1) *For all $j=1,\dots,m$ it holds that for fixed $T,\epsilon>0$*

$$\lim_{n\to\infty}\mathbb{P}^n\big[\sup\nolimits_{t\in[0,T]}\|M_j^n(t)\|_E>\epsilon\big]=0\,. \tag{4.2.4}$$

(C2) *The functions F_j satisfy for all $j=1,\dots,m$*

$$\lim_{n\to\infty}\sup_{\mathcal{S}(n,T)}\int_0^T\big\|\big[\mathcal{A}^n\langle\,\cdot\,, z_j^n(\cdot)\rangle_E\big](u,\theta)-F_j(z^n(\theta),u)\big\|_E^2\,\mathrm{d}t=0\,, \tag{4.2.5}$$

where we have omitted the argument t of the functions u and θ.

(C3) *The initial conditions (Y_0^n,θ_0^n) of the sequence of PDMPs converge in probability* *to the initial conditions of the deterministic limit in the sense that for all* $\epsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}^n\Big[\|Y_0^n - u_0\|_H + \sum_{j=1}^m \|z_j^n(\theta_0^n) - p_j(0)\|_E > \epsilon\Big] = 0\,.$$

Then, for every $\epsilon > 0$ *and every fixed* $T > 0$ *it holds that*

$$\lim_{n\to\infty} \mathbb{P}^n\Big[\sup_{t\in[0,T]}\Big(\|Y_t^n - u(t)\|_H^2 + \sum_{j=1}^m \|z_j^n(\theta_t^n) - p_j(t)\|_E^2\Big) \;>\; \epsilon\Big] = 0\,. \qquad (4.2.6)$$

Remark 4.2.1. The result (4.2.6) implies convergence in probability of the processes $(Y_t^n, z^n(\theta_t^n))_{t\geq 0}$ to the deterministic function $(u(t), p(t))_{t\in[0,T]}$ in the Hilbert space $L^2((0,T), H \times \mathcal{E})$. If the differences of the components are almost surely bounded independent of n the convergence even holds in the mean, cf. the application of the law of large numbers in Theorem 4.5.1.
Further, the conditions (C1)–(C3) are generalisations from Euclidean space to infinite-dimensional Hilbert spaces of those employed in the corresponding theorems for PDMPs in Euclidean space [96] and, in particular, of the original formulation in case of pure jump processes in Euclidean space [79]. In these cases the conditions above reduce to the corresponding assumptions.

4.3 The central limit theorem and the Langevin approximation

We proceed to the presentation of the central limit theorem for associated martingales $(M_t^n)_{t\geq 0}$ defined in (4.1.1). The central limit theorem provides the theoretical basis for an approximation of spatio-temporal PDMPs by Hilbert-space valued diffusion processes where the latter can be represented by solutions of stochastic partial differential equations.

Proving central limit theorems usually involves two tasks: On the one hand, on has to show the existence of a limit and, on the other hand, one has to provide a characterisation of the limit as a certain stochastic process. The former is equivalent to the problem of tightness of the sequence of random variables, cf. [41, Chap. 3, Thms. 2.2,7.8]. For $\mathcal{E}$-valued càdlàg processes this means tightness in the space of càdlàg functions on $\mathcal{E}$, denoted by $D(\mathbb{R}_+, \mathcal{E})$. We note that *necessary and sufficient* conditions for tightness in $D(\mathbb{R}_+, \mathcal{E})$ are that the sequence of E-valued random variables obtained from the processes for t fixed is tight for all t out of a dense subset of $\mathbb{R}_+$ and the convergence in probability of the modulus of continuity, cf., e.g., [69, Sec. 2.1.3] or [41, Chap. 3, Thm. 7.2]. Usually in applications there are sufficient conditions which are easier to validate. Particularly in case of tightness of martingales the author in [91] states sufficient conditions depending on the quadratic variation process. Secondly, in order to characterise the limit there exist different approaches, showing either that the limit solves a given martingale or local martingale problem which is known to have a unique solution (cf. [71, 91]) or proving weak convergence of the finite dimensional distributions, which, due to Levy's Continuity Theorem, is implied by the convergence of the characteristic functions (cf. [75, 96]). We present two central limit theorems employing the two different methods. The two theorems differ in additional technical assumptions to the central condition of the convergence of the quadratic variations. These different conditions may be suitable for different applications and moreover the two theorems show that the extension to infinite-dimensioanl spaces we present are compatible with both approaches to establish central limit theorems.

Finally, we emphasise that in the following the space $\mathcal{E}$ need not necessarily be the same space for which the law of large numbers is satisfied. However, clearly, the space $\mathcal{E}$ in the present section contains the space in the law of large numbers as subspace. In applications, usually, the law of large numbers holds in a space with a stronger norm, for example, for the neuron model considered in Section 4.5 the law of large numbers holds in $L^2(D)$ whereas the central limit theorem holds in the distributions space $H^{-2s}(D)$.[3] This is a major difference to the corresponding results in finite-dimensional space where both limit theorems hold in the same space.[4]

4.3.1 A martingale central limit theorem

In this section we present a central limit theorem for the scaled $\mathcal{E}$–valued martingales $(\sqrt{\alpha_n}\, M_t^n)_{t\geq 0}$ associated with a sequence of PDMPs where $\alpha_n \in \mathbb{R}_+$, $n \in \mathbb{N}$, is a suitable rescaling sequence. In the following let $t \mapsto G(u(t), p(t)) \in L(\mathcal{E}^*, \mathcal{E})$ be a Bochner-integrable operator-valued map such that each $G(u(t), p(t))$ is a symmetric, positive trace class operator. Particularly this implies for all $\Phi \in \mathcal{E}^*$ and all $t > 0$, that it holds that

$$\int_0^t \big\langle \Phi, G(u(s), p(s))\, \Phi \big\rangle_{\mathcal{E}}\, \mathrm{d}s < \infty\,. \tag{4.3.1}$$

Here $(u(t), p(t))_{t\geq 0}$ is the deterministic limit obtained in Theorem 4.2.1 and the use of this notation for the – at this point – arbitrary time-dependent operator G only illustrates that in applications the time-dependence is due to a dependence on the deterministic limit system. The operator-valued map G is used to define a unique centred diffusion process on $\mathcal{E}$, i.e., an $\mathcal{E}$–valued Gaussian process with independent increments, continuous sample paths and zero expectation. In general a centred Gaus-

[3]Here $H^{-2s}(D)$ is the dual space to the Sobolev space $H^{2s}(D)$ where $D \subset \mathbb{R}^d$ and $s > d/2$.

[4]Note also that in finite-dimensional spaces all norms, and hence also all norms on subspaces, are equivalent which does not hold in the case of an infinite-dimensional Hilbert space.

sian process is uniquely defined by its covariance operator and due to a theorem of Itô stated in [75] every family of trace class operators $C^*(t) \in L_1(\mathcal{E},\mathcal{E})$ which are increasing and continuous in t define a centred diffusion process. In the present situation we define C^* in the following way. We denote by $\iota : \mathcal{E} \to \mathcal{E}^*$ the canonical identification of a Hilbert space with its dual, hence we can define for $x, y \in \mathcal{E}$

$$\big(x, C^*(t)\, y\big)_{\mathcal{E}} = \int_0^t \big\langle \iota(x), G(u(s), p(s))\, \iota(y)\big\rangle_{\mathcal{E}}\, \mathrm{d}s\,,$$

which is continuous and increasing for all $x \in \mathcal{E}$ and $C^*(t)$ is a trace class operator on $\mathcal{E}$. Moreover, for operators $C(t) \in L_1(\mathcal{E}^*, \mathcal{E})$, defined by

$$\big\langle \Phi, C(t)\, \Psi\big\rangle_{\mathcal{E}} = \int_0^t \big\langle \Phi, (G(u(s), p(s))\, \Phi\big\rangle_{\mathcal{E}}\, \mathrm{d}s\,, \tag{4.3.2}$$

there is obviously a one-to-one relationship between C^* and C. Hence, we may say that also the latter defines a diffusion process on the space $\mathcal{E}$.

We proceed to the statement of the central limit theorem. The proof of the theorem employs a characterisation of the limit via the local martingale problem. The condition which is essential for the convergence is the convergence of the quadratic variation processes (4.3.6). The second condition (4.3.7) is a technical condition on the jump heights which arises due to the method of proof and is usually satisfied in applications. The remaining conditions are such that (D1) guarantees tightness of the sequence of processes and in combination with (D2) that any limit is a continuous stochastic process. The proof of the following theorem is deferred to Section 4.4.2. Therein we also comment on the differences in the techniques employed compared to previous results.

Theorem 4.3.1. *We assume that the following conditions hold:*

(D1) *For all $t > 0$ it holds that*

$$\sup_{n\in\mathbb{N}} \alpha_n\, \mathbb{E}^n \int_0^t \Big[\Lambda^n(Y^n_s, \theta^n_s) \int_{K_n} \|z^n(\xi) - z^n(\theta^n_s)\|^2_{\mathcal{E}}\, \mu^n\big((Y^n_s, \theta^n_s), \mathrm{d}\xi\big)\, \mathrm{d}s\Big] < \infty\,, \tag{4.3.3}$$

and there exists an orthonormal basis $(\varphi_k)_{k\in\mathbb{N}}$ of $\mathcal{E}^*$ such that for all $k \in \mathbb{N}$

$$\alpha_n\, \mathbb{E}^n \int_0^t \langle \varphi_k, G^n(Y^n_s, \theta^n_s)\varphi_k\rangle_{\mathcal{E}}\, \mathrm{d}s \;\leq\; \gamma_k\, C\,, \tag{4.3.4}$$

where the constants $\gamma_k > 0$, independent of n and t, satisfy $\sum_{k\in\mathbb{N}} \gamma_k < \infty$, and the constant $C > 0$ is independent of n and k but may depend on t.

(D2) *For all $\beta > 0$ and every $\Phi \in \mathcal{E}^*$ it holds that*

$$\lim_{n\to\infty} \mathbb{E}^n\Big[\int_0^t \Lambda^n(Y_s^n,\theta_s^n)\int_{\sqrt{\alpha_n}\,|\langle\Phi,z^n(\xi)-z^n(\theta_s^n)\rangle_{\mathcal{E}}|>\beta} \mu^n\big((Y_s^n,\theta_s^n),\mathrm{d}\xi\big)\,\mathrm{d}s\Big] \;=\; 0\,. \tag{4.3.5}$$

(D3) *Further, for all $\Phi \in \mathcal{E}^*$ and all $t > 0$ it holds that*

$$\lim_{n\to\infty}\int_0^t \mathbb{E}^n\big|\big\langle\Phi, G(u(s),p(s))\,\Phi\big\rangle_{\mathcal{E}} - \alpha_n\big\langle\Phi, G^n(Y_s^n,\theta_s^n)\,\Phi\big\rangle_{\mathcal{E}}\big|\,\mathrm{d}s \;=\; 0\,. \tag{4.3.6}$$

Finally, we assume that the jump heights of the rescaled martingales are almost surely uniformly bounded, i.e., there exists a constant $C < \infty$ such that it holds almost surely for all $n \in \mathbb{N}$ that

$$\sup_{t\geq 0}\,\sqrt{\alpha_n}\,\|z^n(\theta_t^n) - z^n(\theta_{t-}^n)\|_{\mathcal{E}} < C\,. \tag{4.3.7}$$

Then it follows that the process $(\sqrt{\alpha_n}\,M_t^n)_{t\geq 0}$ converges weakly to an $\mathcal{E}$–valued centred diffusion process characterised by the covariance operator (4.3.2).

Below we state a second version of the martingale central limit theorem wherein the limiting process is characterised by the convergence of the characteristic functions of the one-dimensional marginals. The central condition of the convergence of the quadratic variation processes (4.3.6) is unchanged, however, the second, technical condition (4.3.7) in (D3) is changed due to the different method of proof. That is, condition (4.3.8) arises instead of (4.3.7) as an assumption on the distribution of the jump heights employing a characterisation of the limit process using convergence of characteristic functions instead of the local martingale problem. The significance for applications of condition (4.3.8) in contrast to (4.3.7) is that the former avoids the almost sure uniform bound on the jump heights in the latter. That is, arbitrarily large jumps are possible for each martingale in the sequence as long as their probability decreases sufficiently fast. Note that (4.3.8) is stronger than the similar condition (D2) in the preceding theorem. The proof of the following theorem is deferred to Section 4.4.3.

Theorem 4.3.2. *Assume that the laws of the martingales $(\sqrt{\alpha_n}\,M_t^n)_{t\geq 0}$ form a tight sequence, e.g., condition* (D1) *is satisfied, and* (4.3.3) *holds.*

(D3') *The convergence* (4.3.6) *holds and there exists a sequence $\beta_n > 0$ decreasing to* *zero such that for all* $\Phi \in \mathcal{E}^*$

$$\lim_{n\to\infty} \alpha_n \, \mathbb{E}^n \Big[\int_0^t \Lambda^n(Y_s^n, \theta_s^n) \int_{\sqrt{\alpha_n}\,|\langle \Phi, z^n(\xi) - z^n(\theta_s^n)\rangle_{\mathcal{E}}| > \beta_n} \big|\langle \Phi, z^n(\xi) - z^n(\theta_s^n)\rangle_{\mathcal{E}}\big|^2 \mu^n\big((Y_s^n, \theta_s^n), \mathrm{d}\xi\big)\, \mathrm{d}s \Big] = 0 \,. \tag{4.3.8}$$

Then it follows that the process $(\sqrt{\alpha_n}\, M_t^n)_{t\geq 0}$ *converges weakly to an* $\mathcal{E}$*–valued centred diffusion process characterised by the covariance operator* (4.3.2)*.*

4.3.2 Langevin approximation

We have discussed in Section 4.0.1 that the coordinate functions z_i^n, $i = 1, \ldots, m$, can be considered sufficient statistics, i.e., they carry all the information needed for the dynamics of the continuous component $(Y_t^n)_{t\geq 0}$. Usually, e.g., in models of excitable membranes, one is ultimately interested in the dynamics of the continuous component. Therefore, the knowledge of the coordinate process $(z^n(\theta_t^n))_{t\geq 0}$, or a close approximation thereof, is sufficient. From this point of view the significance of the martingale central limit theorem and the law of large numbers is that they provide a justification of an approximation of the processes $(Y_t^n, z^n(\theta_t^n))_{t\geq 0}$ for large enough n by a diffusion process.

To this end we first discuss representations of the limiting diffusion process in the martingale central limit theorem as a stochastic integral. By definition $G(u(s), p(s))$ is a non-negative, self-adjoint trace class operator acting on $\mathcal{E}$, hence there exists a unique non-negative square root, i.e., a non-negative operator $\sqrt{G(u(s), p(s))}$ such that $G(u(s), p(s)) = \sqrt{G(u(s), p(s))} \circ \sqrt{G(u(s), p(s))}$ where $\circ$ denotes the composition of operators. Let $(W_t)_{t\geq 0}$ be a cylindrical Wiener process on $\mathcal{E}$ with covariance operator given by the identity (cf. [35, 100]). Then, as

$$\mathbb{E} \int_0^t \mathrm{Tr}\big(\sqrt{G(u(s), p(s))}\sqrt{I}\big)\big(\sqrt{G(u(s), p(s))}\sqrt{I}\big)^* \,\mathrm{d}s = \int_0^t \mathrm{Tr}\, G(u(s), p(s))\,\mathrm{d}s < \infty \,,$$

the mapping $t \mapsto \sqrt{G(u(s), p(s))}$ is a valid integrand process for a stochastic integral with respect to $(W_t)_{t\geq 0}$. That is, the process $(Z_t)_{t\geq 0}$ defined for all $t \geq 0$ by

$$Z_t := \int_0^t \sqrt{G(u(s), p(s))}\, \mathrm{d}W_s \tag{4.3.9}$$

is an $\mathcal{E}$–valued Gaussian process with continuous sample paths and independent increments which, in addition, is also a square-integrable martingale. Moreover, the process has the covariance given by the operator $\int_0^t G(u(s),p(s))\,\mathrm{d}s$. Therefore, due to unique definition of Gaussian processes via their covariance operators, the process $(Z_t)_{t\geq 0}$ coincides in distribution with the limiting diffusion identified in Theorems 4.3.2 and 4.3.1 for the sequence of martingales $(\sqrt{\alpha_n}\,M_t^n)_{t\geq 0}$.

The *Langevin approximation* $(\widetilde{Y}_t^n,\widetilde{P}_t^n)_{t\geq 0}$ of $(Y_t^n, z^n(\theta_t^n))_{t\geq 0}$ is given by the solution of the system of stochastic partial differential equations

$$\begin{aligned}\mathrm{d}\widetilde{Y}_t^n &= \big(A(\widetilde{P}_t^n)\,\widetilde{Y}_t^n + B(\widetilde{P}_t^n,\widetilde{Y}_t^n)\big)\,\mathrm{d}t\\ \mathrm{d}\widetilde{P}_t^n &= F(\widetilde{P}_t^n,\widetilde{Y}_t^n)\,\mathrm{d}t + \tfrac{1}{\sqrt{\alpha_n}}\sqrt{G(\widetilde{Y}_t^n,\widetilde{P}_t^n)}\,\mathrm{d}W_t\,.\end{aligned} \tag{4.3.10}$$

In order to analyse properties of the Langevin approximation, clearly, well-posedness of the system (4.3.10) has to be addressed first. This is suitably done within the *variational approach* to stochastic partial differential equations. It reflects the approach of using weak solution to abstract evolution equations defining the deterministic inter-jump motion of PDMPs generally taken in this thesis. We refer to [84, Sec. 1.3.1] for a concise introduction to the variational approach to SPDEs containing an existence and uniqueness theorem as well as further references. We do not pursue the issue of well-posedness of the Langevin approximation any further at this point, as we are of the opinion that this question is best addressed when analysing the Langevin approximation for particular models.

Finally, we remark that the sequence of Langevin approximations $(\widetilde{Y}_t^n,\widetilde{P}_t^n)_{t\geq 0}$ possesses the same asymptotic behaviour as the sequence of processes $(Y_t^n, z^n(\theta_n))_{t\geq 0}$ as discussed in the preceding sections. Firstly, it is obvious that for $n\to\infty$ and thus $\alpha_n\to\infty$ the noise term in (4.3.10) vanishes and the system approximates the deterministic solution $(u(t),p(t))_{t\geq 0}$ of the system (4.2.1), just as was proven in the law of large numbers Theorem 4.2.1 for the sequence of PDMPs. It poses no difficulties to make this statement precise in the form of a law of large numbers analogous to Theorem 4.2.1.

Remark 4.3.1. The process (4.3.9) is not necessarily the only stochastic integral process which coincides with the limiting diffusion in distribution. Let U be another separable, real Hilbert space, where $U=\mathcal{E}$ is possible, and assume there exists an operator $Q\in L_1(U,U)$ (or Q cylindrical) and a function[5] $g\in L^2((0,T),L_2(U,\mathcal{E}))$ for all $T>0$ such that $G(u(t),p(t)) = g(u(s),p(s))\circ Q\circ g^*(u(s),p(s))$ for all $t\geq 0$. Then, the process $(Z_t^Q)_{t\geq 0}$ defined by the stochastic integral

$$Z_t^Q := \int_0^t g(u(s),p(s))\,\mathrm{d}W_s^Q\,, \tag{4.3.11}$$

[5]Here, $L^2((0,T),L_2(U,\mathcal{E}))$ denotes the space of square-integrable functions on $(0,T)$ taking values in the Hilbert-space of Hilbert-Schmidt operators from U to $\mathcal{E}$.

where $(W_t^Q)_{t\geq 0}$ is an $\mathcal{E}$-valued Q–Wiener process, has the same quadratic variation as $(Z_t)_{t\geq 0}$ hence the processes coincide in distribution. Further, we remark that starting from the representation (4.3.11) the Langevin approximation is given by (4.3.10) with the obvious changes in the diffusion term. We note that in finite dimensions the non-uniqueness, see, e.g., [5, Chap. 8], of a stochastic integral to a given covariance matrix can be exploited to improve the speed of numerical approximations in Monte-Carlo simulations of diffusion approximations by choosing an optimal diffusion coefficient structure, see [89]. In infinite-dimensions the question of a practical implication of choosing a diffusion approximation based on (4.3.11) over (4.3.10) needs, to the best of our knowledge, still to be addressed.

4.4 Proofs of the main results

4.4.1 Proof of Theorem 4.2.1 (Law of large numbers)

The central argument of the subsequent proof is an appropriate application of Gronwall's Lemma such that the upper bound satisfies the convergence in probability. Here the estimating procedure yielding the estimate to which Gronwall's Lemma is applied necessitates careful attention due to more intricate regularity aspects of solutions to abstract evolution equations in contrast to solutions of ODEs in Euclidean space.

The continuous component Y_t^n of each PDMP is in between successive jump times the weak solution of an abstract evolution equation. Similarly $u(t)$ is the weak solution of the abstract evolution equation (4.2.1). Therefore also the difference of the two paths is in between jump times the weak solution of an abstract evolution equation. Therefore, it holds due to [43, Sec. 5.9, Thm. 3] for almost all t that

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}\|Y_t^n-u(t)\|_H^2\\ &\quad= 2\big\langle A(z^n(\theta_t^n))\,Y_t^n+B(Y_t^n,z^n(\theta_t^n))-A(p(t))\,u(t)-B(u(t),p(t))\,,\,Y_t^n-u(t)\big\rangle_X\,.\end{aligned}$$

Integrating this equation we obtain the integral equation

$$\begin{aligned}\|Y_{t_1}^n-u(t_1)\|_H^2 &= \|Y_{t_0}^n-u_{t_0}\|_H^2\\ &+2\int_{t_0}^{t_1}\big\langle A(z^n(\theta_s^n))Y_s^n+B(Y_s^n,z^n(\theta_s^n))-A(p(s))u(s)-B(u(s),p(s)),Y_s^n-u(s)\big\rangle_X\,\mathrm{d}s,\end{aligned}\tag{4.4.1}$$

which is valid for almost all t_0, t_1 in between two successive jump times. Since both sides of equation (4.4.1) are continuous the equality (4.4.1) even holds *for all* t_0, t_1 between successive jump times. Moreover, as Y_t^n is continuous also at jump times it follows that equation (4.4.1) holds for all $t \in [0,T]$, i.e., we have

$$\|Y_t^n - u(t)\|_H^2 \;=\; \|Y_0^n - u_0\|_H^2$$
$$+2\int_0^t \big\langle A(z^n(\theta_s^n))Y_s^n + B(Y_s^n, z^n(\theta_s^n)) - A(p(s))u(s) - B(u(s), p(s)), Y_s^n - u(s)\big\rangle_X \,\mathrm{d}s\,. \tag{4.4.2}$$

Next we employ the one-sided Lipschitz conditions (4.0.2) to estimate the integral in the right hand side of equation (4.4.2). This yields the inequality

$$\|Y_t^n - u_t\|_H^2 \;\le\; \|Y_0^n - u_0\|_H^2 + 2L_1\int_0^t \|Y_s^n - u(s)\|_H^2 \,\mathrm{d}s + 2L_1\sum_{j=1}^m \int_0^t \|z_j^n(\theta_s^n) - p_j(s)\|_E^2 \,\mathrm{d}s. \tag{4.4.3}$$

The overall aim is to apply Gronwall's inequality to the growth inequality (4.4.3). Therefore, in the next step we derive a control on the terms $\|z_j^n(\theta_s^n) - p_j(s)\|_E^2$ in the right hand side of inequality (4.4.3). As p is a solution of (4.2.1) satisfying (4.2.2) we obtain for every functional $\phi \in E^*$ a decomposition

$$\langle \phi, z_j^n(\theta_t^n) - p_j(t)\rangle_E \;=\; \langle \phi, z_j^n(\theta_0^n) - p_j(0)\rangle_E \tag{4.4.4}$$
$$+\int_0^t \big[\mathcal{A}^n\langle \phi, z_j^n(\cdot)\rangle_E\big](Y_s^n, \theta_s^n)\,\mathrm{d}s - \int_0^t \langle \phi, F_j(p(s), u(s))\rangle_E \,\mathrm{d}s + \langle \phi, M_j^n(t)\rangle_E\,,$$

where the term $\langle \phi, M_j^n(t)\rangle_E$ has precisely the form (4.1.2) for all $t \in [0,T]$. Next we expand the decomposition (4.4.4) to obtain

$$\begin{aligned}
\langle \phi, z_j^n(\theta_t^n) - p_j(t)\rangle_E \;=\;& \\
=\;& \langle \phi, z_j^n(\theta_0^n) - p_j(0)\rangle_E + \langle \phi, M_j^n(t)\rangle_E + \int_0^t \big[\mathcal{A}^n\langle \phi, z_j^n(\cdot)\rangle_E\big](Y_s^n, \theta_s^n)\,\mathrm{d}s\\
&\pm \int_0^t \langle \phi, F_j(z^n(\theta_s^n), Y_s^n)\rangle_E \,\mathrm{d}s - \int_0^t \langle \phi, F_j(p(s), u(s))\rangle_E \,\mathrm{d}s\\
=\;& \langle \phi, z_j^n(\theta_0^n) - p_j(0)\rangle_E + \langle \phi, M_j^n(t)\rangle_E\\
&+ \int_0^t \big[\mathcal{A}^n\langle \phi, z_j^n(\cdot)\rangle_E\big](Y_s^n, \theta_s^n) - \langle \phi, F_j(z^n(\theta_s^n), Y_s^n)\rangle_E \,\mathrm{d}s\\
&+ \int_0^t \big\langle \phi, F_j(z^n(\theta_s^n), Y_s^n) - F_j(p(s), u(s))\big\rangle_E \,\mathrm{d}s\,.
\end{aligned}$$

We take the supremum over all $\phi \in E^*$ with $\|\phi\|_{E^*} \le 1$ on both sides of this equation, square both sides and apply the inequality $|a_1+\ldots+a_k|^2 \le k(|a_1|^2+\ldots+|a_k|^2)$ and the Cauchy-Schwarz inequality to the right hand side which yields

$$\begin{aligned}
&\|z_j^n(\theta_t^n) - p_j(t)\|_E^2 \\
&\le 4\,\|z_j^n(\theta_0^n) - p_j(0)\|_E^2 + 4\,\|M_j^n(t)\|_E^2 + 4\Big(\int_0^t \big\|F_j(z^n(\theta_s^n), Y_s^n) - F_j(p(s), u(s))\big\|_E \,\mathrm{d}s\Big)^2 \\
&\quad + 4T \int_0^t \big\|\big[\mathcal{A}^n\langle\,\cdot\,, z_j^n(\cdot)\rangle_E\big](Y_s^n, \theta_s^n) - F_j(z^n(\theta_s^n), Y_s^n)\big\|_E^2 \,\mathrm{d}s\,.
\end{aligned}$$

We next apply the Lipschitz condition (4.2.3) on F and obtain the estimate

$$\begin{aligned}
&\|z_j^n(\theta_t^n) - p_j(t)\|_E^2 \qquad (4.4.5)\\
&\le\ \|z_j^n(\theta_0^n) - p_j(0)\|_E^2 + 4L_2 \int_0^t \|Y_s^n - u(s)\|_H^2 \,\mathrm{d}s + 4L_2 \sum_{i=1}^m \int_0^t \|z_i^n(\theta_s^n) - p_i(s)\|_E^2 \,\mathrm{d}s \\
&\quad + 4T \int_0^t \big\|\big[\mathcal{A}^n\langle\,\cdot\,, z_j^n(\cdot)\rangle_E\big](Y_s^n, \theta_s^n) - F_j(z^n(\theta_s^n), Y_s^n)\big\|_E^2 \,\mathrm{d}s + \|M_j^n(t)\|_E^2\,.
\end{aligned}$$

To further estimate this inequality we employ the convergence (4.2.4) of the term $\|M_j^n\|_E$ and the uniform convergence (4.2.5) of the generator. It follows by definition of these limits that for every $\epsilon_1 > 0$ and every $\delta > 0$ we can find an $N_{\epsilon_1,\delta}$ such that for all $n \ge N_{\epsilon_1,\delta}$ it holds due to (4.2.4) for all $j = 1,\ldots,m$ and all $t \in [0,T]$ that

$$\|M_j^n(t)\|_E \le \sqrt{\frac{\epsilon_1}{m}}\,,$$

and due to (4.2.5) that

$$\int_0^T \big\|\big[\mathcal{A}^n\langle\,\cdot\,, z_j^n(\cdot)\rangle_E\big](Y_s^n, \theta_s^n) - F_j(z^n(\theta_s^n), Y_s^n)\big\|_E^2 \,\mathrm{d}s \ \le\ \frac{\epsilon_1}{m}$$

on a set $\Omega_1 \subset \Omega$ satisfying $\mathbb{P}^n(\Omega\backslash\Omega_1) \le \delta$ for all $n \ge N_{\epsilon_1,\delta}$. Thus continuing to estimate only for paths on the set Ω_1 we obtain from (4.4.5) the inequality

$$\begin{aligned}
\|z_j^n(\theta_t^n) - p_j(t)\|_E^2 \ &\le\ 4\,\|z_j^n(\theta_0^n) - p_j(0)\|_E^2 + (4T+1)\,\frac{\epsilon_1}{m} \qquad (4.4.6)\\
&\quad + 4L_2 \int_0^t \|Y_s^n - u(s)\|_H^2 \,\mathrm{d}s + 4L_2 \sum_{i=1}^m \int_0^t \|z_i^n(\theta_s^n) - p_i(s)\|_E^2 \,\mathrm{d}s.
\end{aligned}$$

In order to finally obtain the growth estimate suitable for an application of Gronwall's inequality we add inequality (4.4.3) and inequalities (4.4.6) for all $j = 1, \ldots, m$ which yields

$$\|Y_t^n - u(t)\|_H^2 + \sum_{j=1}^m \|z_j^n(\theta_t^n) - p_j(t)\|_E^2 \;\le\; \|Y_0^n - u_0\|_H^2 + C_1 \sum_{j=1}^m \|z_j^n(\theta_0^n) - p_j(0)\|_E^2$$

$$+ C_2\epsilon_1 + C_3 \int_0^t \|Y_s^n - u(s)\|_H^2 \,\mathrm{d}s + C_3 \sum_{j=1}^m \int_0^t \|z_j^n(\theta_s^n) - p_j(s)\|_E^2 \,\mathrm{d}s \tag{4.4.7}$$

with constants given by $C_1 = 4$, $C_2 = 4T + 1$ and $C_3 = 2L_1 + 4L_2 m$. An application of Gronwall's inequality to (4.4.7) yields

$$\sup_{t\in[0,T]} \Big(\|Y_t^n - u(t)\|_H^2 + \sum_{j=1}^m \|z_j^n(\theta_t^n) - p_j(t)\|_E^2 \Big) \;\le\; K_1 \,\mathrm{e}^{K_2 T}\,, \tag{4.4.8}$$

where $K_2 = C_3$ and

$$K_1 = \|Y_0^n - u_0\|_H^2 + C_1 \sum_{j=1}^m \|z_j^n(\theta_0^n) - p_j(0)\|_E^2 + C_2\epsilon_1\,.$$

Finally, due to (C3), i.e., the convergence in probability of the initial conditions, it holds that for every $\epsilon_2 > 0$ we can find to every $\delta > 0$ an $N_{\epsilon_2,\delta}$ such that on a set $\Omega_2 \subset \Omega$ with $\mathbb{P}^n(\Omega\backslash\Omega_2) < \delta$ it holds for all $n \ge N_{\epsilon_2,\delta}$ that

$$\|Y_0^n - u_0\|_H^2 \le \frac{\epsilon_2}{m+2}, \quad \|z_j^n(\theta_0^n) - p_j(0)\|_E^2 \le \frac{\epsilon_2}{C_1(m+2)} \quad \forall\, j = 1, \ldots, m\,. \tag{4.4.9}$$

Let ϵ, $\delta > 0$ be arbitrary. Then we obtain choosing $\epsilon_2 = \epsilon\,\mathrm{e}^{-K_2 T}$ and $\epsilon_1 = \frac{\epsilon_2}{C_2(m+2)}$, thus $K_1 = \epsilon_2$, that for all $n \ge N_{\epsilon,\delta} := N_{\epsilon_1,\delta} \vee N_{\epsilon_2,\delta}$ it holds that

$$\sup\nolimits_{t\in[0,T]} \Big(\|Y_t^n - u(t)\|_H^2 + \sum_{j=1}^m \|z_j^n(\theta_t^n) - p_j(t)\|_E^2 \Big) \;\le\; \epsilon$$

on the set $\Omega_1 \cap \Omega_2$. Therefore it holds for all $n \ge N_{\epsilon,\delta}$ that

$$\mathbb{P}^n \Big[\sup\nolimits_{t\in[0,T]} \Big(\|Y_t^n - u(t)\|_H^2 + \sum_{j=1}^m \|z_j^n(\theta_t^n) - p_j(t)\|_E^2 \Big) \;>\; \epsilon \Big] \;\le\; 2\delta\,.$$

As δ and ϵ are arbitrary the statement (4.2.6) follows and the proof is completed.

4.4.2 Proof of Theorem 4.3.1 (Central limit theorem)

The proof of Theorem 4.3.1 is split into three successive steps. In the first step we proof tightness of the sequence of martingales which guarantees the existence of a limit. Secondly, we show that any limit is a continuous process. Finally, in the last step we prove that the limit is the specific diffusion process as stated in the theorem. The conditions (D1)–(D3) in Theorem 4.3.1 are such that each, in addition, to the preceding is needed in the successive steps of the proof. In particular, that means that (D1) is sufficient for tightness of the martingale sequence.

Tightness

In order to prove tightness of the sequence of $\mathcal{E}$–valued martingales $(\sqrt{\alpha_n}\, M_t^n)_{t\geq 0}$ it suffices to show that the following conditions are satisfied, cf. [91] wherein general conditions for tightness of sequences of Hilber space valued processes and, in particular, martingales are considered:

(T1) The sequence of initial conditions $(\sqrt{\alpha_n}\, M^n(0))_{n\geq 0}$ is tight.

(T2) For all $t \geq 0$ it holds that

$$\lim_{\delta\to\infty} \sup_{n\in\mathbb{N}} \mathbb{P}^n\big[\mathrm{Tr} \ll\!\sqrt{\alpha_n}\, M^n\!\gg_t > \delta\big] = 0\,, \tag{4.4.10}$$

and there exists an orthonormal basis $(\varphi_k)_{k\in\mathbb{N}}$ of E^* such that for each $\epsilon > 0$

$$\lim_{m\to\infty} \sup_{n\in\mathbb{N}} \mathbb{P}^n\Big[\sum\nolimits_{k>m} \langle \varphi_k, \ll\!\sqrt{\alpha_n}\, M^n\!\gg_t \varphi_k\rangle_E > \epsilon\Big] = 0\,. \tag{4.4.11}$$

(A) The sequence of the real-valued trace processes $(\mathrm{Tr} \ll\!\sqrt{\alpha_n}\, M^n\!\gg_t)_{t\geq 0}$, $n \geq \mathbb{N}$, satisfies the *Aldous condition*: For every $T, \epsilon, \delta > 0$ there exists a $h > 0$ and an $N > 0$ such that for any sequence of stopping times[6] $(\sigma^n)_{n\geq 0}$ with $\sigma^n \leq T$ it is valid that

$$\sup_{n\geq N} \sup_{0\leq s\leq h} \mathbb{P}^n\big[\,|\,\mathrm{Tr} \ll\!\sqrt{\alpha_n}\, M^n\!\gg_{\sigma^n+s} - \mathrm{Tr} \ll\!\sqrt{\alpha_n}\, M^n\!\gg_{\sigma^n}\,| \geq \delta\,\big] \leq \epsilon\,. \tag{4.4.12}$$

In connection to the necessary and sufficient conditions for tightness, briefly mentioned in the introduction of Section 4.3, it holds that conditions (T1) and (T2) imply that the sequences $(\sqrt{\alpha_n}\, M_t^n)_{n\in\mathbb{N}}$ are tight in E for almost all $t \in \mathbb{R}_+$ and the Aldous condition (A) implies the convergence of the modulus of continuity.

We proceed to proving the above conditions. First note that condition (T1) is trivially satisfied as $M_0^n = 0$ for all $n > 0$. Hence we proceed to condition (T2). In order to

[6]Here every σ^n is a stopping time on the respective probability space $(\Omega^n, \mathcal{F}^n, \mathbb{P}^n)$ with respect to the given filtration $(\mathcal{F}_t^n)_{t\geq 0}$

establish the first condition (4.4.10) we use Markov's inequality to obtain the estimate

$$\mathbb{P}^n\big[\mathrm{Tr} \ll\!\sqrt{\alpha_n}\, M^n\!\gg_t > \delta\big] \;\leq\; \frac{\alpha_n}{\delta}\,\mathbb{E}^n\Big[\int_0^t \mathrm{Tr}\, G^n(Y_s^n,\theta_s^n)\,\mathrm{d}s\Big]\,,$$

where the right hand side is finite due to (4.3.3) in condition (D1) and taking the supremum on both sides the same condition implies (4.4.10).
Next, to show the second condition (4.4.11) we employ Markov's inequality, the monotone convergence theorem (in order to change the order of expectation and the countable summation over all $k > m$), the form of the quadratic variation (4.1.8) and inequality (4.3.4) to obtain for the term in the left hand side the estimates

$$\begin{aligned}\mathbb{P}^n\big[\textstyle\sum_{k>m}\langle\varphi_k, \ll\!\sqrt{\alpha_n}\, M^n\!\gg_t \varphi_k\rangle_{\mathcal{E}} > \delta\big] \;&\leq\; \frac{\alpha_n}{\delta}\,\mathbb{E}^n\big[\textstyle\sum_{k>m}\langle\varphi_k, \ll\! M^n\!\gg_t \varphi_k\rangle_{\mathcal{E}}\big]\\ &\leq\; \frac{1}{\delta}\Big(\textstyle\sum_{k>m}\gamma_k\Big)\,C\,,\end{aligned}$$

where the upper bound is independent of $n \in \mathbb{N}$. Moreover, the property $\sum_{k\in\mathbb{N}}\gamma_k < \infty$ implies that $\lim_{m\to\infty}\sum_{k>m}\gamma_k = 0$ and hence (4.4.11) holds for all $t \geq 0$.
Finally, it remains to show (A). Let $T, \delta > 0$ and $\sigma^n < T$ be an arbitrary sequence of stopping times, then for all $h > 0$ it holds that for $s \leq h$

$$\begin{aligned}&\mathbb{P}^n\Big[\,\big|\,\mathrm{Tr} \ll\!\sqrt{\alpha_n}\, M^n\!\gg_{\sigma_n+s} - \mathrm{Tr} \ll\!\sqrt{\alpha_n}\, M^n\!\gg_{\sigma_n}\big| \geq \delta\Big]\\ &\qquad= \;\mathbb{P}^n\Big[\alpha_n\int_{\sigma_n}^{\sigma_n+s} \mathrm{Tr}\, G^n(Y_r^n,\theta_r^n)\,\mathrm{d}r \geq \delta\Big]\\ &\qquad\leq\; \frac{\alpha_n}{\delta}\,\Big(\mathbb{E}^n\Big[\int_0^{\sigma^n+h} \mathrm{Tr}\, G^n(Y_r^n,\theta_r^n)\,\mathrm{d}r\Big] - \mathbb{E}^n\Big[\int_0^{\sigma^n} \mathrm{Tr}\, G^n(Y_r^n,\theta_r^n)\,\mathrm{d}r\Big]\Big)\,.\end{aligned}$$

Here the upper bound is finite due to (4.3.3), independent of $s \leq h$ and strictly decreasing in h. Moreover, taking the supremum on both sides over all $n \in \mathbb{N}$ the upper bound is still finite and strictly decreasing in h. Hence, for every $\epsilon > 0$ we can find an h small enough such that (4.4.12) holds. Condition (A) is thus satisfied.

Limit is a continuous process

In the preceding part of the proof we have established that the laws of the sequence of martingales $(\sqrt{\alpha}\, M_t^n)_{t\geq 0}$ are tight which is equivalent to there existence of a weakly convergent subsequence. We now prove that under the additional condition (D2) any cluster point of the sequence is a measure supported on $C^0(\mathbb{R}_+,\mathcal{E})$. That is, the limit is a process with almost surely continuous paths. The method of proof follows the outline of [71, Lemma 3.2] adapted for the stochastic processes being PDMPs on Hilbert spaces, the general setup in this chapter and the particular conditions (D1) and (D2) in Theorem 4.3.1 which differ from [71]. Furthermore, we have extended the result in [71, Lemma 3.2], which only considers convergence on finite time intervals $[0,T]$, to convergence on $D(\mathbb{R}_+,\mathcal{E})$. In the following we employ the abbreviations $Z^n_t := \sqrt{\alpha_n}\, M^n_t$ and $\Delta_t Z^n := Z^n_t - Z^n_{t-}$, i.e., $(\Delta_t Z^n)_{t\geq 0}$ denotes the process of jump heights. Note that $\Delta_t Z^n = \sqrt{\alpha_n}\,\Delta_t z^n(\theta^n)$.

Further, let $\mathbb{P}^*$ denote an accumulation point of the sequence $(\mathbb{P}^n)_{n\in\mathbb{N}}$. Without loss of generality we use $\mathbb{P}^n$, $n\geq 1$, to also denote the subsequence converging weakly to $\mathbb{P}^*$. Furthermore, here $\mathbb{P}^n$ is understood as a law on the Skorokhod space $D(\mathbb{R}_+,\mathcal{E})$ given by the pushforward measure of the process $(\sqrt{\alpha_n}\, M^n_t)_{t\geq 0}$. Then due to the Skorokhod Representation Theorem, e.g., [41, Chap. 3,Thm. 1.8], there exists a probability space $(\Omega^o,\mathcal{F}^o,\mathbb{P}^o)$ supporting $D(\mathbb{R}_+,\mathcal{E})$–valued random variables ζ^n, $n\geq 1$ and ζ^* with distributions $\mathbb{P}^n$ and $\mathbb{P}^*$, respectively, such that ζ^n converges to ζ^* almost surely with respect to $\mathbb{P}^o$. Further, it clearly holds that, e.g., $\mathbb{E}^n f(Z^n) = \mathbb{E}^o f(\zeta^n)$ for suitable functionals f.

We begin the proof with preliminary estimates on functions evaluated along the path of the PDMPs. These ultimately allow to infer that the process of jumps heights converges to the function being constantly zero. Let g be a measurable, bounded, non-negative function $g:\mathbb{R}\to\mathbb{R}$, that vanishes in a neighbourhood of 0 and of ∞, that is, there exists a finite constant $C_g := \sup_{x\in\mathbb{R}} \frac{g(x)^2}{x^2} < \infty$. For such a function g and any $\Phi\in\mathcal{E}^*$ we define the process

$$\begin{aligned} G^n_t(\langle\Phi,Z^n\rangle_{\mathcal{E}}) \;&:=\; \sum_{s\in(0,t]} g\big(\langle\Phi,\Delta_s Z^n\rangle_{\mathcal{E}}\big) \\ &\qquad - \int_0^t \Lambda^n(Y^n_s,\theta^n_s)\int_{K_n} g\big(\sqrt{\alpha_n}\langle\Phi, z^n(\xi)-z^n(\theta^n_s)\rangle_{\mathcal{E}}\big)\,\mu^n\big((Y^n_s,\theta^n_s)\big)\,\mathrm{d}s \\ &= \int_0^t\int_{K_n} g\big(\sqrt{\alpha_n}\langle\Phi, z^n(\xi)-z^n(\theta^n_{s-})\rangle_{\mathcal{E}}\big)\,M^n(\mathrm{d}\xi,\mathrm{d}s)\,, \end{aligned}$$

where M^n is the martingale measure associated with the PDMP. Hence we infer that the $G^n_t(\langle\Phi,Z^n\rangle_{\mathcal{E}})$ is a martingale. Note that the above summation over all $s\in(0,t]$ is well-defined as the PDMPs are regular and thus $g\big(\langle\Phi,\Delta_s Z^n\rangle_{\mathcal{E}}\big)$ is non-zero for only finitely many $s\leq t$.

The proof now proceeds as follows. We first show (a) that for all $t\geq 0$ the random variables $G^n_t(\langle\Phi,\zeta^n\rangle_{\mathcal{E}})$, $n\in\mathbb{N}$, are uniformly integrable and (b) that they converge to $\sum_{s\in(0,t]} g\big(\langle\Phi,\Delta_s\zeta^n\rangle_{\mathcal{E}}\big)$ in probability. This allows to infer that the convergence result also holds as convergence in mean.[7] In part (c) we then use these results to show that

[7] For real-valued random variables convergence in the mean is equivalent to convergence in probability and uniform integrability.

the jump heights of the canonical process of the law $\mathbb{P}^*$ are constantly zero almost surely. This implies that $\mathbb{P}^*\big(C^0([0,t],\mathcal{E})\big) = 1$ for every $t > 0$ where $C^0([0,t],\mathcal{E})$ is understood as the subset of $D(\mathbb{R}_+,\mathcal{E})$ consisting of those càdlàg functions which are continuous up to and including time t. The proof is completed by (d) extending this result to $\mathbb{P}^*\big(C^0(\mathbb{R}_+,\mathcal{E})\big) = 1$.

(a) To show that the sequence of random variables $G^n_t(\langle\Phi,\zeta^n\rangle_{\mathcal{E}})$, $n\in\mathbb{N}$, is uniformly integrable in the space $(\Omega^o,\mathcal{F}^o,\mathbb{P}^o)$ it is sufficient that the second moments are uniformly bounded, cf. [41, Appendix, Prop. 2.2]. The Itô-isometry for real-valued stochastic integrals with respect to the associated martingale measures, which is implied by taking the expectation of the processes in Proposition B.2.5, yields

$$\begin{aligned}
\sup_{n\in\mathbb{N}}\mathbb{E}^o|G^n_t(\langle\Phi,\zeta^n\rangle_{\mathcal{E}})|^2 &= \sup_{n\in\mathbb{N}}\mathbb{E}^n|G^n_t(\langle\Phi,Z^n\rangle_{\mathcal{E}})|^2\\
&= \sup_{n\in\mathbb{N}}\mathbb{E}^n\Big[\int_0^t\Lambda^n(Y^n_s,\theta^n_s)\int_{K_n} g\big(\sqrt{\alpha_n}\langle\Phi,z^n(\xi)-z^n(\theta^n_s)\rangle_{\mathcal{E}}\big)^2\,\mu^n\big((Y^n_s,\theta^n_s),\mathrm{d}\xi\big)\,\mathrm{d}s\Big].
\end{aligned}$$

Therefore, employing the special structure of the map g we obtain the estimate

$$\begin{aligned}
&\sup_{n\in\mathbb{N}}\mathbb{E}^o|G^n_t(\langle\Phi,\zeta^n\rangle_{\mathcal{E}})|^2\\
&\le C_g\sup_{n\in\mathbb{N}}\alpha_n\mathbb{E}^n\Big[\int_0^t\Lambda^n(Y^n_s,\theta^n_s)\int_{K_n}\big|\langle\Phi,z^n(\xi)-z^n(\theta^n_s)\rangle_{\mathcal{E}}\big|^2\,\mu^n\big((Y^n_s,\theta^n_s),\mathrm{d}\xi\big)\,\mathrm{d}s\Big],
\end{aligned}$$

where the right hand side is finite for every $t>0$ due to condition (4.3.3) in (D1).

(b) In this part of the proof we establish convergence in probability of the random variables $G^n_t(\langle\Phi,\Delta\zeta^n\rangle_{\mathcal{E}})$. Let $\beta>0$ be such that $g(x)=0$ for $|x|\le\beta$, i.e., the interval $(-\beta,\beta)$ is contained in the neighbourhood of 0 whereon g vanishes. Then we obtain using Markov's inequality and due to the boundedness of g the estimates

$$\begin{aligned}
&\mathbb{P}^o\Big[\sum\nolimits_{s\in(0,t]}g(\langle\Phi,\Delta_s\zeta^n\rangle_{\mathcal{E}})-G^n_t(\langle\Phi,\zeta^n\rangle_{\mathcal{E}})\;>\;\delta\Big]\\
&= \mathbb{P}^n\Big[\int_0^t\Lambda^n(Y^n_s,\theta^n_s)\int_{K_n}g\big(\sqrt{\alpha_n}\,\langle\Phi,z^n(\xi)-z^n(\theta^n_s)\rangle_E\big)\,\mu^n\big((Y^n_s,\theta^n_s),\mathrm{d}\xi\big)\,\mathrm{d}s\;>\;\delta\Big]\\
&\le \frac{1}{\delta}\,\mathbb{E}^n\Big[\int_0^t\Lambda^n(Y^n_s,\theta^n_s)\int_{K_n}g\big(\sqrt{\alpha_n}\langle\Phi,z^n(\xi)-z^n(\theta^n_s)\rangle_E\big)\,\mu^n\big((Y^n_s,\theta^n_s)\big)\,\mathrm{d}s\Big]\\
&\le \frac{\sup_{x\in\mathbb{R}}|g(x)|}{\delta}\,\mathbb{E}^n\Big[\int_0^t\Lambda^n(Y^n_s,\theta^n_s)\int_{\sqrt{\alpha_n}\,|\langle\Phi,z^n(\xi)-z^n(\theta^n_s)\rangle_E|>\beta}\mu^n\big((Y^n_s,\theta^n_s),\mathrm{d}\xi\big)\,\mathrm{d}s\Big].
\end{aligned}$$

Thus due to condition (4.3.5) in (D2) it holds that

$$\lim_{n\to\infty} \mathbb{P}^o\Big[\sum\nolimits_{s\in(0,t]} g(\langle\Phi,\Delta_s\zeta^n\rangle_{\mathcal{E}}) - G_t^n(\langle\Phi,\zeta^n\rangle_{\mathcal{E}}) \;>\; \delta\Big] = 0\,.$$

Moreover, it holds in $(\Omega^o, \mathcal{F}^o, \mathbb{P}^o)$ almost surely that

$$\lim_{n\to\infty} \sum_{s\in(0,t]} g(\langle\Phi,\Delta_s\zeta^n\rangle_{\mathcal{E}}) = \sum_{s\in(0,t]} g(\langle\Phi,\Delta_s\zeta^*\rangle_{\mathcal{E}})\,.$$

Therefore, combining these two convergence results we obtain that

$$G_t^n(\langle\Phi,\zeta^n\rangle_{\mathcal{E}}) \longrightarrow \sum_{s\in(0,t]} g(\langle\Phi,\Delta_s\zeta^*\rangle_{\mathcal{E}}) \tag{4.4.13}$$

holds as convergence in probability.

(c) From parts (a) and (b) we infer that(4.4.13) also holds as convergence in mean. Together with Jensen's inequality this implies

$$\begin{aligned}\lim_{n\to\infty} \Big|\mathbb{E}^o\Big(G_t^n(\langle\Phi,\Delta_s\zeta^n\rangle_{\mathcal{E}}) - \sum_{s\in(0,T]} g(\langle\Phi,\Delta_s\zeta^*\rangle_{\mathcal{E}})\Big)\Big| \\ \leq \lim_{n\to\infty} \mathbb{E}^o\Big|G_t^n(\langle\Phi,\Delta_s\zeta^n\rangle_{\mathcal{E}}) - \sum_{s\in(0,t]} g(\langle\Phi,\Delta_s\zeta^*\rangle_{\mathcal{E}})\Big| \;=\; 0\,,\end{aligned}$$

and hence we infer that

$$\mathbb{E}^o \sum_{s\in(0,t]} g(\langle\Phi,\Delta_s\zeta^*\rangle_{\mathcal{E}}) \;=\; \lim_{n\to\infty} \mathbb{E}^o G_t^n(\langle\Phi,\zeta^n\rangle_{\mathcal{E}})\,. \tag{4.4.14}$$

Furthermore, $G_t^n(\langle\Phi,Z^n\rangle_{\mathcal{E}})$ is a martingale which satisfies $G_0^n(\langle\Phi,Z^n\rangle_{\mathcal{E}}) = 0$, This, in turn, implies that $\mathbb{E}^n G_t^n(\langle\Phi,Z^n\rangle_{\mathcal{E}}) = 0$ for every $n\in\mathbb{N}$. Therefore we obtain due to (4.4.14)

$$\begin{aligned}\mathbb{E}^* \sum_{s\in(0,t]} g(\langle\Phi,\Delta_s Z\rangle_{\mathcal{E}}) \;&=\; \mathbb{E}^o \sum_{s\in(0,t]} g(\langle\Phi,\Delta_s\zeta^*\rangle_{\mathcal{E}}) \\ &=\; \lim_{n\to\infty} \mathbb{E}^o G_t^n(\langle\Phi,\zeta^n\rangle_{\mathcal{E}}) \;=\; \lim_{n\to\infty} \mathbb{E}^n G_t^n(\langle\Phi,Z^n\rangle_{\mathcal{E}}) \;=\; 0\,.\end{aligned} \tag{4.4.15}$$

In a next step, let g_m be a sequence of functions satisfying the properties for functions g proposed above. Further we assume that the functions $g_m(x)$ increase pointwise to x^2 for $m\to\infty$ (for an example of such functions we refer to [71]). Then due to the monotone convergence theorem it holds that

$$\lim_{m\to\infty} \mathbb{E}^* \sum_{s\in(0,t]} g_m(\langle \Phi, \Delta_s Z\rangle_{\mathcal{E}}) \,=\, \mathbb{E}^* \sum_{s\in(0,t]} |\langle \Phi, \Delta_s Z\rangle_{\mathcal{E}}|^2 .$$

Furthermore, the limiting expectation in the right hand side is zero as each element of the sequence of expectations in the left hand side is zero due to (4.4.15). Next we choose Φ to be an element of an orthonormal basis $(\varphi_k)_{k\in\mathbb{N}}$ of $\mathcal{E}$. Then we sum the expectations for all elements of the basis, i.e., we obtain the term

$$\sum_{k\in\mathbb{N}} \mathbb{E}^* \sum\nolimits_{s\in(0,t]} |\langle \varphi_k, \Delta_s Z\rangle_{\mathcal{E}}|^2 .$$

Due to the dominated convergence theorem we can interchange the countable summation and the expectation and as the PDMP is regular we can afterwards interchange the resulting two summation inside the expectation. Then Parseval's identity yields

$$\mathbb{E}^* \sum_{s\in(0,t]} \|\Delta_s Z\|^2_{\mathcal{E}} \,=\, 0 .$$

As the non-negative random variable inside the expectation is zero only for continuous paths of the process $(Z_s)_{s\in[0,t]}$ we infer that almost all paths are continuous, i.e., $\mathbb{P}^*\big(C^0([0,t],\mathcal{E})\big) = 1$.

(d) To conclude the proof let t_k, $k\in\mathbb{N}$, be a sequence of times increasing to infinity then

$$C^0(\mathbb{R}_+,\mathcal{E}) = \bigcap_{k\in\mathbb{N}} C^0([0,t_k],\mathcal{E}) ,$$

and the events in the right hand side satisfy $C^0([0,t_{k+1}],\mathcal{E}) \subseteq C^0([0,t_k],\mathcal{E})$. The properties of a probability measure thus yield

$$\mathbb{P}^*\big(C^0(\mathbb{R}_+,\mathcal{E})\big) = \lim_{k\to\infty} \mathbb{P}^*\big(C^0([0,t_k],\mathcal{E})\big) = 1 ,$$

that is a process with distribution given by the limit $\mathbb{P}^*$ possesses almost surely continuous paths.

Limit is a diffusion process

In the final part of the proof we uniquely characterise the limit of the sequence of martingales $(\sqrt{\alpha_n}\, M^n_t)_{t\geq 0}$ under the additional assumptions (D3). The method of proof is via the local martingale problem motivated by a proof presented in [91], i.e., the limiting probability measure is the unique solution to a particular martingale problem. The author in [91] considers Hilbert space valued stochastic integral equations driven by Hilbert space valued martingales with state dependent quadratic variation. A central limit theorem for the martingales is presented. The arguments of the subsequent proof are closely related to [91]. This is as the general result on martingales associated with PDMPs, which we have proven in Section 4.1, result in the problem in this part of the proof to be of the same underlying structure as in [91]. One difference, however, is that the present conditions (D1)–(D3) are more general than the conditions in [91] and adapted to the PDMP setup, hence some estimates differ.

In this part of the proof we interpret the sequence of martingales $(\sqrt{\alpha_n}\, M^n_t)_{t\geq 0}$ defined on the probability spaces $(\Omega^n, \mathcal{F}^n, (\mathcal{F}^n_t)_{t\geq 0}, \mathbb{P}^n)$ as random variables on the space $D(\mathbb{R}_+, \mathcal{E})$ equipped with its natural σ-field $\mathcal{D}$. Further, laws on the canonical space are given by the pushforward measure. In order to simplify the notation we denote the laws on the canonical space also by $\mathbb{P}^n$. Due to results in the preceding two parts of the proof we know the sequence $\mathbb{P}^n$, $n \in \mathbb{N}$, admits a limit $\mathbb{P}^*$ supported on $C^0(\mathbb{R}_+, \mathcal{E})$. We use $(\zeta_t)_{t\geq 0}$ to denote the canonical process[8] on $D(\mathbb{R}_+, \mathcal{E})$ which is a version of the martingale $(\sqrt{\alpha_n}\, M^n_t)_{t\geq 0}$ under the push-forward maesure $\mathbb{P}^n$ for all $n \in \mathbb{N}$ or of the weak limit under the measure $\mathbb{P}^*$.

In the following we prove that the limit $\mathbb{P}^*$ is a solution to a local martingale problem the unique solution of which is an $\mathcal{E}$–valued centered diffusion process with covariance operator $C(t) \in L_1(\mathcal{E}^*, \mathcal{E})$ as given in (4.3.2). For any twice continuously differentiable function $f : \mathcal{E} \to \mathbb{R}$ the extended generator $\mathcal{A}f$ of such a diffusion is given by

$$\mathcal{A}f(x,t) = \frac{1}{2}\mathrm{Tr}\left(D^2 f(x) \circ G(t)\right).$$

Then, as in [91] in order to uniquely characterise the solution to the local martingale problem connected with this generator and supported on the space $C^0(\mathbb{R}, \mathcal{E})$ it suffices to consider mappings f of the form $\langle \Phi, \cdot\rangle_{\mathcal{E}}$ and $\langle \Phi, \cdot\rangle^2_{\mathcal{E}}$ for all $\Phi \in \mathcal{E}^*$.[9] That is, we

[8]The canonical process on $D(\mathbb{R}_+, \mathcal{E})$ is similar to the concept of a canonical random counting measure or a canonical marked point process considered in Appendix B.2. To be precise, the *canonical process is a probability on the space* $D(\mathbb{R}+, \mathcal{E})$ equipped with its natural filtration $(\mathcal{D}_t)_{t\geq 0}$ where $\mathcal{D}_t := \sigma(w(s), s \leq t, w \in D(\mathbb{R}_+, \mathcal{E}))$ and $\mathcal{D} := \bigvee_{t\geq 0} \mathcal{D}_t$. Then we define on the canonical space we the $\mathcal{E}$-valued 'process' $(\zeta_t)_{t\geq 0}$ by $\zeta_t : D(\mathbb{R}_+, \mathcal{E}) \to \mathcal{E} : w \mapsto \zeta_t(w) := w(t)$. Each $\mathcal{E}$–valued continuous process defines a unique canonical process via its pushforward measure and the 'process' $(\zeta_t)_{t\geq 0}$ with respect to the canonical process given by the pushforward measure is a version of the original process.

[9]The considerations in the following are generally valid although we use the particular notation of the present chapter. First of all, in the case of an one-dimensional diffusion it is known that the functions $x \to x$ and $x \to x^2$ are sufficient to uniquely characterise the solution of the corresponding martingale problem. Now, suppose that a process $(\zeta_t)_{t\geq 0}$ is a Hilbert space-valued diffusion with covariance operator $C(t)$. This is equivalent to each $(\langle \Phi, \zeta_t\rangle)_{t\geq 0}$, where Φ is an arbitrary element of the dual space, being a diffusion process with covariance operator $\langle \Phi, C(t)\Phi\rangle$. Hence, each of these is the unique solution to the martingale problem such that the processes $t \mapsto \langle \Phi, \zeta_t\rangle$ and

$$t \mapsto \langle \Phi, \zeta_t\rangle^2 - \langle \Phi, C(t)\Phi\rangle$$

are martingales. Conversely, it is now clear that if there exists a probability such that we obtain have to show that the canonical process ζ_t in $C^0(\mathbb{R}_+,\mathcal{E})$ is such that for all $\Phi\in\mathcal{E}^*$ the processes $\langle\Phi,\zeta_t\rangle_{\mathcal{E}}$ and

$$\langle\Phi,\zeta_t\rangle_{\mathcal{E}}^2-\int_0^t\langle\Phi,G(u_s,p_s)\Phi\rangle_{\mathcal{E}}\,\mathrm{d}s \tag{4.4.16}$$

are $\mathbb{P}^*$-local martingales. We start introducing some notation and then show in parts (a) and (b) the martingale properties of the two indicated processes on the canonical space $D(\mathbb{R}_+,\mathcal{E})$. First, however we have to introduce some more notation and a useful result employed in the remainder of the proof.

As before we use $Z^n_t:=\sqrt{\alpha_n}\,M^n_t$ and $\Delta_t Z^n:=Z^n_t-Z^n_{t-}$. Further, as indicated above the notation is such that we use $\mathbb{P}^n$ and $\mathbb{E}^n$ to denote probabilities and expectations on the original given measurable spaces $(\Omega^n,\mathcal{F}^n)$ as well as on the canonical space $(D(\mathbb{R}-_+,\mathcal{E}),\mathcal{D})$. That is, e.g., $\mathbb{E}^n f(Z^n_t)=\mathbb{E}^n f(\zeta_t)$ for any bounded function f, where the former is the expectation taken on the original space $(\Omega^n,\mathcal{F}^n,\mathbb{P}^n)$ and the latter the expectation on the canonical space of càdlàg processes with respect to the pushforward measure.
We also employ the Itô-formula [90, Thm. 25.7] for smooth[10] functions $f\in C^\infty_c(\mathbb{R})$ applied to semi-martingales. For the particular choice of the semi-martingales being the real martingales $\langle\Phi,Z^n_t\rangle_{\mathcal{E}}$ the Itô-formula reads

$$\begin{aligned} f\big(\langle\Phi,Z^n_t\rangle_{\mathcal{E}}\big) &= \frac{1}{2}\int_0^t f''\big(\langle\Phi,Z^n_{s-}\rangle_{\mathcal{E}}\big)\,\big(\langle\Phi,\alpha_n\ll M^n\gg_t\Phi\rangle_{\mathcal{E}}\big)\mathrm{d}s\\ &\quad+\sum_{s\le t}\Big[f(\langle\Phi,Z^n_s\rangle_{\mathcal{E}})-f(\langle\Phi,Z^n_{s-}\rangle_{\mathcal{E}})-\langle\Phi,\Delta_s Z^n\rangle_{\mathcal{E}}\,f'(\langle\Phi,Z^n_{s-}\rangle_{\mathcal{E}})\Big]\\ &\quad-\frac{1}{2}\sum_{s\le t}\Big[\langle\Phi,\Delta_s Z^n\rangle^s_{\mathcal{E}}\,f''(\langle\Phi,Z^n_t\rangle_{\mathcal{E}})\Big]+M^{f,n}_t \end{aligned}\tag{4.4.17}$$

where $(M^{f,n}_t)_{t\ge0}$ is a particular martingale on $(\Omega^n,\mathcal{F}^n,(\mathcal{F}^n_t)_{t\ge0},\mathbb{P}^n)$ depending on Z^n and f.

Next, we introduce on the canonical space for all positive ρ the stopping times $\tau_\rho:=\inf\{t\in\mathbb{R}_+\,|\,\|\zeta_t\|_{\mathcal{E}}>\rho\}$ and note that due to the bound (4.3.7) in (D3) on the jump heights we have that for any law $\mathbb{P}^n$, $n\ge1$, it holds almost surely

$$\|\zeta_{\tau_\rho}\|_{\mathcal{E}}\le\rho+C\,. \tag{4.4.18}$$

Analogously we define the stopping times $\tau^n_\rho:=\inf\{t\in\mathbb{R}_+\,|\,\|Z^n_t\|_{\mathcal{E}}>\rho\}$ on the spaces

a family of diffusion processes $\{(\langle\Phi,\zeta_t\rangle)_{t\ge0},\Phi$ being a bounded, linear fuctional$\}$ then this familiy defines a unique Hilbert space-valued diffusion given by $(\zeta_t)_{t\ge0}$.

[10]The set $C^\infty_c(\mathbb{R})$ denotes the space of infinitely often differentiable real-valued functions with compact support in $\mathbb{R}$.

$(\Omega^n, \mathcal{F}^n, (\mathcal{F}^n_t)_{t\geq 0}, \mathbb{P}^n)$.

Finally, as already mentioned $(\mathcal{D}_t)_{t\geq 0}$ denotes the natural filtration on the canonical space. Then for $A \in \mathcal{D}_t$ we define $A^n := (Z^n)^{-1}F \in \mathcal{F}^n_t$ its preimage with respect to the random variable Z^n. We now proceed to show that the two processes $\langle \Phi, \zeta_t\rangle_{\mathcal{E}}$ and (4.4.16) are indeed martingales with respect to the limit measure $\mathbb{P}^*$.

(a) Let $\Phi \in \mathcal{E}^*$ be fixed and we choose for every ρ a smooth function $f_\rho \in C^\infty_c(\mathbb{R})$ which satisfies $f_\rho(x) = x$ if $|x| \leq \|\Phi\|_{\mathcal{E}^*}(\rho + C)$ and thus $f'(x) = 1$ and $f''(x) = 0$ for $|x| \leq \|\Phi\|_{\mathcal{E}^*}(\rho + C)$. Therefore it holds for $t < \tau^n_\rho$, which implies the estimate $|\langle \Phi, Z^n_{t-}\rangle_{\mathcal{E}}| \leq \|\Phi\|_{\mathcal{E}^*}(\rho + C)$, that

$$f''_\rho(\langle \Phi, Z^n_{t-}\rangle_{\mathcal{E}}) = 0$$

and

$$f_\rho(\langle \Phi, Z^n_t\rangle_{\mathcal{E}}) - f_\rho(\langle \Phi, Z^n_{t-}\rangle_{\mathcal{E}}) - \langle \Phi, \Delta_t Z^n\rangle_{\mathcal{E}}\, f'_\rho(\langle \Phi, Z^n_{s-}\rangle_{\mathcal{E}}) = 0.$$

It follows that applying the Itô-formula (4.4.17) to the function f_ρ and the martingale $Z^n_{t\wedge\tau^n_\rho}$ all terms besides the martingale M^{n,f_ρ} vanish in the the right hand side.

Therefore we obtain for $t_2 \geq t_1$ and all $A \in \mathcal{D}_{t_1}$ that

$$\begin{aligned}
\mathbb{E}^n\Big[\mathbb{I}_A\Big(\langle \Phi, \zeta_{t_2\wedge\tau_\rho}\rangle_{\mathcal{E}} - \langle \Phi, \zeta_{t_1\wedge\tau_\rho}\rangle_{\mathcal{E}}\Big)\Big] &= \mathbb{E}^n\Big[\mathbb{I}_A\Big(f_\rho(\langle \Phi, \zeta_{t_2\wedge\tau_\rho}\rangle_{\mathcal{E}}) - f_\rho(\langle \Phi, \zeta_{t_1\wedge\tau_\rho}\rangle_{\mathcal{E}})\Big)\Big] \\
&= \mathbb{E}^n\Big[\mathbb{I}_{A^n}\Big(f_\rho(\langle \Phi, Z^n_{t_2\wedge\tau^n_\rho}\rangle_{\mathcal{E}}) - f_\rho(\langle \Phi, Z^n_{t_1\wedge\tau^n_\rho}\rangle_{\mathcal{E}})\Big)\Big] \\
&= 0\,.
\end{aligned} \tag{4.4.19}$$

The proof of the first martingale property is concluded as in [91]: The mapping $\zeta \to f_\rho(\langle \Phi, \zeta_{t_2\wedge\tau_\rho}\rangle_{\mathcal{E}})$ is almost surely (with respect to the probability $\mathbb{P}^*$) continuous and as $\mathbb{P}^n$ converges weakly to $\mathbb{P}^*$ it holds due to (4.4.19) that

$$\mathbb{E}^*\Big[\mathbb{I}_A\Big(\langle \Phi, \zeta_{t_2\wedge\tau_\rho}\rangle_{\mathcal{E}} - \langle \Phi, \zeta_{t_1\wedge\tau_\rho}\rangle_{\mathcal{E}}\Big)\Big] = 0\,.$$

We infer from the definition of the conditional expectation that the stopped processes are martingales. Furthermore, as ζ_t possesses continuous paths almost surely under the measure $\mathbb{P}^*$ it holds that τ_ρ diverges to ∞ almost surely for $\rho \to \infty$. Hence, we can find a sequence of stopping times τ_{ρ_k}, $k \in \mathbb{N}$, such that $\tau_{\rho_k} \to \infty$ almost surely for $k \to \infty$. Thus it holds that the process $\langle \Phi, \zeta_t\rangle$ is a local martingale with respect to $\mathbb{P}^*$.

(b) For the second class of processes we consider smooth functions $g_\rho \in C^\infty_c(\mathbb{R})$ such that $g_\rho(x) = x^2$ for all $|x| \leq \|\Phi\|_{\mathcal{E}^*}(\rho + C)$. Starting from the definition of the conditional expectation as in (4.4.19) we obtain

$$\begin{aligned}
&\mathbb{E}^n\Big[\mathbb{I}_A\Big(\langle\Phi,\zeta_{t_2\wedge\tau_\rho}\rangle^2_{\mathcal{E}}-\int_0^{t_2\wedge\tau_\rho}\langle\Phi,G(u(s),p(s))\Phi\rangle_{\mathcal{E}}\,\mathrm{d}s-\langle\Phi,\zeta_{t_1\wedge\tau_\rho}\rangle^2_{\mathcal{E}}\\
&\qquad\qquad+\int_0^{t_1\wedge\tau_\rho}\langle\Phi,G(u(s),p(s))\Phi\rangle_{\mathcal{E}}\,\mathrm{d}s\Big)\Big]\\
&=\ \mathbb{E}^n\Big[\mathbb{I}_A\Big(\langle\Phi,\zeta_{t_2\wedge\tau_\rho}\rangle^2_{\mathcal{E}}-\langle\Phi,\zeta_{t_1\wedge\tau_\rho}\rangle^2_{\mathcal{E}}-\int_{t_1\wedge\tau_\rho}^{t_2\wedge\tau_\rho}\langle\Phi,G(u(s),p(s))\Phi\rangle_{\mathcal{E}}\,\mathrm{d}s\Big)\Big]\\
&=\ \mathbb{E}^n\Big[\mathbb{I}_{A^n}\Big(\langle\Phi,Z^n_{t_2\wedge\tau_\rho}\rangle^2_{\mathcal{E}}-\langle\Phi,Z^n_{t_1\wedge\tau_\rho}\rangle^2_{\mathcal{E}}-\int_{t_1\wedge\tau_\rho}^{t_2\wedge\tau_\rho}\alpha_n\,\langle\Phi,G^n(Y^n_s,\theta^n_s)\Phi\rangle_{\mathcal{E}}\,\mathrm{d}s\Big)\Big]\\
&\qquad+\mathbb{E}^n\Big[\mathbb{I}_{A^n}\Big(\int_{t_1\wedge\tau_\rho}^{t_2\wedge\tau_\rho}\alpha_n\,\langle\Phi,G^n(Y^n_s,\theta^n_s)\Phi\rangle_{\mathcal{E}}-\langle\Phi,G^n(u(s),p(s))\Phi\rangle_{\mathcal{E}}\,\mathrm{d}s\Big)\Big]\,.
\end{aligned}$$

Here the first expectation in the final right hand side vanishes due to the Itô-formula (4.4.17): We apply the Itô-formula for the function g_ρ and the martingales $Z^n_{t\wedge\tau^n_\rho}$ to the terms $\langle\Phi,Z^n_{t_2\wedge\tau_\rho}\rangle^2_{\mathcal{E}}$ and $\langle\Phi,Z^n_{t_1\wedge\tau_\rho}\rangle^2_{\mathcal{E}}$. Then we find – similarly to part (a) – that the summands in the right hand side of the Itô-formula vanish. Therefore we are left with only the martingale M^{n,g_ρ} and the integral term, wherein $g''_\rho(\langle\phi,Z^n_{t-}\rangle_{\mathcal{E}})=2$ for all $t<\tau^n_\rho$. The martingale term vanishes due to the martingale property and the remaining integral is cancelled by the integral in the above expectation. Overall this shows that the first expectation vanishes.

Next we take the absolute value on both sides of the above equality and obtain, estimating the second expectation and extending the integration interval to $[0,T]$, the inequality

$$\begin{aligned}
&\Big|\mathbb{E}^n\Big[\mathbb{I}_A\Big(\langle\Phi,\zeta_{t_2\wedge\tau_\rho}\rangle^2_{\mathcal{E}}-\int_0^{t_2\wedge\tau_\rho}\langle\Phi,G(u(s),p(s))\Phi\rangle_{\mathcal{E}}\,\mathrm{d}s-\langle\Phi,\zeta_{t_1\wedge\tau_\rho}\rangle^2_{\mathcal{E}}\\
&\qquad\qquad+\int_0^{t_1\wedge\tau_\rho}\langle\Phi,G(u(s),p(s))\Phi\rangle_{\mathcal{E}}\,\mathrm{d}s\Big)\Big]\Big|\\
&\qquad\leq\ \int_0^t\mathbb{E}^n\Big|\alpha_n\,\langle\Phi,G^n(Y^n_s,\theta^n_s)\Phi\rangle_{\mathcal{E}}-\langle\Phi,G(u(s),p(s))\Phi\rangle_{\mathcal{E}}\Big|\,\mathrm{d}s\,.
\end{aligned}$$

The convergence of the upper bound to zero for $n\to\infty$ follows by assumption (4.3.6). Hence we have proven an analogous result to (4.4.19) in part (a). The same line of argument that concluded part (a) also concludes part (b). The proof is completed.

4.4.3 Proof of Theorem 4.3.2 (Central limit theorem II)

For the alternative version of the martingale central limit theorem, Theorem 4.3.2, the convergence of the finite-dimensional distributions is employed to characterise the limit as the diffusion process given by the covariance operator $C(t)$ given by (4.3.2). This is in contrast to the use of the local martingale problem for the proof of Theorem 4.3.1. The difference in conditions to the first martingale central limit theorem is that condition (D3') is assumed instead of condition (D3) and thus avoids the assumption that the jump heights of the martingale sequence are uniformly almost surely bounded. First of all note, that as we assume in Theorem 4.3.2 tightness for the sequence $(\sqrt{\alpha_n}\, M_t^n)_{t\geq 0}$, e.g., condition (D1) of Theorem 4.3.1 is satisfied, the existence of a limit is guaranteed.
As we have discussed in the introduction to Section 4.3 the limit is uniquely characterised by the convergence of the finite-dimensional distributions. Due to Lévy's Continuity Theorem, for an appropriate version in Hilbert spaces see [97], the convergence of the finite-dimensional distributions is implied by the convergence of their characteristic functions. In the present situation this task can be further simplified. Analogously to [75, 80, 96] it is sufficient to consider the convergence of the one-dimensional diffusion: Assume that $\sqrt{\alpha_n}\, M_t^n$ converges in distribution to the corresponding one-dimensional marginal of the limiting diffusion for all $t \geq 0$. Completely analogous arguments, as we employ below to establish the convergence of the one-dimensional distributions, establish that for each $t, s \geq 0$ the sequence of random variables $\sqrt{\alpha_n}\,(M_{s+t}^n - M_s^n)$ converges in distribution to the corresponding increment of the limiting diffusion. We can now use this property in combination with the strong Markov property of the sequence of PDMPs to infer that the characteristic function of the two-dimensional vector $(\sqrt{\alpha_n}\, M_s^n, \sqrt{\alpha_n}\, M_{s+t}^n)$ converges to the characteristic function of the corresponding two-dimensional vector of the limiting diffusion. As this holds for all $t, s \geq 0$ we have shown the convergence of the two-dimensional marginals. Repeating this procedure we are able to show the convergence of the finite-dimensional distributions.
Thus, in the following it remains to show the convergence of the one-dimensional characteristic functions of the martingales $(\sqrt{\alpha_n}\, M_t^n)_{t\geq 0}$ to the characteristic function of the one-dimensional marginals of the diffusion process. The latter are given for all $\Phi \in \mathcal{E}^*$ and all $t \geq 0$ by

$$\varphi(t, \Phi) = \exp\Big(-\frac{1}{2}\int_0^t \langle \Phi, G(u(s), p(s))\Phi\rangle_{\mathcal{E}}\, \mathrm{d}s\Big)\,. \tag{4.4.20}$$

The individual steps in the subsequent proof follow the outline of [80, 96] wherein finite-dimensional processes are considered. That this method of proof is also applicable for infinite-dimensional PDMPs relies on the general results for these processes we have derived in Chapter 2. Particularly the characterisation of the extended generator, cf. Theorem 2.2.1 and the resulting Dynkin's formula (2.2.7).

We split the remainder of the proof in two parts. In a first step (a) we derive a formula for characteristic function φ^n of the random variables $\sqrt{\alpha_n}\, M_t^n$, $t \geq 0$ and $n \in \mathbb{N}$. In the second step (b) we show that the conditions (D3') imply the convergence of φ^n to the characteristic function of the diffusion φ given in 4.4.20.

(a) We first derive a formula for the characteristic function of the one-dimensional marginals of the martingales $(\sqrt{\alpha_n}\, M_t^n)_{t\geq 0}$. To this end we consider for every $n \geq 1$ the process $(Y_t^n, \theta_t^n, M_t^n)_{t\geq 0}$ which is a PDMP on the state space $H \times K_n \times \mathcal{E}$. That is, the process $(Y_t^n, \theta_t^n, M_t^n)_{t\geq 0}$ is the original PDMP extended by components containing the associated martingales. By definition the components M_t^n are càdlàg and the jumps in the two discontinuous components θ^n and M^n occur at the same times. Moreover, the jump heights satisfy the relation $M^n_{\tau_n^k} - M^n_{\tau_n^k-} = z^n(\theta^n_\tau) - z^n(\theta^n_{\tau-})$ for τ_n^k denoting the kth jump time of the process $(Y_t^n, \theta_t^n, M_t^n)_{t\geq 0}$.
Let $f : K_n \times \mathcal{E} \to \mathbb{R}$ be a function in the domain of the extended generator of $(Y_t^n, \theta_t^n, M_t^n)_{t\geq 0}$, see Theorem 2.2.1, which is Fréchet-differentiable with respect to the second component taking values in $\mathcal{E}$. In the following this component is denoted by η. As M_t^n is absolutely continuous in $\mathcal{E}$ and hence almost everywhere differentiable, the generator of $(Y_t^n, \theta_t^n, M_t^n)_{t\geq 0}$ applied to such an f, cf. Theorem 2.2.1, is given by

$$
\begin{aligned}
\mathcal{A}^n f(\theta^n,\eta) \;=\;& \\
=\;& -\sum_{j=1}^m \big\langle f_{\eta_j}(\theta^n,\eta), \mathcal{A}^n\langle\,\cdot\,, z_j^n(\theta^n)\rangle_E\big\rangle_E \\
&+\Lambda^n(u,\theta^n)\int_{K_n}\Big[f(\xi, z^n(\xi)-z^n(\theta^n)+\eta) - f(\theta^n,\eta)\Big]\,\mu^n\big((u,\theta^n),\mathrm{d}\xi\big) \\
=\;& \Lambda^n(u,\theta^n)\int_{K_n}\Big[f(\xi, z^n(\xi)-z^n(\theta^n)+\eta) - f(\theta^n,\eta) \\
&\qquad - \big\langle z^n(\xi)-z^n(\theta), f_\eta(\theta^n,\eta)\big\rangle_{\mathcal{E}}\Big]\,\mu^n\big((u,\theta^n),\mathrm{d}\xi\big)\,.
\end{aligned}
\tag{4.4.21}
$$

Here the second equality holds as

$$
\big\langle f_{\eta_j}(\theta^n,\eta), \mathcal{A}^n\langle\,\cdot\,, z_j^n(\theta^n)\rangle_E\big\rangle_E = \mathcal{A}^n\langle f_{\eta_j}(\theta^n,\eta), z_j^n(\theta^n)\rangle_E\,.
$$

We next use the generator (4.4.21) to obtain a formula for the characteristic function.

The characteristic function φ^n of $\sqrt{\alpha_n}\, M_t^n$ is for all $t \geq 0$ and $\Phi \in \mathcal{E}^*$ defined by

$$\varphi^n(t,\Phi) = \mathbb{E}\, \mathrm{e}^{i\langle \Phi, \sqrt{\alpha_n} M_t^n\rangle_{\mathcal{E}}} \tag{4.4.22}$$

with $\langle \Phi, z\rangle_{\mathcal{E}} = \sum_{j=1}^m \langle \phi_j, z_j\rangle_E$ for any $z = (z_1, \ldots, z_m) \in \mathcal{E}$ and $\Phi = (\phi_1, \ldots, \phi_m) \in \mathcal{E}^*$. The chain rule of Fréchet differentiation yields that for all $j = 1, \ldots, m$ the Fréchet derivative of $y \mapsto \mathrm{e}^{i\langle \phi_j, y\rangle}$ at a point $x \in E$ is an element of E^* and given by

$$D\, e^{i\langle \phi_j,\,\cdot\,\rangle_E}[x] : E \to \mathbb{C} :\ h \mapsto i\langle \phi_j, h\rangle_E\, \mathrm{e}^{i\langle \phi_j, x\rangle_E}\,.$$

That is the mapping $(u, \theta^n, \eta) \mapsto \mathrm{e}^{i\langle \Phi, \eta\rangle_{\mathcal{E}}}$ is continuously Fréchet differentiable with respect to η. Moreover, it is also bounded as $|\mathrm{e}^{-ix}| = 1$ for all $x \in \mathbb{R}$ and thus it is in the domain of the extended generator of the process $(Y_t^n, \theta_t^n, M_t^n)_{t\geq 0}$.

Due to the Fréchet differentiability we now obtain from (4.4.21) and Dynkin's formula (2.2.7) that

$$\begin{aligned}\varphi^n(t,\Phi) - 1 \;&=\; \int_0^t \mathbb{E}^n\Big[\Lambda^n(Y_s^n,\theta_s^n) \int \mathrm{e}^{i\sqrt{\alpha_n}\langle \Phi, z^n(\xi) - z^n(\theta_s^n) + M_s^n\rangle_{\mathcal{E}}} - \mathrm{e}^{i\sqrt{\alpha_n}\langle \Phi, M_s^n\rangle_{\mathcal{E}}}\\ &\qquad - i\sqrt{\alpha_n}\, \mathrm{e}^{i\sqrt{\alpha_n}\langle \Phi, M_s^n\rangle_{\mathcal{E}}}\, \big\langle \Phi, z^n(\xi) - z^n(\theta_s^n)\big\rangle_{\mathcal{E}}\, \mu^n\big((Y_s^n,\theta_s^n), \mathrm{d}\xi\big)\Big]\,\mathrm{d}s.\end{aligned}$$

The definition of the operator G^n and the equality (4.1.14), that is,

$$\Lambda^n(u,\theta^n) \int \langle \Phi, z^n(\xi) - z^n(\theta^n)\rangle_{\mathcal{E}}^2\, \mu^n\big((u,\theta^n), \mathrm{d}\xi\big) \;=\; \langle \Phi, G^n(u,\theta^n)\,\Phi\rangle_{\mathcal{E}},$$

yields for the characteristic function φ^n the representation

$$\begin{aligned}\varphi^n(t,\Phi) - 1 \;&=\; \int_0^t \mathbb{E}^n\Big[-\frac{1}{2}\,\alpha_n\, \langle \Phi, G^n(Y_s^n,\theta_s^n)\,\Phi\rangle_{\mathcal{E}}\, \mathrm{e}^{i\sqrt{\alpha_n}\langle \Phi, M_s^n\rangle_{\mathcal{E}}}\Big]\mathrm{d}s + K_n(t,\Phi)\\ &=\; -\int_0^t \frac{1}{2}\big\langle \Phi, G(u(s),p(s))\,\Phi\big\rangle_{\mathcal{E}}\, \varphi_n(s,\Phi)\,\mathrm{d}s + J_n(t,\Phi) + K_n(t,\Phi)\,,\end{aligned} \tag{4.4.23}$$

where the terms $J_n(t,\Phi)$ and $K_n(t,\Phi)$ are given by

$$J_n(t,\Phi) := \frac{1}{2}\int_0^t \mathbb{E}^n\Big[\Big(\big\langle \Phi, G(u(s),p(s))\,\Phi\big\rangle_{\mathcal{E}} - \alpha_n \langle \Phi, G^n(Y_s^n,\theta_s^n)\,\Phi\rangle_{\mathcal{E}}\Big)\mathrm{e}^{i\sqrt{\alpha_n}\langle M_s^n, \Phi\rangle_{\mathcal{E}}}\Big]\,\mathrm{d}s$$

and

$$K_n(t,\Phi) \;:=\; \mathbb{E}^n \int_0^t \Big[\Lambda^n(Y^n_s,\theta^n_s)\int_{K_n} \mathrm{e}^{i\sqrt{\alpha_n}\langle\Phi,M^n_s\rangle_{\mathcal{E}}}\, k\Big(\sqrt{\alpha_n}\,\big\langle\Phi, z^n(\xi)-z^n(\theta^n_s)\big\rangle_{\mathcal{E}}\Big)$$
$$\alpha_n\,\big\langle\Phi, z^n(\xi)-z^n(\theta^n_s)\big\rangle^2_{\mathcal{E}}\,\mu^n\big((Y^n_s,\theta^n_s),\mathrm{d}\xi\big)\Big]\mathrm{d}s\,.$$

The function k in the definition of the term K_n is given by

$$k(y) = \frac{\mathrm{e}^{iy}-1-iy+\frac{1}{2}y^2}{y^2}\,.$$

Note that k is bounded and satisfies $k(y)=o(y)$.

(b) Next we prove the convergence of the characteristic functions. First of all observe that the characteristic function φ of the limiting diffusion given in (4.4.20) satisfies

$$\varphi(t,\Phi)-1 \;=\; -\int_0^t \frac{1}{2}\big\langle\Phi, G(u(s),p(s))\,\Phi\big\rangle_{\mathcal{E}}\,\varphi(s,\Phi)\mathrm{d}s\,.$$

Subtracting this equality from equality (4.4.23) we obtain

$$\varphi^n(t,\Phi)-\varphi(t,\Phi)$$
$$= J_n(t,\Phi)+K_n(t,\Phi)-\int_0^t\frac{1}{2}\big\langle\Phi, G(u(s),p(s)),\Phi\big\rangle_{\mathcal{E}}\big(\varphi^n(s,\Phi)-\varphi(s,\Phi)\big)\,\mathrm{d}s.$$

Next, taking the absolute value on both sides and estimating the right hand side – note that operator $G(u(s),p(s))$ is positive – we obtain the inequality

$$|\varphi^n(t,\Phi)-\varphi(t,\Phi)|$$
$$\le\; |J_n(t,\Phi)+K_n(t,\Phi)|+\frac{1}{2}\int_0^t\big\langle\Phi, G(u(s),p(s))\Phi\big\rangle_{\mathcal{E}}\big|\varphi^n(s,\Phi)-\varphi(s,\Phi)\big|\,\mathrm{d}s.$$

We use Gronwall's Lemma to infer

$$|\varphi^n(t,\Phi)-\varphi(t,\Phi)| \;\le\; \exp\bigg(\frac{1}{2}\int_0^t\big\langle\Phi, G(u(s),p(s))\Phi\big\rangle_{\mathcal{E}}\,\mathrm{d}s\bigg)\,\sup_{s\in[0,t]}\big|J_n(s,\Phi)+K_n(s,\Phi)\big|.$$

The exponential function in the right hand side of this inequality is finite due to condition (4.3.1). The proof is completed showing that the supremum terms, that is, the terms $\sup_{s\in[0,t]}|J_n(s,\Phi)|$ and $\sup_{s\in[0,t]}|K_n(t,\Phi)|$ converge to zero for $n\to\infty$.

On the one hand, the convergence of the first term $\sup_{s\in[0,t]}|J_n(s,\Phi)|$ follows immediately by condition (4.3.6) and the boundedness of the exponential function. On the other hand, for the term $K_n(t,\Phi)$ we employ condition (4.3.8) in (D3').

In more detail, as the exponential is bounded it remains to estimate the term

$$\alpha_n \mathbb{E}^n \int_0^t \Lambda^n(Y_s^n, \theta_s^n) \int_{K_n} |\langle \phi, z^n(\xi) - z^n(\theta^n)\rangle_E|^2 \\ \big|k(\sqrt{\alpha_n}\langle \phi, z^n(\xi) - z^n(\theta^n)\rangle_{\mathcal{E}})\big| \, \mu^n\big((u,\theta^n), \mathrm{d}\xi\big)\, \mathrm{d}s$$

$$\leq \; \overline{k}\, \alpha_n \, \mathbb{E}^n \int_0^t \Lambda^n(Y_s^n, \theta_s^n) \int_{\left\{\theta^n \in K_n \,\middle|\, \sqrt{\alpha_n}|\langle \phi, z^n(\xi) - z^n(\theta_s^n)\rangle_{\mathcal{E}}| > \beta_n\right\}} \\ |\langle \phi, z^n(\xi) - z^n(\theta_s^n)\rangle_{\mathcal{E}}|^2 \, \mu^n\big((Y_s^n, \theta_s^n), \mathrm{d}\xi\big)\, \mathrm{d}s$$

$$+ \, |k(\beta_n)| \, \alpha_n \, \mathbb{E}^n \int_0^t \Lambda^n(Y_s^n, \theta_s^n) \int_{\left\{\theta^n \in K_n \,\middle|\, \sqrt{\alpha_n}|\langle \phi, z^n(\xi) - z^n(\theta_s^n)\rangle_{\mathcal{E}}| \leq \beta_n\right\}} \\ |\langle \phi, z^n(\xi) - z^n(\theta_s^n)\rangle_{\mathcal{E}}|^2 \, \mu^n\big((Y_s^n, \theta_s^n), \mathrm{d}\xi\big)\, \mathrm{d}s$$

where $\overline{k}$ is a bound to $k(y)$. Now the first term in the right hand side converges to zero for $n \to \infty$ due to condition (4.3.8) in (D3'). The second term in the right hand side converges to zero as $|k(\beta_n)| = o(\beta_n)$, $\beta_n \to 0$, and the expectation is bounded for $n \in \mathbb{N}$ due to condition (4.3.3). The proof is completed.

4.5 Application to models of excitable membranes

As an example for the application of the limit theorems presented in Sections 4.2 and 4.3 we consider a compartmental-type stochastic model for spatially extended neuronal membranes, cf. Section 3.2. We note that the law of large numbers that is presented in [7] for a particular one-dimensional exact hybrid model serves as another example for the application of Theorem 4.2.1.

4.5.1 The setup for compartmental type models

We consider the equation for the membrane potential on a bounded domain $D \subset \mathbb{R}^d$ for the physically reasonable dimensions $d \leq 3$ with sufficiently smooth boundary. The most interesting choices for the applied sciences are, in particular, $d = 1$ and $d = 2$, as discussed elsewhere in more detail. One-dimensional domains are regularly used for models of axons, and $d = 2$ is employed in models of cardiac tissue and calcium dynamics. The family of abstract evolution equations defining the dynamics of the PDMP's continuous component Y^n are given for this model by the parabolic, linear, inhomogeneous second order partial differential equations

$$\dot{u} = \sum_{i,j=1}^{d} a_{ij}(x) u_{x_i x_j} + \sum_{i=1}^{m} g_i(x)\, z_i^n(\theta^n)\,(E_i - u) \tag{4.5.1}$$

where the coefficient functions a_{ij} and g_i are smooth on $\overline{D}$ and the differential operator is strongly elliptic. These non-constant values for the diffusion coefficients and the conductances account for spatially inhomogeneous models, e.g., myelinated axons. To conform with the preceding sections of the thesis we equip the equation (4.5.1) with Dirichlet boundary condition. This particular choice is of no particular importance for the considerations that follow and can be readily changed. We define the operators A, B depending on θ^n only via suitable coordinate functions $z^n(\theta^n)$, cf. (4.0.1) in Section 4.0.1, by

$$\begin{aligned} A(z^n(\theta^n))\,u &:= \sum_{i,j=1}^{d} a_{ij}(x) u_{x_i x_j} \\ B(z^n(\theta^n), u) &:= \sum_{i=1}^{m} g_i(x)\, z_i^n(\theta^n)\,(E_i - u)\,. \end{aligned} \tag{4.5.2}$$

We defer the precise definition of the coordinate functions z^n for compartmental-type models and first discuss the deterministic limit system. As the deterministic limit we propose the solution to the membrane equation

$$\dot{u} = \sum_{i,j=1}^{d} a_{ij}(x) u_{x_i x_j} + \sum_{i=1}^{m} g_i(x)\, p_i\,(E_i - u)\,, \tag{4.5.3}$$

with p_i, $i = 1\ldots, m$, given by solutions of the coupled equations

$$\dot{p}_i \;=\; F_i(p,u) \;:=\; \sum_{j\neq i} \Big(p_j\, q_{ji}(u) \;-\; p_i\, q_{ij}(u) \Big)\,. \tag{4.5.4}$$

This deterministic system is the usual deterministic model for axon equations coupled to the *full gating system*, cf. Section 3.3. Analogously as for the PDMP model we choose Dirichlet boundary conditions for the potential component, i.e., $u(t,x) = 0$ for all $t \in [0,T]$ and all $x \in \partial D$. Further, we assume that the rate functions satisfy the conditions (EC3) in Section 3.3.1. We choose initial conditions $u_0 \in H_0^1(D) \cap H^s(D)$ and $p_i(0) \in H^s(D)$ where $s > d/2$. These satisfy $u(0,x) \in [\overline{u}_-, \overline{u}_+]$ for all $x \in D$ and $p_i(0,x) \in [0,1]$, $\sum_{i=1}^m p_i(0,x) = 1$ for all $x \in D$ and $i = 1, \ldots, m$. Here we set $\overline{u}_- := \min_i E_i \leq 0$ and $\overline{u}_+ := \max_i E_i \geq 0$. Finally, in the case $d = 2, 3$ we assume that the domain possesses a C^3–boundary.
Then it holds that the deterministic system (4.5.3), (4.5.4) is well-posed due to Theorem 3.3.1. The unique solution (u,p) is in $C^0([0,T], H^s(D))$ componentwise and, in particular, is pointwise bounded due to Proposition 3.3.1, i.e., $u(t,x) \in [\overline{u}_-, \overline{u}_+]$ and $p_i(t,x) \in [0,1]$ for all $(t,x) \in [0,T] \times \overline{D}$ and all $i = 1, \ldots, m$.

We briefly recollect the essential features for the sequence of PDMPs $(Y_t^n, \theta_t^n)_{t\geq 0}$, $n \in \mathbb{N}$, defined on the probability spaces $(\Omega^n, \mathcal{F}^n, \mathbb{P}^n)$ being the compartmental-type models. First of all, an integral component of the sequence of models is the sequence of compartmentalisation. Thus, for each $n \in \mathbb{N}$ let $p(n) \in \mathbb{N}$ denote the number of compartments of the spatial domain D, i.e., for each n the sets $D_{k,n}, \dots D_{p(n),n}$ form a partition of the domain D which is the basis of the models. Moreover, we assume that each $D_{k,n}$ is convex. The convexity of individual compartments is a technical assumption which allows to employ Poincaré's inequality in the proof of the limit theorems with a known optimal Poincaré constant [1, 98]. Secondly, a fundamental aspect is the channel distribution across the compartments, that is, the coordinate functions $z^n(\theta^n)$ and the stochastic jump dynamics. Recall that in compartmental-type models the individual components $\theta_i^{k,n}(t)$ of the piecewise constant PDMP component θ_t^n denote the number of channels located in the domain $D_{k,n}$ which are in state i at time t. We assume that compartments either contain no channels or a fixed deterministic number. To this end let $l(n) \in \mathbb{N}$ denote this fixed, deterministic number of channels, i.e.,

$$\sum_{i=1}^{m} \theta_i^{k,n} = l(n)\,,$$

for k denoting compartments containing channels. Obviously for k denoting a compartment that does not contain channels it holds that $\theta_i^{k,n} = 0$ for all $i = 1, \dots, m$. Summarising we note that the states $\theta^n = (\theta_i^{k,n})_{i=1,\dots,m,\,k=1,\dots,p(n)}$ for the piecewise constant component of the PDMP are $mp(n)$-dimensional vectors which take only finitely many values the set of which is denoted by K_n. On the set K_n we define for $i = 1, \dots, m$ the coordinate functions

$$z_i^n(\theta^n) = \frac{1}{l(n)} \sum_{k=1}^{p(n)} \theta_i^{k,n}(t)\, \mathbb{I}_{D_{k,n}} \in L^2(D)\,. \tag{4.5.5}$$

The coordinate process $z^n(\theta_t^n)$ is càdlàg with each component taking values in $L^2(D)$. Moreover, each $z_i^n(\theta_t^n)$ is for every $t \geq 0$ a piecewise constant function on the spatial domain D taking values in $[0, 1]$.

As two channel switchings do not occur simultaneously, the only jumps in the piecewise constant component configuration θ_t^n with non-zero probability are transitions concerning one single channel. That is, these are events for which in one particular subdomain one particular channel changes its state. The rate that one channel in subdomain $D_{k,n}$ switches from state i to state j is given by

$$\theta_i^{k,n}\, Q_{ij}^{k,n}(u) \in \mathbb{R}_+\,, \tag{4.5.6}$$

where $Q^n_{ij}(u)$ is a functional of the transmembrane potential u defined as

$$Q^{k,n}_{ij}(u) := q_{ij}\Big(\frac{1}{|D_{k,n}|}\int_{D_{k,n}} u(x)\,\mathrm{d}x\Big).$$

That is, $Q^{k,n}_{ij}(u)$ is the instantaneous rate evaluated at the average value of the transmembrane potential on the subdomain $D_{k,n}$. Hence the instantaneous rate (4.5.6) is the number of channels in state i in domain $D_{k,n}$ times the rate of one channel switching from i to j. This definition yields, by summing over all events, the total instantaneous rate

$$\Lambda^n(Y^n_t,\theta^n_t) := \sum_{i=1,k=1}^{m,p(n)}\sum_{j=1}^{m}\theta^{k,n}_i(t)\,Q^{k,n}_{ij}(Y^n_t).$$

Note that the total instantaneous rate is proportional to the total number of channels in the model, i.e., $\Lambda^n = \mathcal{O}(p(n)l(n))$. Moreover, it is bounded, hence the PDMP is regular

Finally, we note that due to Theorem 3.2.1 the membrane variable of the stochastic model is almost everywhere pointwise bounded , i.e., $Y^n_t(x)\in[\overline{u}_-,\overline{u}_+]$ for almost all $x\in D$ and all $t\geq 0$, analogously to the deterministic limit u.

4.5.2 Limit theorems for compartmental-type models

The limit theorems derived in the preceding sections take the following form for the compartmental models. We recall that it is a general assumption for the hybrid models that the initial conditions satisfy for all $n\in\mathbb{N}$ almost surely that $Y^n_0(x)\in[\overline{u}_-,\overline{u}_+]$ for all $x\in D$.
We denote by $\delta(n)$ the maximal diameter of the elements of the partition corresponding to the nth model containing channels, i.e.,

$$\delta(n) = \max_{k=1,\ldots,p(n)} \mathrm{diam}(D_{k,n})\,\mathbb{I}_{\left[\sum_{j=1}^m \theta^{k,n}_j\neq 0\right]},$$

where the diameter of a set $D_{k,n}$ is defined as $\sup_{x,y\in D_{k,n}}|x-y|$. Obviously, the sequence is bounded by the diameter of the domain D. Moreover, for $\lim_{n\to\infty}\delta(n)=0$, which we always assume, the sequence of models is defined on ever finer partitions. It is clear that a vanishing diameter $\delta(n)\to 0$ for $n\to\infty$ implies that the number of compartments tends to infinity, i.e., $\lim_{n\to\infty}p(n)=\infty$. The law of large numbers in Theorem 4.5.1 provides a connection of the stochastic models to the deterministic excitable membrane system in the following sense: For neuronal membranes with a high channel density and if the contribution of an individual channel to the total conductance of small patches of membranes is small, the stochastic spatio-temporal dynamics of the membrane are close to the deterministic model with a high probability.

Theorem 4.5.1. *Assume that the sequence of partitions satisfy that*

$$\lim_{n\to\infty} \delta(n) = 0, \qquad \lim_{n\to\infty} l(n) = \infty, \tag{4.5.7}$$

and that the initial conditions $(Y_0^n, z^n(\theta_0^n))$ *converge in probability to* (u_0, p_0) *in the space* $L^2(D)^{m+1}$. *Then the compartmental-type models converge in probability to the deterministic solution of the excitable media system in the sense that it holds for all* $\epsilon > 0$ *that*

$$\lim_{n\to\infty} \mathbb{P}\Big[\sup_{t\in[0,T]} \|Y_t^n - u(t)\|_{L^2} + \sum_{i=1}^m \sup_{t\in[0,T]} \|z_i^n(\theta_t^n) - p(t)\|_{L^2} > \epsilon\Big] = 0\,. \tag{4.5.8}$$

Moreover, the convergence also holds in the mean in the space $L^2((0,T), L^2(D))$, *i.e.,*

$$\lim_{n\to\infty} \mathbb{E}^n\Big[\|Y_t^n - u(t)\|_{L^2((0,T),L^2)} + \sum_{i=1}^m \|z_i^n(\theta_t^n) - p(t)\|_{L^2((0,T),L^2)}\Big] = 0\,. \tag{4.5.9}$$

We proceed to present the appropriate quadratic variation process for the martingale central limit theorem applied to compartmental-type models. For the definition of the limiting diffusion we consider for $u, p_i \in C^0(\overline{D})$ the bilinear form

$$\begin{aligned}(\Psi,\Phi) \mapsto \big(G(u,p)\,\Psi,\Phi\big)_{L^2} \;=\; & \sum_{j=1}^m \sum_{i\neq j} \int_D p_i(x)\cdot q_{ij}(u(x))\cdot \psi_j(x)\cdot \phi_j(x)\,\mathrm{d}x \\ & + \sum_{j=1}^m \sum_{i\neq j} \int_D p_j(x)\cdot q_{ji}(u(x))\cdot \psi_j(x)\cdot \phi_j(x)\,\mathrm{d}x \\ & - \sum_{j=1}^m \sum_{i\neq j} \int_D p_j(x)\cdot q_{ji}(u(x))\cdot \psi_i(x)\cdot \phi_j(x)\,\mathrm{d}x \\ & - \sum_{j=1}^m \sum_{i\neq j} \int_D p_i(x)\cdot q_{ij}(u(x))\cdot \psi_i(x)\cdot \phi_j(x)\,\mathrm{d}x\,.\end{aligned} \tag{4.5.10}$$

Note that this is finite for all $\phi_i, \psi_i \in L^2(D)$ as p_i and $q_{ij}(u)$ are bounded functions. Hence, for every given $\Psi \in L^2(D)^m$ the mapping $\Phi \mapsto (G(u,p)\Psi, \Phi)$ is a linear, bounded functional on $L^2(D)^m$ and, conversely, for every given $\Phi \in L^2(D)^m$ the mapping $\Psi \mapsto (G(u,p)\Psi, \Phi)$ is a linear, bounded functional on $L^2(D)^m$.

Proposition 4.5.1. *The operator* $G(u,p)$ *defined via* (4.5.10) *is a trace class operator mapping* $H^s(D)$ *into* $H^{-s}(D)$ *for* $s > d/2$. *Moreover, the operator-valued mapping* $t \mapsto G(u(t), p(t))$ *defines a unique centred diffusion process on* $H^{-s}(D)$.

Finally for the subsequent central limit theorem we define

$$\nu_+(n) := \max_{0\le k\le n} |D_{k,n}|\, \mathbb{I}_{\left[\sum_m \theta_m^{k,n} \neq 0\right]},$$

i.e., $\nu_+(n)$ is the maximum volume of compartments that contain channels. Analogously we define $\nu_-(n)$ being the minimum volume of compartments containing channels. Finally, note that in the following the coordinate functions z^n are considered as maps from K_n into the space $H^{-2s}(D)$.

Theorem 4.5.2. *Let s be the smallest integer such that $s > d/2$. If in addition to* (4.5.7) *and the convergence of the initial conditions the sequence of partitions satisfies*

$$\lim_{n\to\infty} \frac{\nu_-(n)}{\nu_+(n)} = 1 \tag{4.5.11}$$

then the sequence of $H^{-2s}(D)$–valued martingales $\left(\sqrt{\frac{l(n)}{\nu_+(n)}}\, M_t^n\right)_{t\ge 0}$ converges weakly to the $(H(D)^{-2s})^m$–valued diffusion defined by (4.5.10).

Remark 4.5.1. We close this section with brief remarks concerning two aspects of the Theorems 4.5.1 and 4.5.2: the sequence of compartmentalisations and the sequence of initial conditions.
First of all, the conditions guarantee that spatial areas where there are no channels do not matter for the limit behaviour. In the present example, where only spatially homogeneous channel distributions are considered for the limiting model, this is not such a crucial aspect. However, it gains in importance when inhomogeneous structures are considered where there are large membrane areas without channels. For example, this concerns models of myelinated neurons, where channels are clustered at Nodes of Ranvier, or in Calcium dynamics, where channels also regularly form individual clusters on the membrane. The fact that the behaviour of compartments that do not contain channels do not affect the limiting behaviour conforms with the intuitive understanding of a limit in connection with the underlying biological physiology. Next, on the one hand, the condition (4.5.7) in the law of large numbers is intuitively quite accessible: It essentially allows that a sequence of step functions defined via the compartments $(D_{k,n})_{n=1,\dots,p(n)}$, $n \in \mathbb{N}$, taking only finitely many values in $[0,1]$ converges to a continuous function taking values in $[0,1]$. On the other hand, condition (4.5.11) in the central limit theorem seems more technical. Essentially, it imposes the necessity that the sequence of compartmentalisations is homogeneous in the sense that the size of the compartments has to be roughly uniform. Particularly, condition (4.5.11) is not satisfied if for every compartmentalisation there is one compartment which, say, is only half the size as another.
Next, we note that for a large class of domains $D \subset \mathbb{R}^d$ a sequence of compartmen-

talisations that satisfies the conditions of the Theorems 4.5.1 and 4.5.2 exist. Let D be Jordan measurable and $\mathcal{T}_n$ the smallest cover of D by d-dimensional cubes with edges of length $1/n$. We obtain a compartmentalisation of D by the elements of $\mathcal{T}_n$ intersected with D. Then we place channels only in those compartments which are a complete cube. Obviously, the first condition in (4.5.7) and condition (4.5.11) are satisfied. Note that due to the Jordan measurability of D the sequence of subset of D which is generated along its boundary by a union of compartments not containing channels is such that its Lebesgue measure converges to zero. This is important for the necessary convergence of the sequence of initial conditions on which comment next.

In particular, we argue that for all initial conditions of the deterministic limit system a sequence converging in probability trivially exist showing that there exists a suitable sequence of deterministic initial conditions. Firstly, for the membrane variable we can simply use $Y_0^n \equiv u_0$ for all $n \in \mathbb{N}$. Secondly, for the gating variables $p_i(0)$ this is not possible as $z_i^n(\theta)$ is a step function in $L^2(D)$ by definition but higher spatial regularity is assumed for $p_i(0)$. However, for every $p_i(0)$ an approximating sequence of appropriate step functions is easily found. We set

$$\theta_i^{k,n}(0) = \operatorname{argmin}_{j=1,\dots l(n)} \Big| \frac{j}{l(n)} - \frac{1}{|D_{n,k}|} \int_{D_{n,k}} p_i(x)\,\mathrm{d}x \Big|\,.$$

Then we obtain using Poincaré's inequality and denoting the average of $p_i(0)$ over the domain $D^{k,n}$ by $p_i^{D_{k,n}}$ that

$$\begin{aligned} \|z_i^n(\theta_0^n) - p_i(0)\|_{L^2}^2 &= \sum_{k=1}^{p(n)} \int_{D_{n,k}} \Big| \frac{\theta_i^{k,n}(0)}{l(n)} - p_i(0,x) \Big|^2 \mathrm{d}x \\ &\leq 2 \sum_{k=1}^{p(n)} \int_{D_{n,k}} \Big| \frac{\theta_i^{k,n}(0)}{l(n)} - p_i^{D_{n,k}} \Big|^2 + \big| p_i^{D_{n,k}} - p_i(0,x) \big|^2 \mathrm{d}x \\ &\leq 2 \sum_{k=1}^{p(n)} \frac{|D_{n,k}|}{l(n)^2} + \pi^{-2} \delta(n)^2 \|\nabla p_i(0)\|_{L^2(D_{n,k})}^2 \\ &= 2 \frac{|D|}{l(n)^2} + 2\pi^{-2} \delta(n)^2 \|\nabla p_i(0)\|_{L^2}^2, \end{aligned}$$

which converges to zero under the assumptions (4.5.7). Such a sequence can be defined for all $i = 1, \dots, m-1$, and to define $\theta_m^{k,n}$ we use the condition $\sum_{i=1}^m \theta_i^{k,n} = l(n)$. Then, the m sequences converge simultaneously to the respective initial conditions $p_i(0)$. We note that it is clear that the convergence still holds if for each partition there is a collection of compartments not containing channels the measure of which converges to zero for $n \to \infty$. This is the case, e.g., for the sequence of partitions for Jordan measurable domains and corresponding channel distributions we have discussed above.

4.5.3 Proofs of Theorem 4.5.1 (Conditions for the LLN)

We apply Theorem 4.2.1 for the choice of spaces $X = H_0^1(D)$, $H = L^2(D)$ and $E = L^2(D)$. Hence, we have to prove in the following that the assumptions therein are satisfied, i.e., (i) the one-sided Lipschitz condition (4.0.2) on the operators A and B defined by (4.5.2), (ii) the Lipschitz condition on the right hand side of the gating system (4.5.4), (iii) the uniform convergence of the generator and (iv) the martingale convergence. Finally, in (v) we extend the convergence in probability due to Theorem 4.2.1 to convergence in the mean (4.5.9). For simplicity of presentation we assume that all compartments $D_{k,n}$ contain channels. If otherwise, the proof is completely analogous the only difference is that every summation over $k = 1, \ldots, p(n)$ has to be replaced by a summation over only those indices k that refer to a compartment containing channels.

(i) We start considering the one-sided Lipschitz conditions for operators A, B of the evolution equation governing the inter-jump behaviour. For the nonlinear operator B we find that the left hand side in the Lipschitz condition is for almost all t given by a finite sum of terms

$$\langle p_i \cdot (E_i - u) - \widehat{p}_i \cdot (E_i - v), u - v \rangle_{H^1}, \tag{4.5.12}$$

with u, $v \in H_0^1(D)$ and p_i, $\widehat{p}_i \in L^2(D)$. Hence, the duality pairing corresponds to the inner product in $L^2(D)$. We estimate each of the summands of the type (4.5.12) separately. Using the triangle inequality we obtain

$$\big|\langle p_i \cdot (E_i - u) - \cdot \widehat{p}_i \cdot (E_i - v), u - v \rangle_{H^1}\big| \leq |E_i| \, \big|(p_i - \widehat{p}_i, u - v)_{L^2}\big| + \big|(p_i \cdot u - \widehat{p}_i \cdot v, u - v)_{L^2}\big| \, .$$

The first term in the right hand side of this inequality is further estimated using Cauchy-Schwarz and Young's inequality which yield

$$\big|(p_i - \widehat{p}_i, u - v)_{L^2}\big| \leq \tfrac{1}{2}\big\|p_i - \widehat{p}_i\big\|_{L^2}^2 + \tfrac{1}{2}\big\|u - v\big\|_{L^2}^2 \, .$$

For the second term we obtain, making use of the triangle inequality, the Cauchy-Schwarz and Young's inequality and the pointwise bounds on p_i and v, the sequence of estimates

$$\begin{aligned}\big|(p_i \cdot u - \widehat{p}_i \cdot v, u - v)_{L^2}\big| &\le \big|(p_i \cdot (u-v), u-v)_{L^2}\big| + \big|(p_i - \widehat{p}_i, v \cdot (u-v))_{L^2}\big| \\ &\le \big\|p_i \cdot (u-v)\big\|_{L^2}\big\|u-v\big\|_{L^2} + \big\|p_i - \widehat{p}_i\big\|_{L^2}\big\|v \cdot (u-v)\big\|_{L^2} \\ &\le \big\|u-v\big\|^2_{L^2} + \tfrac{\overline{u}^2}{2}\big\|u-v\big\|^2_{L^2} + \tfrac{1}{2}\big\|p_i - \widehat{p}_i\big\|^2_{L^2}\,.\end{aligned}$$

A summation over all these estimates for $i = 1, \ldots m$ yields

$$\langle B(p,u) - B(\widehat{p}, v), u - v\rangle_{H^1} \le m\big(1 + \tfrac{\overline{u}+\overline{u}^2}{2}\big)\,\|u-v\|^2_{L^2} + \frac{1+\overline{u}}{2}\sum_{i=1}^m \big\|p_i - \widehat{p}_i\big\|^2_{L^2}\,.$$

Adding the estimate

$$\langle A(u-v), u-v\rangle_{H^1} \le -\gamma_1\|u-v\|^2_{H^1} + \gamma_2\|u-v\|^2_{L^2} \le \gamma_2\|u-v\|^2_{L^2}$$

for some $\gamma_1, \gamma_2 > 0$, which holds as the linear operator A is coercive and independent of p, we obtain

$$\langle A(u-v), u-v\rangle_{H^1} + \langle B(p,u) - B(\widehat{p}, v), u-v\rangle_{H^1} \le C\Big(\|u-v\|^2_{L^2} + \sum_{i=1}^m \big\|p_i - \widehat{p}_i\big\|^2_{L^2}\Big)$$

for a suitable constant C. Finally, integrating over $(0,T)$ we find the one-sided Lipschitz condition (4.0.2) is satisfied.

(ii) Next we consider the Lipschitz condition (4.2.3) for the deterministic limit (4.5.4). Due to the triangle inequality it suffices to consider differences of the form $\|p_i \cdot q(u) - \widehat{p}_i \cdot q(v)\|_{L^2}$, where q substitutes for an arbitrary rate function q_{jk}. Using the triangle inequality, the pointwise boundedness of $\widehat{p}_i$ and q by 1 and $\overline{q}$, respectively, and the Lipschitz condition on the rate functions q (with common Lipschitz constant L) we obtain

$$\begin{aligned}\|p_i \cdot q(u) - \widehat{p}_i \cdot q(v)\|_{L^2} &\le \|p_i \cdot q(u) - \widehat{p}_i \cdot q(u)\|_{L^2} + \|\widehat{p}_i \cdot q(u) - \widehat{p}_i \cdot q(v)\|_{L^2} \\ &\le \overline{q}\,\|p_i - \widehat{p}_i\|_{L^2} + L\,\|u-v\|_{L^2}\,.\end{aligned}$$

A summation over all such separate estimates, integrating and squaring both resulting sides yield the Lipschitz condition (4.2.3).

(iii) In order to prove the convergence of the generators (4.2.5) we employ in the following two technical results which we collect in a separate proposition. Firstly, the purpose of the formula (4.5.13) is to transform the generator of the PDMP into a form that allows comparison with the deterministic limit system (4.5.4). Secondly, the inequality (4.5.14), which bounds the norm $\|Y^n\|_{L^2((0,T),H^1)}$ by a deterministic constant uniformly over $n \in \mathbb{N}$, is used repeatedly in the subsequent estimation procedures.

Proposition 4.5.2. (a) *The generator of the PDMP satisfies*

$$\Lambda^n(Y_t^n,\theta_t^n)\int_{K_n}\Big(z_i^n(\xi)-z_i^n(\theta_t^n)\Big)\,\mu^n\big((u,\theta^n),\mathrm{d}\xi\big)=\sum_{j\neq i}\Big(z_j^n(\theta_t^n)\cdot q_{ji}^n(Y_t^n)-z_i^n(\theta_t^n)\cdot q_{ij}^n(Y_t^n)\Big) \tag{4.5.13}$$

where

$$q_{ij}^n(Y_t^n)=\sum_{k=1}^{p(n)}Q_{ij}^{k,n}(Y_t^n)\,\mathbb{I}_{D_{k,n}}\in L^2(D)\,.$$

(b) *For all $n\in\mathbb{N}$ and all $T>0$ it holds that*

$$\int_0^T\|Y_t^n\|_{H^1}^2\,\mathrm{d}t\le C_1(1+T)\mathrm{e}^{2C_2T}, \tag{4.5.14}$$

where the constants C_1, C_2 are deterministic and independent of $n\in\mathbb{N}$.

Proof. (a) We denote by $\theta_{k,i\to j}^n(t)$ for all $k=1,\dots,p(n)$ and all $i\neq j$, $i,j=1,\dots m$, the configuration in K_n that arises from the configuration θ_t^n through the event that a channel in state i located in domain $D_{k,n}$ switches to state j. Then simple reorganisation of finite sums yields

$$\begin{aligned}
&\Lambda^n(Y_t^n,\theta_t^n)\int_{K_n} z_i^n(\xi)-z_i^n(\theta_t^n)\,\mu^n\big((Y_t^n,\theta_t^n),\mathrm{d}\xi\big)\\
&=\sum_{k=1}^{p(n)}\sum_{j\neq i}\Big(z_i^n(\theta_{k,j\to i}^n(t))-z_i^n(\theta_t^n)\Big)\,\theta_j^{k,n}\,Q_{ji}^{k,n}(Y_t^n)\\
&\quad+\sum_{k=1}^{p(n)}\sum_{j\neq i}\Big(z_i^n(\theta_{k,i\to j}^n(t))-z_i^n(\theta_t^n)\Big)\,\theta_i^{k,n}\,Q_{ij}^{k,n}(Y_t^n)\\
&=\sum_{k=1}^{p(n)}\sum_{j\neq i}\Big(\frac{1}{l(n)}\mathbb{I}_{D_{k,n}}\Big)\,\theta_j^{k,n}(t)\,Q_{ji}^{k,n}(Y_t^n)+\sum_{k=1}^{p(n)}\sum_{j\neq i}\Big(-\frac{1}{l(n)}\mathbb{I}_{D_{k,n}}\Big)\,\theta_i^{k,n}(t)\,Q_{ij}^{k,n}(Y_t^n)\\
&=\sum_{j\neq i}z_j^n(\theta_t^n)\cdot\Big(\sum_{k=1}^{p(n)}Q_{ji}^{k,n}(Y_t^n)\,\mathbb{I}_{D_{k,n}}\Big)-\sum_{j\neq i}z_i^n(\theta_t^n)\cdot\Big(\sum_{k=1}^{p(n)}Q_{ij}^{k,n}(Y_t^n)\,\mathbb{I}_{D_{k,n}}\Big).
\end{aligned}$$

Thus we obtain that the generator satisfies (4.5.13).

(b) By definition of a PDMP it holds that the component $(Y_t^n)_{t\geq0}$ is the weak solution of the evolution equation

$$\dot{Y}_t^n = AY_t^n + \sum_{i=1}^m g_i\, z_i^n(\theta_t^n)\,(E_i - Y_t^n)$$

with initial condition Y_0^n. Therefore using an estimation procedure as in Section 3.4.3, which yields inequality (3.4.58) in that section, we obtain

$$\int_0^T \|Y_t^n\|_{H^1}^2\,\mathrm{d}t \leq K_1 \mathrm{e}^{2K_2T}\Big(\|Y_0^n\|_{L^2}^2 + 2\overline{u}^2\sum_i \|g_i\|_{L^\infty}\int_0^T \|z_i^n(\theta_t^n)\|_{L^1}\,\mathrm{d}t\Big),$$

where the constants K_1, K_2 are deterministic and depend only on the domain D and the coefficients of A. Further, by definition $\|z_i^n(\theta_t^n)\|_{L_1} \leq |D|$ and the sequence of initial conditions is bounded by assumption as $Y_0^n(x) \in [\overline{u}_-, \overline{u}_+]$ for all $x \in \overline{D}$ almost surely. The inequality (4.5.14) follows.

□

We now proceed to the actual proof of the convergence (4.2.5). To this end we need to consider for almost every t and all $i = 1, \ldots, m$ the convergence in $L^2(D)$ of (4.5.13) to $F_i(z^n(\theta_t^n), Y_t^n)$ where F_i is as defined in (4.5.4). That is, we have to estimate

$$\Big\|\sum_{j\neq i}\Big(z_j^n(\theta^n)\cdot q_{ji}^n(Y_t^n) - z_i^n(\theta^n)\cdot q_{ij}^n(Y_t^n)\Big) - \sum_{j\neq i}\Big(z_j^n(\theta^n)\cdot q_{ji}(Y_t^n) - z_i^n(\theta)\cdot q_{ij}(Y_t^n)\Big)\Big\|_{L^2}. \tag{4.5.15}$$

We find that the single summands in the two summations match up and thus it suffices to consider each of them separately. Employing the boundedness of the coordinate functions, i.e., $\|z_j^n(\theta^n)\|_{L^\infty} \leq 1$ we obtain the estimate

$$\begin{aligned}
\big\|z_j^n(\theta_t^n)\cdot q_{ji}^n(Y_t^n) - z_j^n(\theta_t^n)\cdot q_{ji}(Y_t^n)\big\|_{L^2}^2 &= \big\|z_j^n(\theta_t^n)\big\|_{L^\infty}^2 \big\|q_{ji}^n(Y_t^n) - q_{ji}(Y_t^n)\big\|_{L^2}^2 \\
&\leq \Big\|\sum_{k=1}^{p(n)} \mathbb{I}_{D_{k,n}} Q_{ji}^{k,n}(Y_t^n) - q_{ji}(Y_t^n)\Big\|_{L^2}^2 \\
&= \sum_{k=1}^{p(n)} \int_{D_{k,n}} \big|Q_{ji}^{k,n}(Y_t^n) - q_{ji}(Y_t^n(x))\big|^2\,\mathrm{d}x\,.
\end{aligned} \tag{4.5.16}$$

For the last equality we have used that

$$\Big(\sum_{k=1}^{p(n)} Q_{ji}^{k,n}(Y_t^n)\,\mathbb{I}_{D_{k,n}}(x)\Big) - q_{ji}(Y_t^n(x)) = \sum_{k=1}^{p(n)}\Big(Q_{ji}^{k,n}(Y_t^n) - q_{ji}(Y_t^n(x))\Big)\,\mathbb{I}_{D_{k,n}}(x),$$

and that the summands are mutually orthogonal in $L^2(D)$. Next we estimate each of the remaining integrals in (4.5.16) using the Lipschitz continuity of q_{ji} and Poincaré's inequality in $L^2(D_{k,n})$, i.e.,

$$\int_{D_{k,n}} \Big|q_{ji}\Big(\frac{1}{|D_{k,n}|}\int_{D_{k,n}} Y_t^n(y)\,\mathrm{d}y\Big) - q_{ji}(Y_t^n(x))\Big|^2 \mathrm{d}x$$

$$\leq\; L^2 \int_{D_{k,n}} \Big|\frac{1}{|D_{k,n}|}\int_{D_{k,n}} Y_t^n(y)\,\mathrm{d}y - Y_t^n(x)\Big|^2 \mathrm{d}x$$

$$\leq\; L^2\pi^{-2}\mathrm{diam}(D_{k,n})^2\,\|\nabla Y_t^n\|^2_{L^2(D_{k,n})},$$

where $\|\nabla Y_t^n\|_{L^2}$ is the norm in $L^2(D)$ of the Euclidean norm of the gradient vector ∇Y_t^n. Here we employed that for convex domains the optimal Poincaré constant is given by $\pi^{-1}\mathrm{diam}(D_{k,n})$ [98]. Hence, a summation over all $k = 1,\ldots,p(n)$ and employing that $\|\nabla Y_t^n\|^2_{L^2} \leq \|Y_t^n\|_{H^1}$, yields

$$\big\|q_{ji}^n(Y_t^n) - q_{ji}(Y_t^n)\big\|^2_{L^2} \leq \delta(n)^2\,L^2\pi^{-2}\|Y_t^n\|^2_{H^1}\,.$$

Integrating over $(0,T)$ we therefore obtain for (4.5.15) the estimate

$$\int_0^T \Big\|\big[\mathcal{A}(\phi, z_i^n(\cdot))_{L^2}\big](Y_t^n,\theta_t^n) - F_i(z^n(\theta_t^n), Y_t^n)\Big\|^2_{L^2}\,\mathrm{d}t$$

$$\leq\; \delta(n)^2\,L^2\pi^{-2}\,2(m-1)\int_0^T \|Y_t^n\|^2_{H^1}\,\mathrm{d}t.$$

Finally, the norm $\|Y^n\|_{L^2((0,T),H^1)}$ is bounded independently of $n\in\mathbb{N}$ by a deterministic constant due to Proposition 4.5.2(b). This upper bound then holds for almost all paths of the PDMPs $(Y_t^n,\theta_t^n)_{t\geq 0}$. Thus there exists a constant $C>0$ independent of n such that

$$\sup_{u,\theta^n}\int_0^T \Big\|\big[\mathcal{A}(\phi, z_i^n(\cdot))_{L^2}\big](u(t),\theta^n) - F_i(z^n(\theta^n), u(t))\Big\|^2_{L^2}\,\mathrm{d}t \;\leq\; \delta(n)^2\,C\,, \qquad (4.5.17)$$

where the supremum is taken over a set that contains almost all paths of the PDMP. Due to the assumption (4.5.7) the estimate in the right hand side converges to zero for $n\to\infty$ and the convergence (4.2.5) follows.

(iv) Next we consider convergence in probability of the martingale part. To this end we employ Lemma 4.1.1. As before we denote by $\theta^n_{k,i\to j}(t)$ the channel configuration that arises from the configuration θ_t^n if a channel in compartment k switches from state i to state j. Then it holds that

$$\Lambda^n(Y_t^n,\theta_t^n)\int_{K_n}\|z_i^n(\xi)-z_i^n(\theta_t^n)\|_{L^2}^2\,\mu^n\big((Y_t^n,\theta_t^n),\mathrm{d}\xi\big)$$

$$= \sum_{k=1}^{p(n)}\sum_{j\neq i}\Big(\|z_i^n(\theta_{k,i\to j}^n(t))-z_i^n(\theta_t^n)\|_{L^2}^2\,Q_{ij}^{k,n}(Y_t^n)\,\theta_i^n(t)$$

$$+\|z_i^n(\theta_{k,j\to i}^n(t))-z_i^n(\theta_t^n)\|_{L^2}^2\,Q_{ji}^{k,n}(Y_t^n)\,\theta_j^n(t)\Big)$$

$$\leq \overline{q}\sum_{k=1}^{p(n)}\frac{|D_{k,n}|}{l(n)^2}\sum_{j\neq i}\big(\theta_i^n(t)+\theta_j^n(t)\big).$$

This implies that

$$\mathbb{E}^n\int_0^T\Big[\Lambda^n(Y_t^n,\theta_t^n)\int_{K_n}\|z_i^n(\xi)-z_i^n(\theta_t^n)\|_{L^2}^2\,\mu^n\big((Y_t^n,\theta_t^n),\mathrm{d}\xi\big)\Big]\,\mathrm{d}s=\mathcal{O}\big(l(n)^{-1}\big).$$

Hence, under condition (4.5.7) the condition of Lemma 4.1.1 is satisfied and the martingale convergence follows.

(v) Finally, we extend the convergence in probability to convergence in the mean for the individual components being in the space $L^2((0,T),L^2)$, see the remark following Theorem 4.2.1. First of all note that the components are bounded, i.e.,

$$\|Y_t^n-u(t)\|_{L^2}\leq 2\overline{u}\,|D|,\qquad \|z_i^n(\theta_t^n)-p_i(t)\|_{L^2}\leq 2\,|D|\,.$$

Therefore it holds that

$$\|X^n-X\|:=\|Y^n-u\|_{L^2((0,T),L^2)}+\sum_{i=1}^m\|z_i^n(\theta^n)-p_i\|_{L^2((0,T),L^2)}\ \leq\ C$$

for a suitable deterministic bound $C<\infty$ independent of $n\in\mathbb{N}$. Then for all $\epsilon_0>0$ it holds that

$$\mathbb{E}^n\|X_n-X\| \;=\; \mathbb{E}^n\big[\|X_n-X\|\,\mathbb{I}_{[\|X_n-X\|\leq\epsilon_0]}\big]+\mathbb{E}^n\big[\|X_n-X\|\,\mathbb{I}_{[\|X_n-X\|>\epsilon_0]}\big]$$

$$\leq\ \epsilon_0+M\,\mathbb{P}^n\big[\|X_n-X\|>\epsilon_0\big]\,.$$

Next choose $\epsilon_0<\epsilon/2$ and note that due to the convergence in probability there exists an N_ϵ such that $M\,\mathbb{P}^n\big[\|X_n-X\|>\epsilon_0\big]\leq\epsilon/2$ for all $n>N_\epsilon$. Hence, for every $\epsilon>0$ there exists an N_ϵ such that $\mathbb{E}^n\|X_n-X\|<\epsilon$ for all $n>N_\epsilon$. Convergence in the mean is proven.

4.5.4 Proof of Theorem 4.5.2 (Conditions for the CLT)

In order to prove Theorem 4.5.2 we employ Theorem 4.3.1 for the space $E = H^{-2s}(D)$ where s is the smallest integer such that $s > d/2$. We usually employ the simpler notation E and $\mathcal{E} = E^m$ throughout the proof however occasionally switch to $H^{-2s}(D)$ if we want to emphasise the specific choice of the Hilbert space. The reason choosing this particular s is that it is the smallest integer such that the embedding of $H^{2s}(D)$ into $H^s(D)$ is of Hilbert-Schmidt type[11] due to Maurin's Theorem [2, Thm. 6.61] and $H^s(D)$ is embedded in $C^0(\overline{D})$ due to the Sobolev Embedding Theorem. These two properties are essential in order to prove the conditions (4.3.3) – (4.3.7) of Theorem 4.3.1. All conditions except (4.3.6), which establishes the convergence of the quadratic variation, are straightforward consequences of the assumptions of the theorem. These are shown in part (i) of the subsequent proof. For condition (4.3.6) more involved estimation procedures are necessary which are presented in part (ii). Analogously to the proof of Theorem 4.5.1 we assume for simplicity of presentation that all compartments $D_{k,n}$ contain channels and the adaptations to the proof for the general case are as explained in the proof of Theorem 4.5.1.

(i) We first show condition (4.3.3). Recall that $\theta^n_{k,i\to j}(t)$ denotes the element of K_n that differs from θ^n_t by one channel in the kth compartment being in state i instead of state j. Then, the Sobolev Embedding Theorem yields the estimate

$$\|z^n_i(\theta^n_{k,i\to j}(t)) - z^n(\theta^n_t)\|_E = \sup_{\|\phi\|_{H^{2s}}=1} \big| l(n)^{-1}\langle \phi, \mathbb{I}_{D^{k,n}}\rangle_{H^{2s}}\big| \leq \frac{C}{l(n)}\,|D^{k,n}|, \qquad (4.5.18)$$

where C is a constant resulting from the continuous embedding of $H^{2s}(D)$ into $C^0(\overline{D})$. Using this estimate for the jump heights in the space $H^{-2s}(D)$ we find similarly to part (iv) of the proof of Theorem 4.5.1 that it holds

$$\alpha_n\,\mathbb{E}^n \int_0^T \Big[\Lambda^n(Y^n_t,\theta^n_t)\int_{K_n} \|z^n(\xi) - z^n(\theta^n_t)\|^2_{\mathcal{E}}\,\mu^n\big((Y^n_t,\theta^n_t),\mathrm{d}\xi\big)\,\mathrm{d}t\Big] = \mathcal{O}(1)\,.$$

Hence, condition (4.3.3) is satisfied. Moreover, we infer from (4.5.18) that the rescaled jump sizes are bounded almost surely uniformly, i.e., condition (4.3.7) is satisfied. Particularly, it holds that $\sqrt{\alpha_n}\,\|z^n_i(\theta^n_{k,i\to j}(t)) - z^n_i(\theta^n_t)\|_E = \mathcal{O}(l(n)^{-1/2})$. This implies that for arbitrary $\beta > 0$ and any $\Phi \in (H^{2s}(D))^m$ there exists an N_β such that for all $n \geq N_\beta$

$$\int_{\sqrt{\alpha_n}|\langle \Phi, z^n_i(\xi) - z^n_i(\theta^n)\rangle_{\mathcal{E}}|>\beta} \mu^n\big((u,\theta^n),\mathrm{d}\xi\big) \;=\; 0$$

holds for all values (u,θ^n) the PDMP attains. Therefore, due to the dominated conver-

[11] The embedding of a Hilbert space X into another Hilbert space H is of *Hilbert-Schmidt type* if $\sum_{k\in\mathbb{N}} \|\varphi_k\|^2_H < \infty$ for every orthonormal basis $(\varphi_k)_{k\in\mathbb{N}}$ of X.

gence theorem we infer that also condition (4.3.5) is satisfied. It remains to consider condition (4.3.4). To this end let $(\varphi_k)_{k\in\mathbb{N}}$ be an orthonormal basis in $(H^{2s}(D))^m$ where $\varphi_k = (\varphi_k^1,\dots,\varphi_k^m)$ and hence $(\varphi_k^i)_{k\in\mathbb{N}}$ is an orthonormal basis in $H^{2s}(D)$ for all $i=1,\dots,m$. Then we obtain the estimate

$$\begin{aligned}
&\langle\varphi_k, G^n(Y_t^n,\theta_t^n)\varphi_k\rangle_{\mathcal{E}}\\
&\quad= \Lambda^n(Y_t^n,\theta_t^n)\int_{K_n}\Big(\sum_{i=1}^m\langle\varphi_k^i, z_i^n(\xi)-z_i^n(\theta_t^n)\rangle_{H^{-2s}}\Big)^2\mu^n\big((Y_t^n,\theta_t^n),\mathrm{d}\xi\big)\\
&\quad\le m\sum_{i=1}^m\|\varphi_k^i\|_{H^s}^2\Big(\Lambda^n(Y_t^n,\theta_t^n)\int_{K_n}\|z_i^n(\xi)-z_i^n(\theta_t^n)\|_{H^{-s}}^2\,\mu^n\big((Y_t^n,\theta_t^n),\mathrm{d}\xi\big)\Big).
\end{aligned}$$

Here we have employed for the single summands in the right hand side that for $z_i^n(\xi)-z_i^n(\theta^n)\in L^2(D)$ the duality pairing in $H^{2s}(D)$ equals the duality pairing in $H^s(D)$. Further, note that $\|z_i^n(\theta_{k,i\to j}^n(t))-z_i^n(\theta_t^n)\|_{H^{-s}}$ satisfies an estimate analogous to (4.5.18) due to the continuous embedding of $H^s(D)$ in $C^0(\overline{D})$. Therefore we overall obtain that

$$\alpha_n\,\langle\varphi_k, G^n(Y_t^n,\theta_t^n)\varphi_k\rangle_{\mathcal{E}}\le C\sum_{i=1}^m\|\varphi_k^i\|_{H^s}$$

for a suitable non-random constant C independent of n. Finally, set $\gamma_k:=\sum_{i=1}^m\|\varphi_k^i\|_{H^s}$ then it holds that $\sum_{k\in\mathbb{N}}\gamma_k<\infty$ as the embedding $H^{2s}(D)\hookrightarrow H^s(D)$ is of Hilbert-Schmidt type. We infer that condition (4.3.4) is satisfied.

(ii) In this second part of the proof we show that also the central condition (4.3.6) holds, which is the convergence of the quadratic variation. For simplicity of notation we omit the time argument t of the PDMP paths and the deterministic solution and note that the following estimates hold for almost all t. First of all we expand the quadratic variation of the martingales into the finite sum

$$\begin{aligned}
&\Lambda^n(Y^n,\theta^n)\int_{K_n}\langle\Phi, z^n(\xi)-z^n(\theta^n)\rangle_{\mathcal{E}}^2\,\mu^n\big((Y^n,\theta^n),\mathrm{d}\xi\big)\\
&\quad=\Lambda^n(Y^n,\theta^n)\int_{K_n}\sum_{j=1}^m\langle\phi_j, z_j^n(\xi)-z_j^n(\theta^n)\rangle_E^2\,\mu^n\big((Y^n,\theta^n),\mathrm{d}\xi\big)\\
&\qquad+\Lambda^n(Y^n,\theta^n)\int_{K_n}\sum_{\substack{i,j=1\\ i\neq j}}^m\langle\phi_j, z_j^n(\xi)-z_j^n(\theta^n)\rangle_E\langle\phi_i, z_i^n(\xi)-z_i^n(\theta^n)\rangle_E\,\mu^n\big((Y^n,\theta^n),\mathrm{d}\xi\big)\\
&\quad=\sum_{j=1}^m\sum_{\substack{i=1\\ i\neq j}}^m\sum_{k=1}^{p(n)}\Big(\frac{\theta_j^{k,n}}{l(n)^2}\,Q_{ji}^{k,n}(Y^n)+\theta_i^{k,n}\,Q_{ij}^{k,n}(Y^n)\Big)\,\langle\phi_j,\mathbb{I}_{D_{k,n}}\rangle_E^2\\
&\qquad-\sum_{\substack{i,j=1\\ i\neq j}}^m\sum_{k=1}^{p(n)}\Big(\frac{\theta_i^{k,n}}{l(n)^2}\,Q_{ij}^{k,n}(Y^n)+\theta_j^{k,n}\,Q_{ji}^{k,n}(Y^n)\Big)\,\langle\phi_i,\mathbb{I}_{D_{k,n}}\rangle_E\langle\phi_j,\mathbb{I}_{D_{k,n}}\rangle_E.
\end{aligned}$$

We find that the terms in this summation match with the integral terms in the definition of the operator $G(u,p)$ in (4.5.10). Thus, due to the triangle inequality it suffices to consider the convergence of the single summands separately. That is we consider, on the one hand, for all $j = 1,\ldots,m$ and $i \neq j$ the differences

$$\Big|\int_D p_j(x)\, q_{ji}(u(x))\, \phi_j^2(x)\,\mathrm{d}x \;-\; \alpha_n \sum_{k=1}^{p(n)} \frac{\theta_j^{k,n}}{l(n)^2}\, Q_{ji}^{k,n}(Y^n)\langle \phi_j, \mathbb{I}_{D_{k,n}}\rangle_E^2\Big|\,, \tag{4.5.19}$$

and, on the other hand, for all $i,j = 1,\ldots,m$ such that $i \neq j$ the differences

$$\Big|\int_D p_i(x)\, q_{ij}(u)\, \phi_i(x)\phi_j(x)\,\mathrm{d}x - \alpha_n \sum_{k=1}^{p(n)} \frac{\theta_i^{k,n}}{l(n)^2}\, Q_{ij}^{k,n}(Y^n)\langle \phi_i, \mathbb{I}_{D_{k,n}}\rangle_E\langle \phi_j, \mathbb{I}_{D_{k,n}}\rangle_E\Big|\,. \tag{4.5.20}$$

We next estimate these terms separately in parts (ii.1) and (ii.2). Finally, in part (ii.3) the estimates are combined to prove the convergence of the quadratic variation.

(ii.1) A further application of the triangle inequality to the term (4.5.19) yields

$$\begin{aligned}
&(4.5.19)\\
&= \Big|\int_D p_j(x)\, q_{ji}(u(x))\, \phi_j^2(x)\,\mathrm{d}x \;-\; \int_D z_j^n(\theta^n)(x)\, q_{ji}(Y^n(x))\, \phi_j^2(x)\,\mathrm{d}x\Big|\\
&\quad + \Big|\sum_{k=1}^{p(n)} \frac{\theta_j^{k,n}}{l(n)} \int_{D_{k,n}} q_{ji}(Y^n(x))\, \phi_j^2(x)\,\mathrm{d}x \;-\; \alpha_n \sum_{k=1}^{p(n)} \frac{\theta_j^{k,n}}{l(n)^2}\, Q_{ji}^{k,n}(Y^n)\langle \phi_j, \mathbb{I}_{D_{k,n}}\rangle_E^2\Big|\,.
\end{aligned} \tag{4.5.21}$$

We estimate the two resulting differences separately and obtain for the first term in the right hand side of (4.5.21) that

$$\begin{aligned}
&\Big|\int_D p_j(x)\, q_{ji}(u(x))\, \phi_j^2(x)\,\mathrm{d}x \;-\; \int_D z_j^n(\theta^n)(x)\, q_{ji}(Y^n(x))\, \phi_j^2(x)\,\mathrm{d}x\Big|\\
&\qquad \leq \Big|\int_D p_j(x)\, q_{ji}(u(x))\, \phi_j^2(x)\,\mathrm{d}x - \int_D z_j^n(\theta^n)(x)\, q_{ji}(u(x))\, \phi_j^2(x)\,\mathrm{d}x\Big|\\
&\qquad\quad + \Big|\int_D z_j^n(\theta^n)(x)\, q_{ji}(u(x))\, \phi_j^2(x)\,\mathrm{d}x - \int_D z_j^n(\theta^n)(x)\, q_{ji}(Y^n(x))\, \phi_j^2(x)\,\mathrm{d}x\Big|\\
&\qquad \leq \overline{q}\,\|\phi_j\|_{L^\infty}^2\,\|p_j - z_j^n(\theta^n)\|_{L^1} \;+\; L\,\|\phi_j\|_{L^\infty}^2\,\|u - Y^n\|_{L^1}\,.
\end{aligned} \tag{4.5.22}$$

For the second term in the right hand side of (4.5.21) we obtain by using $\theta_j^{k,n}/l(n) \leq 1$

the estimate

$$\sum_{k=1}^{p(n)} \Big| \int_{D_{k,n}} q_{ji}(Y^n(x))\, \phi_j^2(x)\, \mathrm{d}x - \frac{\alpha_n}{l(n)}\, q_{ji}\Big(\frac{1}{|D_{k,n}|}\int_{D_{k,n}} Y^n(x)\,\mathrm{d}x\Big)\Big(\int_{D_{k,n}} \phi_j(x)\,\mathrm{d}x\Big)^2\Big|. \tag{4.5.23}$$

We continue estimating each summand therein separately. We begin employing the Mean Value Theorem to expand the rate function q_{ij} in the integral in the left hand side such that

$$q_{ji}(Y^n(x)) = q_{ji}\Big(\frac{1}{|D_{k,n}|}\int_{D_{k,n}} Y^n(y)\,\mathrm{d}y\Big) + q'_{ji}(\vartheta^{k,n}(x))\Big(Y^n(x) - \frac{1}{|D_{k,n}|}\int_{D_{k,n}} Y^n(y)\,\mathrm{d}y\Big), \tag{4.5.24}$$

where $\vartheta^{k,n}(x)$ denotes an appropriate mean value. For now we omit the remainder term, i.e., the second term in the right hand side of (4.5.24), a consideration of which is deferred. Hence, we obtain for the absolute value in each summand in (4.5.23) the estimate

$$q_{ji}\Big(\frac{1}{|D_{k,n}|}\int_{D_{k,n}} Y^n(y)\,\mathrm{d}y\Big)\Big|\int_{D_{k,n}} \phi_j^2(x)\,\mathrm{d}x - \frac{\alpha_n}{l(n)}\Big(\int_{D_{k,n}} \phi_j(x)\,\mathrm{d}x\Big)^2\Big|.$$

We note that q_{ji} is bounded by $\overline{q}$ and continue to estimate

$$\begin{aligned}
&\le\ \overline{q}\,|D_{k,n}|\,\Big|\frac{1}{|D_{k,n}|}\int_{D_{k,n}} \phi_j^2(x)\,\mathrm{d}x - \frac{\alpha_n|D_{k,n}|^2}{l(n)|D_{k,n}|}\Big(\frac{1}{|D_{k,n}|}\int_{D_{k,n}} \phi_j(x)\,\mathrm{d}x\Big)^2\Big| \\
&\le\ \overline{q}\int_{D_{k,n}}\Big(\phi_j(x) - \frac{1}{|D_{k,n}|}\int_{D_{k,n}} \phi_j(y)\,\mathrm{d}y\Big)^2\mathrm{d}x \qquad (4.5.25)\\
&\qquad + \overline{q}\,|D_{k,n}|\,\Big|\Big(1 - \frac{\alpha_n|D_{k,n}|^2}{l(n)|D_{k,n}|}\Big)\Big(\frac{1}{|D_{k,n}|}\int_{D_{k,n}} \phi_j(x)\,\mathrm{d}x\Big)^2\Big|. \qquad (4.5.26)
\end{aligned}$$

The term (4.5.25) is now estimated by Poincaré's inequality which yields an upper bound by $\overline{q}\,\pi^{-2}\,\mathrm{diam}^2(D_{k,n})\,\|\nabla\phi\|^2_{L^2(D_{k,n})}$. For the term (4.5.26) a summation over all $k = 1,\dots,p(n)$ yields

$$\overline{q}\sum_{k=1}^{p(n)} |D_{k,n}|\,\Big|1 - \frac{\alpha_n|D_{k,n}|^2}{l(n)|D_{k,n}|}\Big|\Big(\frac{1}{|D_{k,n}|}\int_{D_{k,n}} \phi_j(x)\,\mathrm{d}x\Big)^2 \le \overline{q}\,\Big|1 - \frac{\nu_-(n)}{\nu_+(n)}\Big|\,\|\phi_j^n\|^2_{L^2}\,, \tag{4.5.27}$$

where ϕ_j^n is a piecewise constant approximation to ϕ_j defined by

$$\phi_j^n := \sum_{k=1}^{p(n)}\Big(\frac{1}{|D_{k,n}|}\int_{D_{k,n}} \phi_j(x)\,\mathrm{d}x\Big)\,\mathbb{I}_{D_{k,n}}\,.$$

As ϕ_j^n converges to ϕ_j in $L^2(D)$ it holds that the sequence of norms converge, hence

$\|\phi_j^n\|_{L^2}$ is a bounded sequence. Therefore the right hand side in (4.5.27) is a componentwise product of convergent sequences. The sequence $|1-\nu_-(n)/\nu_+(n)|$ converges to zero, cf. condition (4.5.11), thus the right hand side in (4.5.27) converges to zero for $n \to \infty$.

Finally, it remains to consider the term arising from the remainder in the expansion of q_{ji}, see (4.5.24), inserted into (4.5.23). By assumption q'_{ji} is bounded (by a constant $\overline{q}$). Therefore we obtain an upper bound on the respective term by

$$\begin{aligned}
&\overline{q}\,\|\phi_j\|^2_{L^\infty}\sum_{k=1}^{p(n)}\int_{D_{n,k}}\Big|Y^n(x)-\frac{1}{|D_{k,n}|}\int_{D_{k,n}}Y^n(y)\,\mathrm{d}y\Big|\,\mathrm{d}x\\
&\leq\ \overline{q}\,\|\phi_j\|^2_{L^\infty}\sum_{k=1}^{p(n)}\frac{\delta(n)}{2}\|\nabla Y^n\|_{L^1(D_{n,k})}\\
&\leq\ \overline{q}\,\|\phi_j\|^2_{L^\infty}\frac{\delta(n)}{2}\|\nabla Y^n\|_{L^1}\,,
\end{aligned}$$

where we have employed the Poincaré inequality in L^1 with the optimal Poincaré constant is given by $\mathrm{diam}(D_{k,n})/2$ [1].

A combination of the estimates yields an estimate to (4.5.19) by

$$(4.5.19)\ \leq\ C_\Phi\Big(\|p_j-z_j^n(\theta^n)\|_{L^1}+\|u-Y^n\|_{L^1}+\delta(n)^2\|\nabla Y^n\|_{L^1}+\delta^2(n)+\delta(n)+R(n)\Big)\,,$$

where the term $R(n)$ is given by the right hand side of (4.5.27) and converges to zero for $n\to\infty$. The constant $C_\Phi<\infty$ is a suitable deterministic constant independent of $n\in\mathbb{N}$ which depends on $\Phi\in(H^{2s}(D))^m$ via the norm in $H^s(D)$ of the components of Φ.

(ii.2) Next we consider the mixed terms (4.5.20). Analogously to part (ii.1) we apply the triangle inequality and obtain

$$\begin{aligned}
(4.5.20)\ \leq\ &\Big|\int_D p_j(x)\,q_{ji}(u(x))\,\phi_j(x)\,\phi_i(x)\,\mathrm{d}x-\int_D z_j^n(\theta^n)(x)\,q_{ji}(Y^n(x))\,\phi_j(x)\,\phi_i(x)\,\mathrm{d}x\Big|\\
&+\Big|\sum_{k=1}^{p(n)}\frac{\theta_j^{k,n}}{l(n)}\int_{D_{k,n}}q_{ji}(Y^n(x))\,\phi_j(x)\,\phi_i(x)\,\mathrm{d}x\\
&\qquad-\ \alpha_n\sum_{k=1}^{p(n)}\frac{\theta_j^{k,n}}{l(n)^2}\,Q_{ji}^{k,n}(Y^n)\langle\phi_j,\mathbb{I}_{D_{k,n}}\rangle_E\langle\phi_i,\mathbb{I}_{D_{k,n}}\rangle_E\Big|.
\end{aligned}$$

As in (ii.1) we obtain for the first term in this right hand side an upper bound by

$$\overline{q}\,\|\phi_i\|_{L^\infty}\|\phi_j\|_{L^\infty}\|p_j - z_j^n(\theta^n)\|_{L^1} + L\,\|\phi_i\|_{L^\infty}\|\phi_j\|_{L^\infty}\|u - Y^n\|_{L^1}\,.$$

Also the second term is treated as in (ii.1), i.e., applying the Mean Value Theorem and estimating the resulting terms accordingly. In particular the remainder term is estimated completely analogously. Therefore, the only term we are left to estimate is

$$\overline{q}\,|D_{k,n}|\,\Big|\frac{1}{|D_{k,n}|}\int_{D_{k,n}}\phi_i(x)\,\phi_j(x)\,\mathrm{d}x - \Big(\frac{1}{|D_{k,n}|}\int_{D_{k,n}}\phi_i(x)\,\mathrm{d}x\Big)\Big(\frac{1}{|D_{k,n}|}\int_{D_{k,n}}\phi_j(x)\,\mathrm{d}x\Big)\Big| \tag{4.5.28}$$

$$+\,\overline{q}\,|D_{k,n}|\,\Big|\Big(1-\frac{\alpha_n|D_{k,n}|^2}{l(n)|D_{k,n}|}\Big)\Big(\frac{1}{|D_{k,n}|}\int_{D_{k,n}}\phi_i(x)\,\mathrm{d}x\Big)\Big(\frac{1}{|D_{k,n}|}\int_{D_{k,n}}\phi_j(x)\,\mathrm{d}x\Big)\Big|. \tag{4.5.29}$$

First of all, using Young's inequality we obtain for the second term the estimate

$$(4.5.29)\;\le\;\frac{\overline{q}}{2}\Big|1-\frac{\nu_-(n)}{\nu_+(n)}\Big|\,\big(\|\phi_i^n\|^2_{L^2}+\|\phi_j^n\|^2_{L^2}\big), \tag{4.5.30}$$

which converges to zero for $n\to\infty$.

We next estimate the term (4.5.28). Firstly, we note that as in part (a) we find using Poincaré's inequality an upper bound to the term

$$|D_{k,n}|\,\Big|\frac{1}{D_{k,n}}\int_{D_{k,n}}\big(\phi_i(x)-\phi_j(x)\big)^2\,\mathrm{d}x - \Big(\frac{1}{D_{k,n}}\int_{D_{k,n}}\phi_i(x)-\phi_j(x)\,\mathrm{d}x\Big)^2\Big| \tag{4.5.31}$$

and the upper bound is proportional to $\delta(n)^2$. Next, expanding the two squared terms in (4.5.31) we find using the reverse triangle inequality that the term (4.5.31) is an upper bound to

$$\begin{aligned}
|D_{k,n}|\,\Big|\Big|\frac{1}{D_{k,n}}\int_{D_{k,n}}\phi_i(x)^2\,\mathrm{d}x + \frac{1}{D_{k,n}}\int_{D_{k,n}}\phi_j(x)^2\,\mathrm{d}x&\\
-\Big(\frac{1}{D_{k,n}}\int_{D_{k,n}}\phi_i(x)\,\mathrm{d}x\Big)^2 - \Big(\frac{1}{D_{k,n}}\int_{D_{k,n}}\phi_j(x)\,\mathrm{d}x\Big)^2\Big|&\\
-\,2\Big|\frac{1}{|D_{k,n}|}\int_{D_{k,n}}\phi_i(x)\phi_j(x)\,\mathrm{d}x - \Big(\frac{1}{D_{k,n}}\int_{D_{k,n}}\phi_i(x)\,\mathrm{d}x\Big)\Big(\frac{1}{D_{k,n}}\int_{D_{k,n}}\phi_j(x)\,\mathrm{d}x\Big)\Big|\Big|&.
\end{aligned}$$

Thus also this term possesses an upper bound which is proportional to $\delta(n)^2$. For $n\to\infty$ the upper bound converges to zero. As for $\delta(n)\to 0$ also the term spanning the first and second line converges to zero which was established in (ii.1), necessarily also the term in the third line converges to zero. Therefore we infer that the term (4.5.28) converges to zero proportional to $\delta(n)^2$.

Now, a combination of these estimates yields analogously to (4.5.4) in (ii.1) that

$$(4.5.20) \;\leq\; C_\Phi \Big(\|p_j - z_j^n(\theta^n)\|_{L^2} + \|u - Y^n\|_{L^2} + \delta(n)^2 \|\nabla Y^n\|_{L^1} + \delta^2(n) + \delta(n) + R(n) \Big). \tag{4.5.32}$$

Here $R(n)$ is a term converging to zero for $n \to \infty$ arising from (4.5.30) and it is of the same type as the term $R(n)$ in (ii.1). The deterministic constant C_Φ is independent of $n \in \mathbb{N}$ and depends on Φ via the norm in $H^s(D)$ of the components of Φ.

(ii.3) A combination of the final results (4.5.4) and (4.5.32) in (ii.1) and (ii.2) yields that there exists a constant $C_\Phi < \infty$ such that for almost all t it holds that

$$\begin{aligned} &\big|\langle \Phi, G(u(t), p(t))\, \Phi\rangle_{\mathcal{E}} - \alpha_n \big\langle \Phi, G^n(Y_t^n, \theta_t^n)\, \Phi\big\rangle_{\mathcal{E}}\big| \\ &\leq C_\Phi \Big(\sum_{i=1}^m \|p_i(t) - z_i^n(\theta_t^n)\|_{L^2} + \|u(t) - Y_t^n\|_{L^2} + \delta(n)^2 \big(1 + \|\nabla Y_t^n\|_{L^2}\big) + \delta(n) + R(n) \Big). \end{aligned}$$

Here we have also employed the continuous embedding of $L^2(D) \hookrightarrow L^1(D)$. We next square both sides of this inequality and integrate over $(0,T)$. Afterwards we take the square root of the integral terms and further take the expectation of the resulting inequality. Finally, appropriate applications of Jensen's inequality yields that

$$\begin{aligned} \int_0^T \mathbb{E}^n \Big| \langle \Phi, G(u(t), p(t))\, \Phi\rangle_{\mathcal{E}} - \alpha_n \big\langle \Phi, G^n(Y_t^n, \theta_t^n)\, \Phi\big\rangle_{\mathcal{E}} \Big| \, \mathrm{d}s \\ \leq\; C_{\Phi,T} \Big(\delta^2(n) + \delta(n) + R(n) \\ + \mathbb{E}^n \Big[\|u - Y^n\|_{L^2((0,T),L^2)} + \textstyle\sum_{i=1}^m \|p_i - z_i^n(\theta^n)\|_{L^2((0,T),L^2)} \Big] \Big) \end{aligned} \tag{4.5.33}$$

for an appropriate constant $C_{T,\Phi} < \infty$. Note that in order to arrive at the estimate (4.5.33) we have further employed that the random term $\|\nabla Y^n\|_{L^2((0,T),L^2)}$ can be estimated by a deterministic bound independent of $n \in \mathbb{N}$ due to Proposition 4.5.2 (b). Finally, due to the law of large numbers, i.e., Theorem 4.5.1, the sequence of PDMPs converges to the deterministic limit in the mean. Hence the expectation in the right hand side in (4.5.33) converges to zero for $n \to \infty$. Furthermore, $\delta(n)$ converges to zero by assumption (4.5.7), as does the term $R(n)$. Thus, overall the right hand side in (4.5.33) converges to zero. The convergence of the quadratic variation is proved which completes the proof of Theorem 4.5.1.

4.6 Section Appendix

4.6.1 Proof of Theorem 4.1.1 (Itô-isometry)

In this proof we show that under condition (4.1.3) the processes M_j^n, $j = 1, \ldots, m$, $n \in \mathbb{N}$, defined in (4.1.1) are square-integrable, càdlàg martingales which satisfy the Itô-isometry (4.1.4). Throughout the proof we fix a $j = 1, \ldots, m$ and $n \in \mathbb{N}$ and the results holds for any such j and n. Therefore, speaking of a PDMP in the following always refers to the PDMP $(Y_t^n, \theta_t^n)_{t\geq 0}$ corresponding to the fixed n. Further, for notational simplicity we omit the indices n and j discriminating processes and characteristics of PDMPs, i.e., M_j^n and $z_j^n(\theta^n)$ are denoted simply by M and $z(\theta)$. Finally, recall that τ_k, $k = 1, 2, \ldots$, denotes the sequence of increasing random jump times of the PDMP which are stopping times satisfying $\lim_{k\to\infty} \tau_k = \infty$ almost surely.

First of all, note that the process M is càdlàg by definition. The proof of the remaining open results is split into three parts. In the first, part (a), we prove the martingale property for the real process $(\langle \phi, M(t)\rangle_E)_{t\geq 0}$ for every $\phi \in E^*$. Then, the first main statement of Theorem 4.1.1, the square-integrability of the process $M(t)$, is proved in part (b). Moreover, as square-integrability implies integrability, the Hilbert space martingale property follows. Finally, the second main statement, the Itô-Isometry (4.1.4), is established in part (c). The proof we present in part (b) is motivated by the proof of [64, Prop. 4.5.3] which states the corresponding results for real-valued martingales associated with PDMPs, cf. Proposition B.2.4. In the extension of the method of proof employed therein to the present setup one has to ensure, on the one hand, that the employed results and estimation procedures all have corresponding analoga in the infinite-dimensional setting. On the other hand, one has to carefully make sure that only the weaker regularity results available in infinite-dimensions are used. Finally, the introduction of random initial conditions also necessitates some adaptations.

(a) First note that for all $\phi \in E^*$ the real-valued processes $\langle \phi, M(t)\rangle_E$ satisfy

$$\begin{aligned}\langle \phi, M(t)\rangle_E \;=\; & \langle \phi, z(\theta_t)\rangle_E - \langle \phi, z(\theta_0)\rangle_E \\ & - \int_0^t \Lambda(Y_{s-}, \theta_{s-}) \int_K \langle \phi, z(\xi)\rangle_E - \langle \phi, z(\theta_{s-})\rangle_E \, \mu\big((Y_{s-}, \theta_{s-}), \mathrm{d}\xi\big)\, \mathrm{d}s\,.\end{aligned} \tag{4.6.1}$$

Equation (4.6.1) is obtained from (4.1.2) due to the regularity of the PDMP as the set of jump times in $[0, t]$ is almost surely finite for all $t \geq 0$. Therefore the integrands in the right hand sides of (4.1.2) and (4.6.1) differ only on a set of Lebesgue measure zero almost surely. Moreover, the integrand in the right hand side of (4.6.1) has the form of the extended generator, cf. Theorem 2.2.1, applied to the map

$$\langle \phi, \cdot \rangle_E : K_n \to \mathbb{R} : \theta \mapsto \langle \phi, z(\theta) \rangle_E \tag{4.6.2}$$

which is indepndent of u. Thus it follows that the process $\langle \phi, M(t) \rangle_E$ is a local martingale if the map (4.6.2) is in the domain of the extended generator. Obviously, path-differentiability almost everywhere, cf. condition (i) in Theorem 2.2.1, is trivially satisfied as the mapping $t \mapsto \langle \phi, z(\theta_t) \rangle_E$ is piecewise constant. Furthermore, as there are no jumps upon hitting a boundary, the boundary condition (ii) in Theorem 2.2.1 is redundant. Hence, for (4.6.2) to be in the domain of the extended generator it remains to consider the integrability condition, cf. (iii) in Theorem 2.2.1, for which it is a sufficient condition that the expectation

$$\mathbb{E} \int_0^t \Lambda(Y_{s-}, \theta_{s-}) \int_K \big| \langle \phi, z(\xi) - z(\theta_{s-}) \rangle_E \big| \, \mu\big((Y_{s-}, \theta_{s-}), \mathrm{d}\xi\big) \, \mathrm{d}s \tag{4.6.3}$$

is finite for all $t > 0$, cf. Section B.2.4. Using Young's inequality we obtain an upper bound to (4.6.3) by

$$\begin{aligned} &\frac{1}{2} \mathbb{E} \int_0^t \Lambda(Y_{s-}, \theta_{s-}) \, \mathrm{d}s \\ &\qquad + \frac{1}{2} \mathbb{E} \int_0^t \Lambda(Y_{s-}, \theta_{s-}) \int_K \big| \langle \phi, z(\xi) - z(\theta_{s-}) \rangle_E \big|^2 \, \mu\big((Y_{s-}, \theta_{s-}), \mathrm{d}\xi\big) \, \mathrm{d}s \, . \end{aligned}$$

Here the first expectation in the right hand side is finite due to the PDMP being regular and the second is finite by an immediate consequence of assumption (4.1.3).

Next, we show that the process is not only a local martingale but even a martingale. As we have mentioned above the process $\langle \phi, M(t) \rangle_E$ satisfies

$$\langle \phi, M(t) \rangle_E = \int_0^t \int_K \langle \phi, z(\xi) - z(\theta_{s-}) \rangle_E \, \widetilde{M}(\mathrm{d}s, \mathrm{d}\xi) \, ,$$

where $\widetilde{M} := N - \widehat{N}$ is the random martingale measure associated with the PDMP with counting measure N and compensator $\widehat{N}(\mathrm{d}\xi, \mathrm{d}s) = \Lambda(Y_{s-}, \theta_{s-}) \, \mu\big((Y_{s-}, \theta_{s-}), \mathrm{d}\xi\big) \, \mathrm{d}s$. We have just shown above that the process $\langle \phi, z(\xi) - z(\theta_{s-}) \rangle_E$ is a valid integrand for $\widetilde{M}$. Thus the process $\langle \phi, M(t) \rangle_E$ has the form of a stochastic integral with respect to the martingale measure associated with the PDMP. Furthermore, due to Theorem B.2.6 it holds that the process is a martingale if (4.6.3) is finite for all $t \geq 0$. We have already shown that this holds due to the regularity of the PDMP and assumption 4.1.3.

(b) We now prove the square-integrability of the processes M. In a first step we prove in (b.1) that M stopped at the first jump τ_i is square-integrable. Subsequently in part (b.2) this result is extended to M stopped at any jump time τ_k, $k \in \mathbb{N}$. Then we are able to infer square-integrability of the process M. As square-integrability implies integrability it follows from the results in part (a) that M is a Hilber space valued martingale.

(b.1) Note that prior to τ_1 the θ-component of the PDMP remains constant. We introduce the notation

$$\widetilde{N}(s) := \int_0^s \Lambda(Y_r,\theta_0)\int_K z(\xi) - z(\theta_0)\,\mu\big((Y_r,\theta_0),\mathrm{d}\xi\big)\,\mathrm{d}r$$

which implies that $s \mapsto \|\widetilde{N}(s)\|_E^2$ is almost surely absolutely continuous with derivative

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}\|\widetilde{N}(s)\|_E^2 &= 2\big(\tfrac{\mathrm{d}}{\mathrm{d}t}\widetilde{N}(s),\widetilde{N}(s)\big)_E\\ &= 2\,\Lambda(Y_s,\theta_0)\int_K \big(z(\xi)-z(\theta_0),\widetilde{N}(s)\big)_E\,\mu\big((Y_s,\theta_s),\mathrm{d}\xi\big)\,. \qquad (4.6.4)\end{aligned}$$

Due to the structure of a PDMP we obtain for the conditional expectation with respect to the initial condition

$$\begin{aligned}\mathbb{E}\big[\|M(\tau_1\wedge t)\|_E^2\,|\,\mathcal{F}_0\big] &= \|\widetilde{N}(t)\|_E^2\,\exp\Big(-\int_0^t \Lambda(Y_r,\theta_0)\,\mathrm{d}r\Big)\\ &\quad+\int_0^t\Big[\int_K \big\|z(\vartheta)-z(\theta_0)-\widetilde{N}(s)\big\|_E^2\,\mu\big((Y_s,\theta_0),\mathrm{d}\vartheta)\big)\Big]\,\Lambda(Y_s,\theta_0)\\ &\qquad\exp\Big(-\int_0^s \Lambda(Y_r,\theta_0)\,\mathrm{d}r\Big)\,\mathrm{d}s\,.\end{aligned}$$

That is, the first term in the right hand side is the position of the stopped process $\|M(\tau_1\wedge t)\|_E^2$ at time t if $t<\tau_1$ times the conditional probability that the first jump does not occur before t. The second term is its position after the jump integrated over the conditional density that a jump occurs in $[0,t]$. We apply integration by parts to the first term (note that $\widetilde{N}(0)=0$) and find that

$$\begin{aligned}&\|\widetilde{N}(t)\|_E^2\,\exp\Big(-\int_0^t \Lambda(Y_s,\theta_0)\,\mathrm{d}s\Big)\\ &\quad=\int_0^t\Big[2(\tfrac{\mathrm{d}}{\mathrm{d}t}\widetilde{N}(s),\widetilde{N}(s))_E\,\exp\Big(-\int_0^s \Lambda(Y_r,\theta_0)\,\mathrm{d}r\Big)\\ &\qquad\qquad-\|\widetilde{N}(s)\|_E^2\,\Lambda(Y_s,\theta_0)\,\exp\Big(-\int_0^s \Lambda(Y_r,\theta_0)\,\mathrm{d}r\Big)\Big]\,\mathrm{d}s\,.\end{aligned}$$

Therefore we obtain

$$\begin{aligned}
&\mathbb{E}\big[\|M(\tau_1\wedge t)\|_E^2\,|\,\mathcal{F}_0\big]\\
&= \int_0^t \Big[2\big(\tfrac{\mathrm{d}}{\mathrm{d}t}\widetilde{N}(s),\widetilde{N}(s)\big)_E \exp\Big(-\int_0^s \Lambda(Y_r,\theta_0)\,\mathrm{d}r\Big)\Big]\,\mathrm{d}s\\
&\quad + \int_0^t \Big[\Big(\int_K \|z(\vartheta)-z(\theta_0)-\widetilde{N}(s)\|_E^2 - \|\widetilde{N}(s)\|_E^2\,\mu\big((Y_s,\theta_0),\mathrm{d}\vartheta\big)\Big)\\
&\qquad\qquad \Lambda(Y_s,\theta_0)\,\exp\Big(-\int_0^s \Lambda(Y_r,\theta_0)\,\mathrm{d}r\Big)\Big]\,\mathrm{d}s\,.
\end{aligned}$$

Note that $\|z(\vartheta)-z(\theta_0)-\widetilde{N}(s)\|_E^2 = \|z(\vartheta)-z(\theta_0)\|_E^2+\|\widetilde{N}(s)\|_E^2-2(z(\vartheta)-z(\theta_0),\widetilde{N}(s))_E$ and thus

$$\begin{aligned}
&\mathbb{E}\big[\|M(\tau_1\wedge t)\|_E^2\,|\,\mathcal{F}_0\big]\\
&= \int_0^t \Big[2\big(\tfrac{\mathrm{d}}{\mathrm{d}t}\widetilde{N}(s),\widetilde{N}(s)\big)_E \exp\Big(-\int_0^s \Lambda(Y_r,\theta_0)\,\mathrm{d}r\Big)\Big]\,\mathrm{d}s\\
&\quad - 2\int_0^t \Big[\Big(\int_K \big(z(\vartheta)-z(\theta_0),\widetilde{N}(s)\big)_E\,\mu\big((Y_s,\theta_0),\mathrm{d}\vartheta\big)\Big)\\
&\qquad\qquad \Lambda(Y_s,\theta_0)\,\exp\Big(-\int_0^s \Lambda(Y_r,\theta_0)\,\mathrm{d}r\Big)\Big]\,\mathrm{d}s\\
&\quad + \int_0^t \Big[\Big(\int_K \|z(\vartheta)-z(\theta_0)\|_E^2\,\mu\big((Y_s,\theta_0),\mathrm{d}\vartheta\big)\Big)\\
&\qquad\qquad \Lambda(Y_s,\theta_0)\,\exp\Big(-\int_0^s \Lambda(Y_r,\theta_0)\,\mathrm{d}r\Big)\Big]\,\mathrm{d}s\,.
\end{aligned}$$

Due to the form of the derivative (4.6.4) the first two terms cancel and we are left with the equality

$$\begin{aligned}
\mathbb{E}\big[\|M(\tau_1\wedge t)\|_E^2\,|\,\mathcal{F}_0\big] &= \int_0^t \Lambda(Y_s,\theta_0)\int_K \|z(\vartheta)-z(\theta_0)\|_E^2\,\mu\big((Y_s,\theta_0),\mathrm{d}\vartheta\big) \qquad (4.6.5)\\
&\qquad \exp\Big(-\int_0^s \Lambda(Y_r,\theta_0)\,\mathrm{d}r\Big)\mathrm{d}s\,.
\end{aligned}$$

Next we calculate the expectation of the real-valued process

$$\widetilde{N}_2(s) := \int_0^s \Lambda(Y_r,\theta_0)\int_K \|z(\vartheta)-z(\theta_0)\|_E^2\,\mu\big((Y_r,\theta_0),\mathrm{d}\vartheta\big)\,\mathrm{d}r$$

stopped at τ_1. The process $\widetilde{N}_2$ is connected to the process $\widetilde{N}$ defined at the beginning of part (b.1) inasmuch as the integrand of the former is the squared norm of the latter. Furthermore note that $\widetilde{N}_2$ is the term inside the expectation in the right hand side of the Itô-isometry (4.1.4). Thus the aim is now to show that the conditional expectation of $\widetilde{N}_2(t\wedge\tau_1)$ equals the conditional expectation of $\|M(t\wedge\tau_1)\|_E^2$. Again due to the particular structure of the PDMP we obtain for the conditional expectation

$$\begin{aligned}\mathbb{E}\big[\widetilde{N}_2(\tau_1\wedge t)\,|\,\mathcal{F}_0\big] \;&=\; \widetilde{N}_2(s)\,\exp\Big(-\int_0^t \Lambda(Y_r,\theta_0)\,\mathrm{d}r\Big)\\ &\quad + \int_0^t \Big[\widetilde{N}_2(s)\,\Lambda(Y_s,\theta_0)\,\exp\Big(-\int_0^s \Lambda(Y_r,\theta_0)\,\mathrm{d}r\Big)\Big]\,\mathrm{d}s\,.\end{aligned}$$

Integration by parts applied to the integral term yields

$$\begin{aligned}\int_0^t \Big[\widetilde{N}_2(s)\,\Lambda(Y_s,\theta_0)\,\exp\Big(-\int_0^s \Lambda(Y_r,\theta_0)\,\mathrm{d}r\Big)\Big]\,\mathrm{d}s \;&=\; -\widetilde{N}_2(t)\,\exp\Big(-\int_0^t \Lambda(Y_r,\theta_0)\,\mathrm{d}r\Big)\\ + \int_0^t \Big[\Lambda(Y_s,\theta_0)\int_K \|z(\vartheta)-z(\theta_0)\|_E^2\,\mu\big((Y_s,\theta_0),\mathrm{d}\vartheta\big)&\,\exp\Big(-\int_0^s \Lambda(Y_r,\theta_0)\,\mathrm{d}r\Big)\Big]\,\mathrm{d}s\,.\end{aligned}$$

Therefore we obtain that

$$\begin{aligned}\mathbb{E}\big[\widetilde{N}_2(\tau_1\wedge t)\,|\,\mathcal{F}_0\big] \;&=\; \int_0^t \Lambda(Y_s,\theta_0)\int_K \|z(\vartheta)-z(\theta_0)\|_E^2\,\mu\big((Y_s,\theta_0),\mathrm{d}\vartheta\big)\\ &\qquad \exp\Big(-\int_0^s \Lambda(Y_r,\theta_0)\,\mathrm{d}r\Big)\,\mathrm{d}s\,.\end{aligned} \tag{4.6.6}$$

A comparison of the right hand sides in equalities (4.6.5) and (4.6.6) shows that they are equal and thus we obtain after taking the expectation of both conditional expectation that

$$\mathbb{E}\|M(\tau_1\wedge t)\|_E^2 \;=\; \mathbb{E}\widetilde{N}_2(\tau_1\wedge t)\,. \tag{4.6.7}$$

As $\widetilde{N}_2$ is increasing and thus $\widetilde{N}_2(\tau_1\wedge t)\leq \widetilde{N}_2(t)$ almost surely, we obtain that the right hand side in this equation is finite due to condition (4.1.3). Note that (4.6.7) is the Itô-Isometry (4.1.4) for the stopped process $M(t\wedge\tau_1)$.

(b.2) In this part of the proof we show the square-integrability for the process M stopped at an arbitrary jump time τ_k, $k\in\mathbb{N}$, and finally for the non-stopped process M. To this end we first note that Analogously to part (b.1) we find that

$$\mathbb{E}\Big[\big\|M(\tau_{k+1}\wedge t)-M(\tau_k\wedge t)\big\|_E^2\,\big|\,\mathcal{F}_{\tau_k}\Big] \;=\; \mathbb{E}\Big[\widetilde{N}_2(\tau_{k+1}\wedge t)-\widetilde{N}_2(\tau_k\wedge t)\,\Big|\,\mathcal{F}_{\tau_k}\Big]\,.$$

Thus taking the expectation on both sides of this equality yields

$$\mathbb{E}\,\big\|M(\tau_{k+1}\wedge t)-M(\tau_k\wedge t)\big\|_E^2 \;=\; \mathbb{E}\,\widetilde{N}_2(\tau_{k+1}\wedge t)-\mathbb{E}\,\widetilde{N}_2(\tau_k\wedge t)\;<\infty\,, \qquad (4.6.8)$$

where the right hand side is finite as due to(4.1.3) both expectations are finite.
By induction we next show that each $M(\tau_k\wedge t)$ is square-integrable. Assume that $\mathbb{E}\|M(\tau_k\wedge t)\|_E^2<\infty$, where the induction basis for $k=1$ holds due to part (b.1). Then the reverse triangle inequality yields that

$$\mathbb{E}\|M(\tau_{k+1}\wedge t)\|_E^2+\mathbb{E}\|M(\tau_k\wedge t)\|_E^2-2\mathbb{E}\big(\|M(\tau_{k+1}\wedge t)\|_E\,\|M(\tau_k\wedge t)\|_E\big)$$
$$\leq\;\mathbb{E}\|M(\tau_{k+1}\wedge t)-M(\tau_k\wedge t)\|_E^2\,.$$

Here the right hand side is finite due to (4.6.8) and an application of Young's inequality to the product in the left hand side yields that for all $\epsilon>0$

$$(1-2\epsilon)\mathbb{E}\|M(\tau_{k+1}\wedge t)\|_E^2+\big(1-\tfrac{1}{2\epsilon}\big)\mathbb{E}\|M(\tau_k\wedge t)\|_E^2\;<\;\infty\,.$$

Assume that $\mathbb{E}\|M(\tau_{k+1}\wedge t)\|_E^2=\infty$. Then choosing $\epsilon<1/2$ we obtain a contradiction due to the induction hypotheses.

In a final step of this part of the proof we show square-integrability for the non-stopped process. Using Fatou's Lemma and monotone convergence for interchanging limits and expectation we obtain the following upper estimate

$$\begin{aligned}\mathbb{E}\|M(t)\|_E^2 \;&=\; \mathbb{E}\liminf_{k\to\infty}\|M(\tau_k\wedge t)\|_E^2 \qquad (4.6.9)\\ &\leq\; \liminf_{k\to\infty}\mathbb{E}\|M(\tau_k\wedge t)\|_E^2 \;=\; \lim_{k\to\infty}\mathbb{E}\widetilde{N}_2(\tau_k\wedge t)\;=\;\mathbb{E}\widetilde{N}_2(t)\end{aligned}$$

where the final term is finite due to condition (4.1.3). Moreover, as square-integrability implies integrability, the martingale property for the Hilbert space valued process M now follows due to part (a).

(c) Finally, in the last part of the proof we establish the Itô-Isometry. To this end we first show that equality (4.6.7) holds for all $\tau_k\wedge t$, $k\in\mathbb{N}$. Again we proceed by induction with the induction basis given by (4.6.7). We observe that

$$\begin{aligned}\|M(\tau_{k+1}\wedge t)-M(\tau_k\wedge t)\|_E^2 \;&=\; \|M(\tau_{k+1}\wedge t)\|_E^2-\|M(\tau_k\wedge t)\|_E^2 \qquad (4.6.10)\\ &\quad-2\big(M(\tau_k\wedge t),M(\tau_{k+1}\wedge t)-M(\tau_k\wedge t)\big)_E\,.\end{aligned}$$

Taking the conditional expectation with respect to the stopped σ-field $\mathcal{F}_{\tau_k\wedge t}$ we find that the second term in the right hand side of (4.6.10) vanishes as it holds

$$\mathbb{E}\big[\big(M(\tau_k\wedge t), M(\tau_{k+1}\wedge t)-M(\tau_k\wedge t)\big)_E\,\big|\,\mathcal{F}_{\tau_k\wedge t}\big] =$$

$$= \big(M(\tau_k\wedge t), \mathbb{E}\big[M(\tau_{k+1}\wedge t)-M(\tau_k\wedge t)\big)\,\big|\,\mathcal{F}_{\tau_k\wedge t}\big]\big)_E = 0$$

due to the following property of the conditional expectation: Firstly, for E–valued random variables X, Y such that $\mathbb{E}\|X\|_E\|Y\|_E<\infty$ it holds for $\mathcal{G}$–measurable X that $\mathbb{E}\big[(X,Y)_E|\mathcal{G}\big] = \big(X, \mathbb{E}[Y|\mathcal{G}]\big)_E$ [112, Lemma 2.1.2]. Secondly, the Optional Sampling Theorem, i.e., $\mathbb{E}\big[M(\tau_{k+1}\wedge t)\,\big|\,\mathcal{F}_{\tau_k\wedge t}\big] = M(\tau_k\wedge t)\big)$ in the above application, also holds for Hilbert space-valued martingales[12]. Thus we obtain

$$\mathbb{E}\big[\|M(\tau_{k+1}\wedge t)\|_E^2\big|\mathcal{F}_{\tau_k\wedge t}\big]-\mathbb{E}\big[\|M(\tau_k\wedge t)\|_E^2\big|\mathcal{F}_{\tau_k\wedge t}\big]$$

$$= \mathbb{E}\big[\widetilde{N}_2(\tau_{k+1}\wedge t)\big|\mathcal{F}_{\tau_k\wedge t}\big]-\mathbb{E}\big[\widetilde{N}_2(\tau_k\wedge t)\big|\mathcal{F}_{\tau_k\wedge t}\big]\,.$$

Taking the expectation on both sides of this equality and using the induction hypotheses, i.e., the second expectations on both sides of the above equality equate, yields

$$\mathbb{E}\|M(\tau_{k+1}\wedge t)\|_E^2 = \mathbb{E}\widetilde{N}_2(\tau_{k+1}\wedge t)\,. \tag{4.6.11}$$

We conclude the proof extending the Itô-isometry (4.6.11) from the stopped processes to the non-stopped process. We have already obtained the upper estimate $\mathbb{E}\|M(t)\|_E^2\leq\mathbb{E}\widetilde{N}_2(t)$, cf. (4.6.9). Hence it remains to prove that a lower bound is given by the same term. As $\|M(t)\|_E^2$ is a real-valued submartingale it holds for all $k\geq 1$ due to the standard Optional Sampling Theorem for càdlàg submartingales that

$$\mathbb{E}\|M(t)\|_E^2 \geq \mathbb{E}\|M(\tau_k\wedge t)\|_E^2 = \mathbb{E}\widetilde{N}_2(\tau_k\wedge t)\,.$$

Hence, for $k\to\infty$ we obtain by monotone convergence $\mathbb{E}\|M(t)\|_E^2\geq\mathbb{E}\widetilde{N}_2(t)$ which, combined with the upper bound (4.6.9), yields the Itô-isometry (4.1.4). The proof is completed.

4.6.2 Proof of Proposition 4.5.1

We show that the operator $G(u,p)$ defined by (4.5.10) defines a centred Gaussian random variable via being its covariance and that the map $t\mapsto G(u(t),p(t))$ defines a diffusion process. To this end we have to show that $G(u,p)$ is self-adjoint, of trace class and positive, cf. Section 4.3.1. To this end, first note that the mappings

[12]The Optional Sampling Theorem can be proved similarly to the methods employed for [112, Lemma 2.1.2] relying on the linearity properties of the Bochner integral and the monotone convergence theorem.

$G_{ij}(u,p) : L^2(D) \to L^2(D)$ defined by

$$\begin{aligned} G_{ij}(u,p)\,\psi &= \mathbb{I}_{[i=j]}\Big(\sum_{k\neq i} p_k\cdot q_{ki}(u)\cdot\psi + \sum_{k\neq i} p_i\cdot q_{ik}(u)\cdot\psi\Big) \\ &\quad - \mathbb{I}_{[i\neq j]}\Big(p_j\cdot q_{ji}(u)\cdot\psi + p_i\cdot q_{ij}(u)\cdot\psi\Big) \qquad (4.6.12) \end{aligned}$$

are linear and bounded. Further, the mapping $(\Psi,\Phi)\mapsto \big(G(u,p)\,\Psi,\Phi\big)_{L^2}$ is a summation of the operators (4.6.12) acting on the components of Ψ and Φ, i.e.,

$$\big(G(u,p)\,\Psi,\Phi\big)_{L^2} = \sum_{i,j=1}^{m}\big(G_{ij}(u,p)\,\psi_i,\phi_j\big)_{L_2}.$$

Obviously, the operators G_{ij} are symmetric, i.e., $(G_{ij}(u,p)\,\psi,\phi)_{L_2} = (G_{ij}(u,p)\,\phi,\psi)_{L_2}$, and moreover the symmetry $G_{ij}(u,p) = G_{ji}(u,p)$ holds. Hence, it follows that $G(u,p)$ is a symmetric operator on $L^2(D)^m$ and thus self-adjoint due to the Hellinger-Toeplitz Theorem. Clearly, the same holds for G considered as an operator mapping $H^s(D)$ into $H^{-s}(D)$ due to the continuous embeddings $H^s(D)\hookrightarrow L^2(D)\hookrightarrow H^{-s}(D)$.

We proceed to show the trace class property of $G(u,p)$ as an operator mapping $H^s(D)$ into $H^{-s}(D)$ where $s > d/2$. As usual the identification of an element of $H^s(D)$ with a linear functional is via the inner product in $L^2(D)$, i.e., $\langle v,u\rangle_{H^s} = (v,u)_{L^2}$ for all $u,v\in H^s(D)$. Let $(\varphi_k)_{k\in\mathbb{N}}$ denote an orthonormal basis in $H^s(D)$. Then an application of the Cauchy-Schwarz inequality yields for all $i,j=1,\dots m$, that

$$\sum_{k=1}^{\infty}|\langle G_{ij}(u,p)\varphi_k,\varphi_k\rangle_{H^s}| \leq C\sum_{k=1}^{\infty}\|\varphi_k\|_{L^2}^2$$

for a suitable constant C depending only on a pointwise bound of the rates q_{ij}. This upper bound is finite as the embedding $H^s(D)\hookrightarrow L^2(D)$ is of Hilbert-Schmidt type due to Maurin's Theorem. Hence the operators G_{ij} are of trace class. Next we extend the trace class property from its summands to $G(u,p)$. To this end let $\overline{\varphi}_k = (\varphi_k^1,\dots,\varphi_k^m)$ denote the elements of an orthonormal basis of $H^s(D)^m$ where each component is an orthonormal basis in the component space $H^s(D)$. Then we obtain

$$\begin{aligned} \langle G(u,p)\overline{\varphi}_k,\overline{\varphi}_k\rangle_{(H^s)^m} &= \sum_{k=1}^{\infty}\sum_{i,j=1}^{m}|\langle G_{ij}(u,p)\varphi_k^i,\varphi_k^j\rangle_{H^s}| \\ &= \sum_{i=1}^{m}\sum_{k=1}^{\infty}|\langle G_{ii}(u,p)\varphi_k^i,\varphi_k^i\rangle_{H^s}| + \sum_{i\neq j}^{m}\sum_{k=1}^{\infty}|\langle G_{ij}(u,p)\varphi_k^i,\varphi_k^j\rangle_{H^s}|, \end{aligned}$$

where the finiteness of the first sum follows immediately from G_{ii} being trace class operators. For the second sum we observe that the summands satisfy due to the Cauchy-Schwarz and Young's inequality

$$|\langle G_{ij}(u,p)\varphi_k^i, \varphi_k^j\rangle_{H^s}| \leq C\,\|\varphi_k^i\|_{L^2}\|\varphi_k^j\|_{L^2} \leq \frac{C}{2}\|\varphi_k^i\|_{L^2}^2 + \frac{C}{2}\|\varphi_k^j\|_{L^2}^2$$

for a suitable constant C. Hence, a summation over these terms is again finite due to the Hilbert-Schmidt embedding. Therefore, we infer that G is a trace class operator.

Finally, it only remains to show that the operator G is positive, that is, it satisfies $\langle G(u,p)\Phi, \Phi\rangle_{(H^s)^m} \geq 0$ for all $\Phi \in H^s(D)^m$. This is easily seen by an appropriate reorganisation of summands:

$$\begin{aligned}
\big(G(u,p)\,\Phi, \Phi\big)_{L^2} &= \sum_{i,j=1}^m \big(G_{ij}(u,p)\,\phi_i, \phi_j\big)_{L_2} \\
&= \sum_{i=1}^m \sum_{\substack{j=1\\ j\neq i}}^m \big(\phi_i^2, p_i q_{ij} + p_j q_{ji}\big)_{L^2} - \sum_{i=1}^m \sum_{j>i} \big(2\phi_i\phi_j, p_i q_{ij} + p_j q_{ji}\big)_{L^2} \\
&= \sum_{i=1}^m \sum_{j>i}^m \big(\phi_i^2 + \phi_j^2, p_i q_{ij} + p_j q_{ji}\big)_{L^2} - \sum_{i=1}^m \sum_{j>i} \big(2\phi_i\phi_j, p_i q_{ij} + p_j q_{ji}\big)_{L^2} \\
&= \sum_{i=1}^m \sum_{j>i}^m \big((\phi_i - \phi_j)^2, p_i q_{ij} + p_j q_{ji}\big)_{L^2} \\
&\geq 0\,.
\end{aligned}$$

The proof of Proposition 4.5.1 is completed.

Chapter 5

Numerical Methods for PDMPs

Due to the high complexity of hybrid models, particularly in biochemical applications, the former are studied extensively and almost exclusively by numerical means. To this end researchers either employ statistically exact algorithms or use statistically approximate algorithms, also called 'pseudo-exact' algorithms. Here *statistically exact* means that the simulation method produces paths that are samples of the distribution of the underlying stochastic process in contrast to pseudo-exact methods which produce sample paths from a distribution approximating the distribution of the underlying stochastic process. However, even statistically exact algorithms are *exact only in theory due to numerical errors which arise inevitably* in solving most systems of differential equations in actual implementations. As the main source of error researchers are so far primarily interested in model accuracy and thus considerations regarding the numerical error are in general neglected, cf., e.g., [3, 58, 70]. Yet ultimately, even if a theoretically exact PDMP formulation of the model is considered or the model is highly accurate, *numerical studies are conducted by numerical approximations* to the PDMPs as in general an analytic representation of the paths is not available. Despite the importance and widespread use of numerical studies an analysis of the numerical error, in particular, a thorough analytical investigation of the convergence and error behaviour of any algorithms used, is still missing. The aim of this chapter is to provide a – to the best of our knowledge – first contribution towards this goal.

We note that the class of numerical methods we consider in the following contains these methods which have been presented in the literature when addressing exact numerical simulation algorithms, cf. [3, 123]. However, in the present study a convergence analysis is carried out which has not been attempted by the aforementioned authors. In particular, we are interested in the convergence of numerical approximations to PDMPs in a pathwise sense which corresponds to the fact that numerical simulations are carried out path by path. Thus, let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ be a probability space and

$(X(t,\omega))_{t\in[0,T]}$ and $(\widehat{X}(t,\omega))_{t\in[0,T]}$ denote a PDMP[1] and its numerical approximation defined thereon. We then consider the errors

$$|X(T,\omega) - \widehat{X}(T,\omega)| \quad \text{and} \quad \max_{n=1,\dots,N(T,\omega)} |X(\tau_n(\omega),\omega) - \widehat{X}(\widehat{\tau}_n(\omega),\omega)|$$

for almost all $\omega \in \Omega$, where $\tau_n(\omega)$, $\widehat{\tau}_n(\omega)$ are the random jump times of the exact and approximate process and $N(T,\omega)$ the number of jumps of the PDMP in $[0,T]$. As, in general, $\tau_n(\omega)$ and $\widehat{\tau}_n(\omega)$ do not coincide, we also consider the error in the jump times, i.e.,

$$\max_{n=1,\dots,N(T,\omega)} |\tau_n(\omega) - \widehat{\tau}_n(\omega)| \,.$$

We postpone a discussion of the reason to consider these error measures to Section 5.2.

The methods we present for approximating a PDMP incorporate *continuous ODE methods*, also called methods with dense output. Apart from numerically solving the deterministic inter-jump dynamics the key problem in simulating a PDMP is simulating the random, path-dependent jump times. As shown in Section 5.2 this problem can be reformulated and then, combined with the numerical solution of the inter-jump dynamics, yields a *hitting time problem with random threshold*. This we solve using continuous ODE methods. The main feature of such a method is that it does not only provide a numerical approximation to the exact solution at discrete grid points but provides an approximation of the whole path over the whole interval. That is, continuous methods are essentially approximations on a discrete grid with an interpolation formula for the intervals between the grid points. Hence, these methods are naturally suited for solving hitting time problems. The main result of this chapter is that numerical approximations of PDMPs built on continuous ODE methods conserve the order of convergence of the underlying continuous method. That is, if an approximation is constructed using, e.g., a continuous Runge-Kutta method of order p, then also the almost sure convergence of the stochastic approximation to the path of the PDMP is of order p.

The remainder of this chapter is organised as follows. In Section 5.1 we present a theoretically exact simulation algorithm based on independent, identically distributed (i.i.d.) standard uniform random variables which takes the role of an exact solution for our convergence analysis. Section 5.2 presents the approximate algorithms we consider in this chapter and also contains the main convergence theorem. The proof of the

[1]Throughout this thesis we have employed the notational convention that the time variable for a stochastic process is given as a lower index whereas for deterministic functions as an argument in brackets. However, throughout this chapter we denote the time argument in brackets also for stochastic processes. As in the following we have distinguish frequently between time arguments τ_n, $\widetilde{\tau}_n$ and $\widehat{\tau}_n$ we are of the opinion that the present notation is an improvement in legibility and intelligibility.

convergence theorem can be found in Section 5.3. We extend the convergence theorem to a larger class of PDMPs in Section 5.4. Finally, in Section 5.5 we present some numerical experiments using examples from the neuroscience literature to illustrate the theoretical findings and draw some conclusions for the usage and implementations of such algorithms.

5.1 Definition of an exact 'solution'

In this section we present Algorithm A1 based on Davis construction procedure, cf. Definition 2.1.2, which provides the 'exact solution' numerical approximations have to converge to. Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ be a filtered probability space and let $(X(t))_{t\in[0,T]} = (X(t,\omega))_{t\in[0,T]}$ be a standard finite-dimensional PDMP with state space E defined thereon, see Definition 3.0.2. That is, $E \subseteq \mathbb{R}^d \times K$ where K is an at most countable and bounded subset of $\mathbb{R}^m$ and thus $X(t) = (Y(t), \theta(t)) \in \mathbb{R}^{d+m}$. Here $Y(t)$ possesses continuous paths in $\mathbb{R}^d$ and $\theta(t)$ has piecewise constant, right-continuous paths in K. For the sake of simplicity we restrict ourselves to PDMPs with deterministic initial condition $X(0) \equiv x_0 \in E$. We consider this particular structure of processes, whose jumps only occur in a fixed subset of variables which are otherwise constant on inter-jump intervals, as these correspond to the stochastic processes encountered in applications in mathematical neuroscience, cf. Chapter 3. However, it is straightforward to extend the results in Section 5.2 to a broader class of PDMPs which contain processes encountered in multiscale models of chemical reaction systems, cf. Section 5.4.

Recall that the PDMP is uniquely defined by the characteristics (g, λ, μ), cf. the remark following Definition 3.0.2. That is, the deterministic evolution of the continuous component $Y(t)$ in between jumps is governed by a family of ordinary differential equations

$$\dot{y} = g(y, \theta) \tag{5.1.1}$$

with parameter $\theta \in K$. Hence, as the component $\theta(t)$ of a PDMP is constant in between jumps it follows that in between jumps a PDMP's paths satisfy a system of ODEs of the form

$$\begin{pmatrix} \dot{y} \\ \dot{\theta} \end{pmatrix} = \begin{pmatrix} g(y,\theta) \\ 0 \end{pmatrix}. \tag{5.1.2}$$

We denote by $t \mapsto \phi(t, x_0)$ the unique, global solution of the system (5.1.2) with respect to the initial value $x_0 = (y_0, \theta_0) \in E$. We assume that it satisfies $\phi(t, x_0) \in E$ for all $t \in [0, T]$ and all initial values $x_0 \in E$. Further, the transition rate λ defines a survivor function S as in (3.0.2), i.e.,

$$S(r, X(t)) = \exp\Big(-\int_0^r \lambda\big(\phi(s, X(t))\big)\,\mathrm{d}s\Big)\,. \tag{5.1.3}$$

and μ is a Markov kernel from E into K such that $\mu((x,\theta),\{\theta\}) = 0$ for all $(x,\theta) \in E$. In addition to the Definition 3.0.2 of a standard PDMP we assume that μ is continuous in x, i.e., $x \mapsto \mu(x, A)$ is continuous for every Borel set A of K.

Example. We present a concrete example for a PDMP arising in mathematical neuroscience as a model of a space-clamped membrane, see Section 3.1.2. This model is also considered as a numerical example in Section 5.5. The component $Y(t)$ is one-dimensional taking values in $[0, E_{\mathrm{Na}}]$ for initial conditions therein and the component $\theta(t)$ is 8–dimensional where $K = \{\theta \in \{0,\dots,N\}^8 : \sum_{k=1}^8 \theta_k = N\}$ with $N \in \mathbb{N}$. Thus the phase space is $E = [0, E_{\mathrm{Na}}] \times K$ and the characteristics are given as follows. The family of ODEs (5.1.1) is given by

$$C\dot{y} = -\overline{g}_{\mathrm{Na}}\,\theta_8\,(y - E_{\mathrm{Na}}) - \overline{g}_{\mathrm{L}} y\,,\quad \theta \in K$$

with constants E_{Na}, $\overline{g}_{\mathrm{Na}}$, $\overline{g}_{\mathrm{L}} > 0$. The stochastic dynamics are given by the jump rate

$$\lambda((y,\theta)) = \begin{pmatrix} a_m(y)\\ b_m(y)\\ a_h(y)\\ b_h(y)\end{pmatrix}^T \begin{pmatrix} 3&2&1&0&3&2&1&0\\ 0&1&2&3&0&1&2&3\\ 1&1&1&1&0&0&0&0\\ 0&0&0&0&1&1&1&1\end{pmatrix} \begin{pmatrix}\theta_1\\ \vdots\\ \theta_8\end{pmatrix}.$$

and the transition measure μ which is given by point probabilities of the form

$$\mu((y,\theta)),\{\theta + (-1,1,0,0,0,0,0,0)^T\}) = \frac{3\,a_m(y)\,\theta_1}{\lambda((y,\theta))}\,.$$

Davis' construction procedure, given as Algorithm A1 below, provides a theoretically exact method for simulating paths of a PDMP to given characteristics on a probability space that supports a sequence of i.i.d. standard uniform random variables, cf. Section 2.1.1. Here 'exact' denotes the fact that the distribution of the process defined by Algorithm A1 equals the distribution of a PDMP to the triple (g, λ, μ). However, the adverb 'theoretically' should emphasise the fact that in general neither IVP (5.1.2) nor the implicit equation (5.1.4) arising in Algorithm A1 can be solved exactly, thus this algorithm cannot be employed in practice to simulate trajectories exactly. Yet, it can be used as a comparison to approximate algorithms for a convergence analysis and thus, in typical terms of numerical analysis, plays the role of an exact solution.

Algorithm A1. *An exact simulation algorithm for a PDMP is given by:*

Step 1. *Set a jump counter* $n = 0$ *and fix the initial time* $\tau_0 = 0$ *and initial condition* $X(0) = x_0 = (y_0, \theta_0) \in E$.

Step 2. *Simulate a uniformly distributed random variable* U *and solve*

$$S(\sigma, X(\tau_n)) = U \tag{5.1.4}$$

with respect to σ *to obtain the waiting time until the next jump time, i.e.,* $\sigma = \tau_{n+1} - \tau_n$. *Then for* $t \in (\tau_n, \tau_{n+1}]$ *set*

$$X(t) = \phi(t - \tau_n, X(\tau_n))\,.$$

If $\tau_{n+1} \geq T$, *stop at* $t = T$.

Step 3. *Else, simulate a post-jump value* θ_{n+1} *for the piecewise constant component according to the distribution* $\mu(\phi(\tau_{n+1} - \tau_n, X(\tau_n)), \cdot\,)$ *and set*

$$X(\tau_{n+1}) = \begin{pmatrix} \phi(\tau_{n+1} - \tau_n, X(\tau_n)) \\ \theta_{n+1} \end{pmatrix}.$$

Step 4. *Set* $n = n + 1$ *and start again with Step 2.*

To conclude this section, we briefly comment on prominent classes of jump Markov processes that are special cases of PDMPs and their simulation. A first special case is given by ODEs with Markovian switching. For these processes the jump rate is given by jumps of a Poisson process that is independent of the paths followed in between jumps, i.e., $\lambda \equiv \mathit{const}$. Such equations can be written as stochastic differential equations driven by a Poisson process, i.e., stochastic differential equations of jump type (JSDEs) without a diffusion term, cf. [86]. In general, JSDEs without a diffusion part are special cases of PDMPs [64]. Secondly, PDMPs are generalisations of continuous-time Markov chains. These are piecewise constant processes and thus follow a trivial evolution between successive jump times, i.e., $g \equiv 0$. However, in contrast to switching ODEs the jump intensity may be state dependent. Such processes are most importantly used in stochastic models of chemical reaction kinetics and are treated numerically by the Stochastic Simulation Algorithm (SSA) [53]. We note that the numerical advantage of these special cases is that inter-jump times can be simulated exactly by sampling from an exponential distribution, that is, the implicit equation (5.1.4) can be solved exactly. Accordingly there exists a large amount of literature on simulation methods for these processes and a numerical analysis thereof, see, for instance, references in the recent studies [26, 82].

5.1.1 A note on simulation methods from uniform random variables

By virtue of Algorithm A1 and the properties of the characteristics (g, λ, μ) we can construct a PDMP uniquely from a sequence $(U_n)_{n\geq 1}$ of i.i.d. standard uniform random variables. For each inter-jump interval two random variables have to be calculated from samples of uniform random variables: first, the waiting time until the next jump σ and, secondly, the post jump values θ_n. Serially processing these tasks the random variables U_{2n-1} are used to define the nth inter-jump waiting time via equation (5.1.4) and the random variables U_{2n} are used to simulate the nth post-jump value of the PDMP. We briefly comment on the methods for simulating these two types of random variables as they are an integral part of a simulation method for the path of a PDMP.

Firstly the simulation method (5.1.4) employed in Step 2 stems from the fact that every continuous distribution function F on $\mathbb{R}$ with generalised inverse $F^{-1}(u) = \inf\{t : F(t) \geq u\}$ satisfies for a uniformly distributed random variable U

$$\mathbb{P}[F^{-1}(U) \leq t] = F(t)\,,$$

cf. the *inverse CDF method*, e.g., [51]. Therefore, as

$$\inf\{t : F(t) \geq u\} = \inf\{t : S(t) \leq 1 - u\}$$

for $S(t) = 1 - F(t)$ being the survivor function to F, it follows that in the case of a continuous survivor function the solution σ of (5.1.4) is distributed according to the survivor function (5.1.3). Note that $1 - U$ is also uniformly distributed and thus without loss of generality we may use U instead of $1 - U$ which yields (5.1.4). Further, we note that equation (5.1.4) does not necessarily possess a unique solution or a solution at all. However, both these cases are excluded by the conditions on λ in the main convergence theorem and hence for the remainder of the paper we can assume that a unique solution to (5.1.4) always exists.
Secondly, for simulating the post-jump values we exploit the fact that there exist deterministic functions

$$\Theta_n : E \times (0, 1) \to \mathbb{R}^m : (x, u) \mapsto \Theta_n(x, u) \tag{5.1.5}$$

defining the random nth jump height conditional on the pre-jump value x in terms of realisations u of uniform random variables, cf., e.g., [37, Cor. 23.4]. That is,

$$\theta(\tau_n) - \theta(\tau_n-) = \Theta_n(X(\tau_n-), U_{2n}),$$

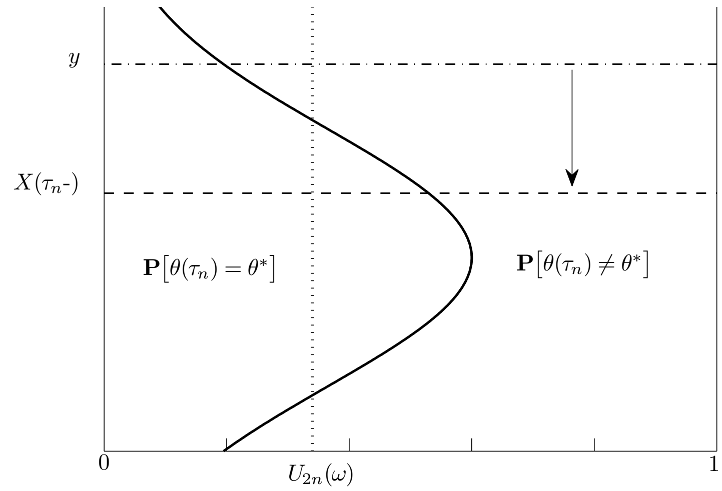


Figure 5.1: Schematic view of the domain of Θ_n. The curve marks points of discontinuity. The function Θ_n is constant θ^* for arguments (x, u) left of the curve and piecewise constant with different values right of it. θ^* is the realised nth jump height and we see $\Theta_n(y, U_{2n}(\omega)) = \theta^*$ for y close enough to $X(\tau_n-)$ as the probability that U_{2n} lies on a discontinuity is zero.

if τ_n denotes the PDMP's nth jump time. In particular we have that for all $n \geq 1$

$$\mathbb{P}[\theta + \Theta_n((y, \theta), U) \in A] = \mu((y, \theta), A)$$

for U standard uniformly distributed and all Borel sets A of K. As K is an at most countable set we can choose $\Theta_n(x, \cdot\,)$ to be piecewise constant on $(0, 1)$ for all $x \in E$ with possibly countably many discontinuities. More explicitly, each $\mu(x, \cdot\,)$ is a discrete probability distribution on K and to construct a function $\Theta_n(x, \cdot\,)$ we partition the interval $(0, 1)$ into bins with lengths the size of the point probabilities $\mu(x, \{\theta - \widehat{\theta}\})$, $\theta, \widehat{\theta} \in K$. Then the realisation of a random jump height with distribution $\mu(x, \cdot\,) - x$ is given by the distance $\theta - \widehat{\theta}$ which corresponds to the bin a realisation $U(\omega)$ falls into. This is the standard method of simulating discrete random variables [51]. We note, that clearly such a function Θ_n is not uniquely defined, as, e.g., every permutation of the partition intervals yields a method producing samples from the same original distribution $\mu(x, \cdot\,)$. Thus different choices of functions Θ_n produce different paths for the same sequence $(U_n)_{n \in \mathbb{N}}$. Nevertheless, the stochastic processes defined by these different choices have the same distribution in path space and hence are versions of each other defined on the same probability space. That means, the choice of the functions Θ_n does not matter for simulating paths of a PDMP.

Finally, we derive in (5.1.7) a consequence of the assumption that the transition probabilities $\mu(x, \cdot)$ depend continuously on the pre-jump value x. This result is of importance for the convergence of approximate simulation algorithms.
First note that the continuity of μ with respect to x implies that the point probabilities

$\mu(x, \theta - \widehat{\theta})$ and thus the boundaries of the partition intervals depend continuously on x. Hence, for all $n > 0$ we have that $\Theta_n(x, \cdot)$ is a piecewise constant function with discontinuities in u that depend continuously on x, cf. Fig. 5.1. Thus for fixed $(x, u) \in E \times (0, 1)$, where u is not a point discontinuity of $\Theta_n(x, \cdot)$, it follows that

$$|\Theta_n(x, u) - \Theta_n(y, u)| = 0$$

for y close enough to x, say for $|x - y| < \kappa$, with constant $\kappa = \kappa(x, u)$ depending only on the fixed pair (x, u). We denote the set of all discontinuity points of $\Theta_n(x, \cdot)$ by $D_n(x)$ and, as $D_n(x)$ is at most countable for all x in E, it follows that

$$\mathbb{P}\big[U_n \in D_n(X(\tau_n-)) \;\forall\; n \geq 1\big] = 0\,. \tag{5.1.6}$$

Hence, for all $n \geq 1$ it is valid that

$$|\Theta_n(X(\tau_n-), U_{2n}) - \Theta_n(y, U_{2n})| = 0 \quad a.s. \tag{5.1.7}$$

for y sufficiently close to $X(\tau_n-)$. That is, for y sufficiently close to the exact value $X(\tau_n-)$ the simulation method for the post-jump value is a.s. exact.

5.2 The main convergence theorem

In this section we present the numerical methods employed to approximate a PDMP defined via an i.i.d. sequence of standard random variables and state the central result at the end of the section: the almost sure convergence of the numerical approximations and their asymptotic order of convergence. We first precisely state and discuss the convergence concept we are interested in before we continue with constructing the numerical methods by discretising Algorithm A1.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space which is not the same as in Section 5.1. Further, U_n, $n \geq 1$ is a sequence of i.i.d. standard uniform random variables defined on this probability space and we fix a finite time interval $[0, T]$. Then for given characteristics (g, λ, μ) as in Section 5.1 we use the sequence of uniform random variables to construct a PDMP $(X(t, \omega))_{t \in [0,T]}$ on this probability space via Algorithm A1. A suitable filtration $(\mathcal{F}_t)_{t \in [0,T]}$ does exist [37]. That is, we recover the structure as initially assumed in Section 5.1.
Moreover, we assume that a numerical approximation $(\widehat{X}(t, h, \omega))_{t \in [0,T]}$ is defined on the same probability space. Here h denotes a defining parameter of the numerical approximation, such as a discretisation step size. In the following we refer to the PDMP $(X(t, \omega))_{t \in [0,T]}$ as the *exact PDMP* and to its numerical approximation $(\widehat{X}(t, h, \omega))_{t \in [0,T]}$ as the *approximate PDMP*. Let $\tau_n(\omega)$ and $\widehat{\tau}_n(h, \omega)$, $n \geq 1$, denote

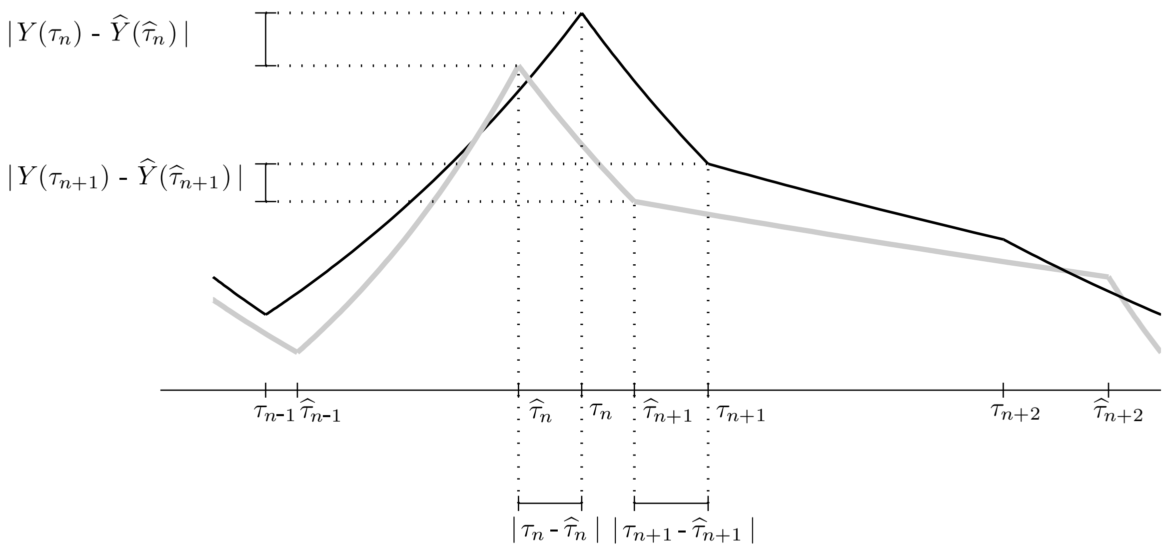


Figure 5.2: Schematic relation of the continuous component $Y(t)$ of the PDMP (black line) to its approximation $\widehat{Y}(t)$ (grey line) in phase space and the approximation errors we need to control to obtain path-wise convergence of Algorithm A2.

the jump times of the exact PDMP and its approximation, respectively. Finally, the number of jumps of the exact PDMP in the time interval $[0,T]$ is denoted by $N(T,\omega)$, where $N(T,\omega) < \infty$ almost surely.
As stated at the beginning of the Chapter, we are interested in the *pathwise convergence of global errors* of the form

$$\lim_{h\to 0} \max_{n=1,\dots,N(T,\omega)} |X(\tau_n(\omega),\omega) - \widehat{X}(\widehat{\tau}_n(h,\omega),h,\omega)| = 0 \tag{5.2.1}$$

and

$$\lim_{h\to 0} |X(T,\omega) - \widehat{X}(T,h,\omega)| = 0 \tag{5.2.2}$$

for almost all $\omega \in \Omega$. That is, (5.2.1) and (5.2.2) state the almost sure convergence of the approximate PDMP to the exact PDMP at their jump times and at the endpoint, cf. Fig. 5.2. It is a standard approach in numerical analysis, whether deterministic or stochastic, to define the error of a numerical approximation by a measure of the difference of the exact and approximate process at discrete time points in the approximation interval $[0,T]$ [57, 72, 92]. Here the jump times of the PDMP and its approximation are used as a discretisation of the time interval. However, in general, the exact and approximate jump times do differ and hence yield different time discretisations for the exact and approximate PDMP. This is in contrast to error definitions typically used in numerics. Moreover, typically, the number and locations of grid points in error definitions are the same for each trajectory and increase in number for decreasing step size h. It is exactly the opposite in the error definition (5.2.1) as $N(T,\omega)$ varies with $\omega \in \Omega$ and but is fixed over variations in h.

As the numerical approximation is continuous over time we are in principle able to use a 'typical' error definition. However, for ordinary or stochastic differential equations the time points the error is evaluated at arise naturally as the grid points of the discretisation used in the numerical method. For PDMP approximations as considered in this study this is not the case. The methods we consider, in general, do not possess the same grid points for any trajectory of one PDMP to approximate and they possibly do not even possess the same number of grid points. Therefore, setting particular but arbitrary points apart in using them in an error definition without a mathematical reasoning does not seem preferable to us. However, for a PDMP all numerical methods necessarily need to approximate the jump times. Therefore, relating the values of the exact and approximate PDMP at their respective jump times is reasonable insofar as an accurate approximation is expected to behave after the nth just as the exact PDMP behaves after the nth jump. We note that this is related to the idea of discontinuity tracking in connection with numerical solution of delay differential equations [11].
We have already mentioned that, in general, the jump times of the exact and approximate PDMP differ. Therefore it is not sufficient to consider the errors (5.2.1) and (5.2.2) alone. They may become arbitrarily small with the exact and approximate trajectories still substantially differing as possibly only the values at the jump times converge but the jump times themselves remain separated. Thus, we additionally require from a convergent method that also the jump times of the approximate PDMP converge to the jump times of the exact PDMP, that is, for almost all $\omega \in \Omega$

$$\lim_{h\to 0} \max_{n=1,\dots,N(T,\omega)} |\tau_n(\omega) - \widehat{\tau}_n(h,\omega)| = 0\,. \tag{5.2.3}$$

Definition 5.2.1. We say an approximation is pathwise convergent if (5.2.1)–(5.2.3) hold. In addition, an approximation is of order p if p is the largest integer such that the asymptotic behaviour of (5.2.1)–(5.2.3) is $\mathcal{O}(h^p)$.

To construct an approximation algorithm for the numerical simulation of a PDMP's paths we start from the theoretically exact Algorithm A1. We obtain an approximate algorithm by discretising the problems (5.1.2) and (5.1.4) and solving them numerically. A numerical solution of (5.1.2) can be obtained by a standard ODE method, thus substituting Step 3 with an approximation is straightforward. However, the delicate part is numerically solving (5.1.4) for which obviously a numerical solution of (5.1.2) is needed as an integrand which is integrated until an *a priori unspecified time* σ, cf. the definition of the survivor function (5.1.3). Hence, the two problems have to be solved in parallel including a mechanism for detecting the time σ conditional on a realisation of U. For an efficient solution we transform the equation for the next transition time (5.1.4) into an equivalent problem. This can then be combined with the IVP (5.1.2) yielding a merged formulation of the two problems which allows for a numerical solution by continuous ODEs methods. The resulting numerical approximation is finally given by Algorithm A2 below.

Taking the logarithm of equation (5.1.4) we obtain the equivalent equation

$$w(\sigma, x) = \int_0^\sigma \lambda(\phi(s,x))\mathrm{d}s = -\log U\,. \tag{5.2.4}$$

Thus $w(\sigma, x) = -\log S(\sigma, x)$ for the survivor function S. Differentiation of w with respect to σ yields

$$\dot{w}(t,x) = \lambda(\phi(t,x))\,.$$

This in turn yields that calculating the next jump time from (5.1.4), i.e., Step 2 in Algorithm A1, is equivalent to solving the IVP

$$\begin{pmatrix} \dot{y} \\ \dot{\theta} \\ \dot{w} \end{pmatrix} = \begin{pmatrix} g(y,\theta) \\ 0 \\ \lambda(y,\theta) \end{pmatrix}, \quad \begin{pmatrix} y(0) \\ \theta(0) \\ w(0) \end{pmatrix} = \begin{pmatrix} x_0 \\ 0 \end{pmatrix}, \; t \in [0,\infty)\,, \tag{5.2.5}$$

which is integrated until the component $w(t)$ hits the threshold $-\log U$. The hitting time

$$\sigma = \inf\{t > 0 : w(t) = -\log U\}$$

is equivalent to the waiting time until the next jump, where, in general, $\sigma = \infty$ may be possible unless λ is bounded away from zero.

We need to remark on two aspects at this point. Firstly, the setup of the hitting time problem (5.2.5) is as in [3]. The authors therein also state how it can be solved using continuous methods and also present an ad-hoc implementation of an event detection for this specific problem. However, as the authors are primarily focused on the modelling and simulation results, they suppose that (5.2.5) can be computed "up to any desired accuracy and neglect the discretisation error" [3, p. 6]. This may be a reasonable assumption for the solution of any standard ODE without jumps in the parameters as numerical methods are well studied in this case. However, *additional assumptions are needed and have to be considered* such that an event detection is possible with arbitrary accuracy, cf. [108]. This aspect is not discussed by the authors. An algorithm for simulating a PDMP's path repeatedly solves an ODE system starting at the point the previous event detection had returned. Thus we can only assume that the algorithm produces a path with any desired accuracy if also the event detection locates the hitting times with any desired accuracy. To repeat the purpose of the present analysis, it is precisely this point we address and present for a large class of algorithms a thorough numerical analysis in terms of pathwise convergence and the conditions this presupposes.

Secondly, we note that there are different but equivalent setups for the event detection problem (5.2.5), i.e., also different to [3]; one as stated in [123] and a second, which is essentially analogous to the setup in [123], we briefly introduce now. Instead of manipulating equation (5.1.4) to obtain (5.2.4) we differentiate the survivor function S with respect to σ and obtain

$$\dot{S}(\sigma, x) = -S(\sigma, x)\, \lambda(\phi(\sigma, x))\,.$$

This yields instead of (5.2.5) an IVP to solve given by

$$\begin{pmatrix} \dot{y} \\ \dot{\theta} \\ \dot{S} \end{pmatrix} = \begin{pmatrix} g(y,\theta) \\ 0 \\ -S\,\lambda(y,\theta) \end{pmatrix}, \quad \begin{pmatrix} y(0) \\ \theta(0) \\ S(0) \end{pmatrix} = \begin{pmatrix} x_0 \\ 0 \end{pmatrix}, \ t \in [0,\infty)\,. \tag{5.2.6}$$

This system is solved until the component $S(t)$ hits the threshold given by the random variable U. Once again the hitting time is exactly the time until the next transition. Clearly the two systems (5.2.5) and (5.2.6) are equivalent in the sense that the y and θ components of their respective solutions coincide as also do the hitting times of the respective thresholds for a given realisation of U.

However, with respect to a theoretical analysis the setup (5.2.5) is the most feasible due to the simpler right hand side in the last component. Also we conjecture that for actual implementations it is more efficient as (5.2.5) is a simpler type of ODE system than (5.2.6). For this reason, we consider in what follows IVP (5.2.5) and denote its solution with respect to the initial value $(x_0, 0)$ by $\psi(t, x_0)$. Concerning the different notations introduced so far we summarise

$$\psi(t,x_0) = \begin{pmatrix} \phi(t,x_0) \\ w(t,x_0) \end{pmatrix} = \begin{pmatrix} y(t,x_0) \\ \theta_0 \\ w(t,x_0) \end{pmatrix}, \quad \begin{pmatrix} x_0 \\ 0 \end{pmatrix} = \begin{pmatrix} y_0 \\ \theta_0 \\ 0 \end{pmatrix}$$

as we make use all of these notations whenever brevity or clarity demands.

By $\widehat{\psi}(t,x,h), t \geq 0$, we denote a continuous approximate solution to the IVP (5.2.5) with initial condition $(x, 0)$ obtained by a continuous numerical ODE method with step size h. An example of a continuous method is the *continuous trapezoidal rule*, which applied to (5.1.1) takes the form

$$\widehat{y}((n+1)h) \;=\; \widehat{y}(nh) + h\,\tfrac{1}{2}\,g\big(\widehat{y}(nh)\big) + h\,\tfrac{1}{2}\,g\big(\widehat{y}((n+1)h)\big)\,, \tag{5.2.7}$$

with $n = 0, \dots, N-1$, $h = T/N$ and for $\xi \in [0,1]$

$$\widehat{y}(nh+\xi h) \;=\; \widehat{y}(nh) + h\,\tfrac{1}{2}\xi(2-\xi)\,g\big(\widehat{y}(nh)\big) + h\,\tfrac{1}{2}\xi^2\,g\big(\widehat{y}((n+1)h)\big)\,. \qquad (5.2.8)$$

That is, (5.2.7) is the usual trapezoidal rule for ODEs supplemented with an interpolation formula (5.2.8) to obtain an approximation over the intervals $[nh,(n+1)h]$ between the discrete grid points. We remark that although for the trapezoidal rule or the numerical methods implemented for the examples in Section 5.5 the parameter h denotes an equidistant step size this need not necessarily be the case. On the one hand, h may denote the maximal step size used in case of variable step size methods, or, on the other hand, a given error tolerance, cf. a discussion of deterministic event detection in [108]. Essentially, h is a defining parameter of the method which, if convergent, converges to the exact solution for $h \to 0$. In favour of linguistic simplicity we keep referring to h as the step size in the remainder of the chapter. Further, to keep the presentation and notation simple we restrict ourselves to continuous one-step methods, the implementations in Section 5.5 all being based on continuous Runge-Kutta methods discussed in [11]. However we expect the subsequent results, i.e., Theorem 5.2.1 and Corollary 5.4.1, to remain valid in the case of continuous multi-step methods.

We refer to [11, 57] for a general discussion of continuous methods and just briefly collect properties of these methods which we always assume to hold when employed to solve a standard ordinary differential equation of the general form

$$\dot{y} = f(t,y) \qquad (5.2.9)$$

on the interval $[0,T]$. To discuss these properties we employ in this paragraph the following notation. We denote by $y(t,x)$ the exact solution of (5.2.9) with respect to the initial condition $y(0,x) = x$ and $\widehat{y}(t,x,h)$ denotes the numerical approximation with respect to the same initial condition and step-size h.

Firstly, the approximate solution obtained by a continuous ODE method satisfies a stability-type estimate of the form

$$\max_{t\in[0,T]} |y(t,x) - \widehat{y}(t,z,h)| \le \mathrm{e}^{C_1 T}|x-z| + \mathrm{err}(T,h)\,, \qquad (5.2.10)$$

with an error function $\mathrm{err}(T,h)$ that satisfies $\mathrm{err}(T,h) \to 0$ for $h \to 0$. Moreover, the positive constant C_1 and the function err depend only on the right hand side of (5.2.9). When the right hand side of (5.2.9) is Lipschitz continuous (cf. conditions of Theorem 5.2.1) both depend only on the corresponding Lipschitz constant. In particular, C_1 in (5.2.10) can be chosen to be the Lipschitz constant. Further, err does not depend on the initial condition $x \in E$ and without loss of generality we assume monotonicity for the error, i.e., $\mathrm{err}(T_1, h) \geq \mathrm{err}(T_2, h)$ if $T_1 > T_2$. For the trapezoidal rule (5.2.7), (5.2.8) such a function err is given for small enough h by

$$\mathrm{err}(T, h) = C_2\big(1 + C_1^{-1}\mathrm{e}^{C_1 T}\big)h^2$$

with some appropriate constant $C_2 > 0$.
Particularly, the stability condition (5.2.10) implies that the method is uniformly convergent on $[0, T]$ in the sense that

$$\lim_{h\to 0}\max_{t\in[0,T]} |y(t, x) - \widehat{y}(t, x, h)| = 0 \qquad (5.2.11)$$

for all initial conditions x. We say that the method is convergent of order p if p is the largest integer such that for $h \to 0$

$$\max_{t\in[0,T]} |y(t, y) - \widehat{y}(t, y, h)| = \mathcal{O}(h^p), \qquad (5.2.12)$$

in which case $err(T, h) = \mathcal{O}(h^p)$. Finally, we assume that the numerical approximation satisfies a Lipschitz condition with respect to t on [0,T], i.e.,

$$|\widehat{y}(t, x, h) - \widehat{y}(s, x, h)| \leq C_3\, |t - s| \quad \forall\, t, s \in [0, T], \qquad (5.2.13)$$

where the Lipschitz constant C_3 is uniform with respect to the initial condition x. In particular, these conditions are satisfied by the one-step methods developed in [11] for the IVP (5.2.5) with properties as specified in Theorem 5.2.1.

Thus using a continuous ODE method we obtain an approximation $(\widehat{X}(t))_{t\in[0,T]}$ for the PDMP $(X(t))_{t\in[0,T]}$ by the subsequent algorithm.

Algorithm A2. *An algorithm simulating an approximation to a PDMP is given by:*

Step 1. *Set a jump counter $n = 0$ and fix the initial time $\widehat{\tau}_0 = 0$ and initial condition $\widehat{X}(0) = x_0 = (y_0, \theta_0) \in E$.*

Step 2. *Simulate a uniformly distributed random variable U and solve the IVP* (5.2.5) *with initial condition $(\widehat{X}(\widehat{\tau}_n), 0)$ numerically until*

$$\widehat{\sigma} = \inf\{t > 0 : \widehat{w}(t) = -\log U\}.$$

We take $\widehat{\tau}_{n+1} = \widehat{\tau}_n + \widehat{\sigma}$ as the numerical approximation of the next jump time. Hence, we set

$$\widehat{X}(t) = \widehat{\phi}(t - \widehat{\tau}_n, \widehat{X}(\widehat{\tau}_n)) \text{ for } t \in (\widehat{\tau}_n, \widehat{\tau}_{n+1}).$$

If $\widehat{\tau}_{n+1} \geq T$ then stop.

Step 3. *Otherwise, simulate a post-jump value* $\widehat{\theta}_{n+1}$ *for the discrete component according to the distribution* $\mu(\widehat{\phi}(\widehat{\tau}_{n+1}-\widehat{\tau}_n, \widehat{X}(\widehat{\tau}_n)), \cdot)$ *and set*

$$\widehat{X}(\widehat{\tau}_{n+1}) = \begin{pmatrix} \widehat{\phi}(\widehat{\tau}_{n+1}-\widehat{\tau}_n, \widehat{X}(\widehat{\tau}_n)) \\ \widehat{\theta}_{n+1} \end{pmatrix}.$$

Step 4. *Set* $n = n + 1$ *and start again at Step 2.*

Just as hybrid models mix stochastic jump models with continuous deterministic models, hybrid algorithms are essentially constructed by combining a simulation algorithm for the discrete stochastic events, i.e., the SSA, and a numerical method for solving differential equations. Thus, in the SSA terminology Algorithm A2 simulates the stochastic events by Gillespie's direct method, hence such an algorithm may also be called *direct hybrid method*, cf. [3]. The above kind of algorithm, though arising naturally from the model problem, is also related to numerical methods developed for JSDEs. In the case of $\lambda \equiv const.$ Algorithm A2 is exact and turns into jump-adapted methods for JSDEs, see, e.g., [99] for a discussion of jump-adapted Taylor methods, in particular for the pure jump case. We note that derivation or analysis of numerical methods for PDMPs along the lines of Itô-Taylor expansions employed in the JSDE case is not possible as general PDMPs lack a representation as a solution of a stochastic differential equation. Finally, Algorithm A2 obviously reduces to the SSA or a purely deterministic ODE method in the degenerate case if either the PDMP is piecewise constant, i.e., $g \equiv 0$, or there are no jumps present at all, i.e., $\lambda \equiv 0$.

The following theorem states conditions on the functions g and λ, which are, together with a valid numerical method for solving (5.2.5), sufficient to guarantee almost sure convergence in the sense of Definition 5.2.1. The measure μ is assumed to be as specified in Section 5.1.

Theorem 5.2.1. *Let* $(X(t))_{t\in[0,T]}$ *be a regular PDMP with phase space* $E \subseteq \mathbb{R}^d \times K$*. Let* g *be bounded, Lipschitz continuous and continuously differentiable on* E*, i.e., there exist constants* M, L *such that for all* $(y,\theta), (z,\vartheta) \in E$ *it holds*

$$|g(y,\theta)| \leq M\,, \tag{5.2.14}$$

$$|g(y,\theta) - g(z,\vartheta)| \leq L\,|(y,\theta)-(z,\vartheta)|\,. \tag{5.2.15}$$

Further, assume that the jump rate λ *is bounded, bounded away from zero, Lipschitz continuous and continuously differentiable, i.e., there exist constants* $\lambda_{\min}, \lambda_{\max}, L$

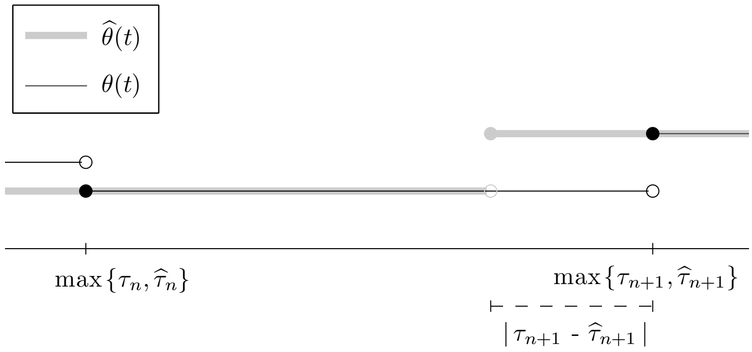


Figure 5.3: Even if the step size h is chosen small enough such that the post jump values of the piecewise constant component of the exact PDMP and its approximation coincide, we have that $\sup_{t\in[\max\{\tau_n,\widehat{\tau}_n\},\max\{\tau_{n+1},\widehat{\tau}_{n+1}\}]} |\theta(t)-\widehat{\theta}(t)| = |\theta(\tau_n)-\theta(\tau_{n+1})| > 0$ as in general $|\tau_{n+1}-\widehat{\tau}_{n+1}| > 0$ for all $h > 0$.

such that for all $(y,\theta),(z,\vartheta)\in E$ *it holds*

$$0 < \lambda_{\min} \le \lambda(y,\theta) \le \lambda_{\max} < \infty\,, \tag{5.2.16}$$

$$|\lambda(y,\theta)-\lambda(z,\vartheta)| \le L\,|(y,\theta)-(z,\vartheta)|\,. \tag{5.2.17}$$

Then the algorithms A1 *and* A2 *are well defined and the numerical approximation* $(\widehat{X}(t))_{t\in[0,T]}$ *defined by* Algorithm A2 *converges almost surely to the exact PDMP constructed by* Algorithm A1 *for all initial values* $x_0\in E$ *in the sense of* Definition 5.2.1. *Moreover, the continuous components of the PDMP and its approximation converge a.s. uniformly for* $h\to 0$*, i.e.,*

$$\lim_{h\to 0}\sup_{t\in[0,T]} |Y(t,\omega)-\widehat{Y}(t,h,\omega)| = 0 \tag{5.2.18}$$

almost surely. If, in addition, the continuous ODE method is of order p *then also the asymptotic behaviour of* (5.2.1)–(5.2.3) *and* (5.2.18) *is of order* p*.*

We briefly comment on the nature of the conditions on g and λ in the above theorem. First of all, note that without loss of generality we can choose the same Lipschitz constant in (5.2.15) and in (5.2.17). Particularly this implies that the right hand side of (5.2.5) satisfies a Lipschitz condition with constant L, hence the global existence of a unique solution is guaranteed. Moreover, the solution is Lipschitz continuous with respect to the initial condition and, as g, λ are continuously differentiable, it is also differentiable with respect to the initial condition [55]. Secondly, the assumptions that λ is bounded away from zero is necessary in order for Step 2 in Algorithm A1 to be well-posed as an event detection problem which presupposes, on the one hand, that the derivative of the event function is not zero [108], i.e., in the present case

$$\frac{\mathrm{d}}{\mathrm{d}\sigma}\Big(\int_0^\sigma \lambda(\phi(s,x))\,\mathrm{d}s + \log U\Big) \;=\; \lambda(\phi(\sigma,x)) \;\neq\; 0\,.$$

On the other hand, it is further necessary for the right hand side of (5.2.5) to be bounded in a ball around the event location in phase space [108], that is, in a ball around the point $z \in E \times \mathbb{R}_+$ where the solution ψ is in the moment its last component w hits the threshold. However, for the stochastic problem we are considering this event can occur at any point in phase space as the initial condition of the IVP (5.2.5) as well as the threshold $-\log U$ are random. Therefore the global boundedness of the right hand side of (5.2.5), i.e., the global boundedness of g and λ, as well as λ being bounded away from zero is required.

We have already mentioned the connection to a convergence result for deterministic event detection [108]. Therein the authors prove that continuous methods conserve their order for problems where one jump event has to be approximated by event detection algorithms over the approximation interval. This result, however, is extended by Theorem 5.2.1 in two ways. First, random instead of deterministic thresholds need to be considered for the simulation of PDMPs and secondly, jump times have to be calculated serially thus a simulation algorithm consists of a sequence of random event detection problems that take the result of the previous event detection as initial condition for the next one. Hence, for an error analysis we have to provide an analysis of the way the single errors of each event detection problem accumulate and are brought forward to a global level.

5.3 Proof of Theorem 5.2.1

We begin by recalling some notation which is used throughout the proof. Let the finite interval $[0,T]$ be fixed. We also fix an $\omega \in \Omega$ such that $N = N(T,\omega) < \infty$, i.e., the realised path $(X(t,\omega))_{t\in[0,T]}$ obtained by Algorithm A1 has only finitely many jumps in the interval $[0,T]$. By definition of a regular PDMP this includes all ω except a potential set of measure zero. By $\tau_n(\omega),\ n \geq 1$, we denote the jump times of the exact PDMP. Further, for a fixed step size h of the continuous ODE method used to solve the system (5.2.5) we denote by $(\widehat{X}(t,h,\omega))_{t\in[0,T]}$ the resulting path of the continuous approximation obtained by Algorithm A2 with jump times denoted by $\widehat{\tau}_n(h,\omega),\ n \geq 1$. Finally, we denote by $U_n(\omega), V_n(\omega),\ n \geq 1$, the realisations of the i.i.d. standard uniform random variables defining the nth inter-jump time and the nth post-jump value, respectively.[2] In the remainder of the proof we generally

[2]Note that in the rest of the chapter we use U_{2n-1} and U_{2n} to denote the elements of the sequence of uniform random variables defining inter-jump times and post-jump values. However, for the clarity of this long and detailed proof we think it is preferable to use different symbols U_n and

omit the dependency of any random variable on ω, as well as the dependency of the approximation on h.

First of all, we briefly discuss why both algorithms A1 and A2 are well-defined, i.e., we discuss the existence and uniqueness of trajectories for a.e. $\omega \in \Omega$. This is derived immediately from conditions (5.2.15)–(5.2.17). The Lipschitz conditions (5.2.15) and (5.2.17) guarantee the existence of a unique global solution $\psi(t,x)$ to the IVP (5.2.5) for all initial conditions $x \in E$ and the existence of a numerical approximation $\widehat{\psi}(t,x)$ follows from the fact that the continuous ODE method is well-defined. Moreover, (5.2.16) implies that $w(t,x)$ given by (5.2.4) is strictly increasing with $w(0,x) = 0$ and $w(t,x) \to \infty$ for $t \to \infty$ and hence (5.1.4) has a unique solution for all $U \in (0,1)$. For the corresponding property of the numerical approximation we note that a method may not produce continuous solutions that are increasing. However, the embedded discrete methods do produce strictly increasing and diverging iterates, which, e.g., immediately follows from the first order condition for Runge-Kutta methods, cf. [11], and condition (5.2.16). By the continuity of $\widehat{w}$ it follows that there exists a finite hitting time to every finite threshold. Uniqueness of this hitting time follows by the use of the generalised inverse in Step 2 of Algorithm A2.

We continue by explaining the basic idea of the proof before we present the arguments for the single steps in detail. As usual for the convergence analysis of numerical methods we compare the approximation and the exact PDMP at discrete time points in $[0,T]$, cf. the definition of the global error (5.2.1). Analogously to a standard approach in the convergence analysis we consider a 'local' error and show how the error is transported from the local to the global level. Then convergence follows from consistency and numerical stability of the method. However, the 'local' error we consider is not the error resulting over one discrete time step in the continuous ODE method but the error that arises over one interval between two successive jumps, which corresponds to the global error of the ODE method over one inter-jump interval.
Recall that we use a discretisation of the interval $[0,T]$ with grid points given by the jump times τ_n and $\widehat{\tau}_n$, $n = 0, \ldots, N$, for the exact and the approximate PDMP, respectively. Therefore the time discretisation for the two processes we have to compare differ in the location of their grid points as well as possibly in the number of grid points, as in general $\tau_n \neq \widehat{\tau}_n$ a.s. and a priori $T < \widehat{\tau}_N$ may be possible with positive probability for some $h > 0$. The former is not a problem as we relate the nth inter-jump interval of the exact PDMP to the nth inter-jump interval of the approximation regardless of the interval endpoints, see Fig. 5.2. For this to be well defined the latter is a problem as we need the same number of inter-jump intervals in each approximation as there are in the exact solution. To overcome this obstacle we can

V_n to distinguish the uniform random variables used to the define to two types of stochasticity in a PDMP.

- without loss of generality - assume that Algorithm A2 is stopped only after it had at least N jumps. In this way we obtain an approximation with the same number of grid points and which is defined over an interval including $[0, T]$ (at least for small enough h, cf. Section 5.3.4).

The proof is organised as follows. In a first step of the proof we derive an error bound for the error in the approximation over one inter-jump interval, i.e., the 'local' error of our approximation method. To obtain such a bound we first need a bound on the error in the time until the next jump, i.e., a bound for

$$|(\tau_{n+1} - \tau_n) - (\widehat{\tau}_{n+1} - \widehat{\tau}_n)| . \tag{5.3.1}$$

Two distinct causes introduce errors in this approximation. On the one hand, the error due to the fact that we numerically approximate the hitting times $\tau_{n+1} - \tau_n$ by $\widehat{\tau}_{n+1} - \widehat{\tau}_n$ and, on the other hand, there are perturbations in the initial conditions as in general $X(\tau_n) \neq \widehat{X}(\widehat{\tau}_n)$. Therefore, we first consider in Section 5.3.1 consistency and numerical stability with respect to perturbations in the initial data for the hitting time approximation. Then, in Section 5.3.2 we consider the 'local' error of the method in the phase space of the PDMP whereas in Section 5.3.3 we prove the first global convergence result, i.e., the limits (5.2.1) and (5.2.3). The convergence at the interval endpoint, i.e., (5.2.2), is proven in Section 5.3.4. Finally, in Section 5.3.5 we extend the discrete convergence result to a continuous result for the continuous component of the PDMP, i.e., we prove (5.2.18).

5.3.1 Consistency of the hitting time approximation

In this first part of the proof we consider (5.3.1), i.e., the error in the time until the next jump of the exact PDMP started in x to the time until the next jump obtained by the numerical solution of (5.2.5) started from a perturbed initial value $\widehat{x}$. Note that we consider the PDMP and its approximation just on one inter-jump interval and hence the components $\theta(t)$, $\widehat{\theta}(t)$ remain constant, although possibly distinct.
To keep the presentation simple we omit the index n in the following, only to be reintroduced in Section 5.3.3 when we consider the global error of the method. Further, without loss of generality we set for the exact and approximate inter-jump interval the left endpoint to 0 and denote by τ and $\widehat{\tau}$ the jump time of the exact PDMP and its approximation, respectively. Hence the error (5.3.1) can be written as $|\tau - \widehat{\tau}|$ and the two time points τ, $\widehat{\tau}$ satisfy the equations

$$-\log U = w(\tau, x) \quad \text{and} \quad -\log U = \widehat{w}(\widehat{\tau}, \widehat{x}) .$$

That is, τ and $\widehat{\tau}$ are the respective hitting times of the random threshold $-\log U$ of the component w of the exact and approximate solution of the IVP (5.2.5).

In order to estimate $|\tau - \widehat{\tau}|$ we introduce the additional auxiliary time $\widetilde{\tau}$ as the time until the next jump of the exact process started in the perturbed initial value $\widehat{x}$ which is given by

$$-\log U = w(\widetilde{\tau}, \widehat{x}) \,.$$

Then the triangle inequality yields the initial estimate

$$|\tau - \widehat{\tau}| \leq |\tau - \widetilde{\tau}| + |\widetilde{\tau} - \widehat{\tau}| \,. \tag{5.3.2}$$

Here the first term in the right hand side measures the error in the exact hitting time with respect to perturbations in the initial data whereas the second denotes the numerical error of the method without perturbations in the initial condition. We continue estimating the two terms separately and start with the latter.

We introduce $r(\tau, h, x)$ as the error in the last component of of the continuous method defined by

$$w(\tau, x) = \widehat{w}(\tau, x) + r(\tau, h, x),$$

where obviously due to (5.2.10) it is valid that $|r(s, h, x)| \leq \mathrm{err}(\tau, h)$ for all $s \leq \tau$ and all $x \in E$. By definition it follows that

$$\begin{aligned} 0 &= \widehat{w}(\widehat{\tau}, \widehat{x}) + \log U \\ &= w(\widehat{\tau}, \widehat{x}) - r(\widehat{\tau}, \widehat{x}, h) + \log U \\ &= w(\widetilde{\tau}, \widehat{x}) + (\widehat{\tau} - \widetilde{\tau})\, \nabla_t w(\vartheta, \widehat{x}) - r(\widehat{\tau}, \widehat{x}, h) + \log U \end{aligned}$$

for some $\vartheta \in [\min\{\widetilde{\tau}, \widehat{\tau}\}, \max\{\widetilde{\tau}, \widehat{\tau}\}]$. Here we have employed the Mean Value Theorem for the last equality. As $w(\widetilde{\tau}, \widehat{x}) + \log U = 0$, this is equivalent to

$$(\widetilde{\tau} - \widehat{\tau})\, \nabla_t w(\vartheta, \widehat{x}) = -r(\widehat{\tau}, \widehat{x}, h),$$

and we obtain due to the definition of w (5.2.4) and condition (5.2.16) the estimate

$$|\widetilde{\tau} - \widehat{\tau}| \leq \frac{|r(\widehat{\tau}, \widehat{x}, h)|}{\lambda_{\min}} \leq \frac{\mathrm{err}(\widehat{\tau}, h)}{\lambda_{\min}} \,. \tag{5.3.3}$$

Next we have to consider the term $|\tau - \widetilde{\tau}|$ in the right hand side of (5.3.2). To bound this term we aim for an estimate on the distance in time by the distance of the initial conditions in phase space of pairs $(\tau, x), (\widetilde{\tau}, \widehat{x})$ satisfying $w(\tau, x) = w(\widetilde{\tau}, \widehat{x})$, i.e., an estimate of the form

$$|\tau - \widetilde{\tau}| \leq C_1\, |x - \widehat{x}| \tag{5.3.4}$$

for some positive random constant C_1 depending only on τ and U.
Due to the assumptions in Theorem 5.2.1 the solution of the IVP (5.2.5) depends differentiably on the initial value. Thus, proceeding analogously as for error estimate (5.3.3) we obtain by the Mean Value Theorem that

$$\begin{aligned} 0 &= w(\widetilde{\tau},\widehat{x}) - w(\tau,x) \\ &= w(\tau,x) + \nabla_t w(\vartheta,\zeta)\,(\widetilde{\tau}-\tau) + \nabla_x w(\vartheta,\zeta)\cdot(\widehat{x}-x) - w(\tau,x) \end{aligned}$$

for some mean values $\vartheta \in [\min\{\widetilde{\tau},\widehat{\tau}\},\max\{\widetilde{\tau},\widehat{\tau}\}]$, $\zeta \in \{y : y = tx+(1-t)\widetilde{x},\ t\in[0,1]\}$.
By the Cauchy-Schwarz inequality it follows that

$$|\tau-\widetilde{\tau}| \le \Big|\frac{\nabla_x w(\vartheta,\zeta)}{\lambda(\phi(\vartheta,\zeta))}\Big| \cdot |x-\widehat{x}|. \tag{5.3.5}$$

This yields an estimate of the form (5.3.4) if we bound the first factor in the right hand side of (5.3.5) uniformly. As the right hand side of the IVP (5.2.5) is Lipschitz continuous in y and θ with constant L it follows that $\phi(t,x)$, which denotes the components y and θ of the solution of (5.2.5), depends Lipschitz continuously on the initial condition with Lipschitz constant e^{Lt}. Moreover, as by the assumptions of the theorem the solution is continuously differentiable with respect to the initial condition it follows that the derivatives satisfy

$$|\nabla_x w(s,x)| \le \mathrm{e}^{Ls} \quad \forall\, x\in E,\ s\ge 0\,.$$

Hence we obtain for the first factor in the right hand side of (5.3.5) the estimate

$$\Big|\frac{\nabla_x w(\vartheta,\zeta)}{\lambda(\phi(\vartheta,\zeta))}\Big| \le \frac{\mathrm{e}^{\,L\max\{\widetilde{\tau},\tau\}}}{\lambda_{\min}} \quad \forall\,\zeta\in E\,.$$

Finally, we consider an a priori bound for the maximum $\max\{\tau,\widetilde{\tau}\}$ in the exponent. Obviously, it is valid that $\max\{\widetilde{\tau},\tau\} \le \tau + |\widetilde{\tau}-\tau|$ and as $\tau,\widetilde{\tau}$ are hitting times of the exact solution their difference is globally bounded due to (5.2.16): A straightforward calculation shows that $|\tau-\widetilde{\tau}| \le \frac{\lambda_{\max}-\lambda_{\min}}{\lambda_{\max}\lambda_{\min}}\,(-\log U)$ independently of the initial conditions x and $\widehat{x}$. Thus, these estimates applied to (5.3.5) yield

$$|\tau-\widetilde{\tau}| \le \frac{\mathrm{e}^{L\tau}}{\lambda_{\min}} U^{-\delta}\,|x-\widehat{x}|, \tag{5.3.6}$$

where $\delta = L\frac{\lambda_{\max}-\lambda_{\min}}{\lambda_{\max}\lambda_{\min}}$ and hence we have arrived at an estimate of the form (5.3.4).
Overall an application of the estimates (5.3.3) and (5.3.6) to the right hand side of

(5.3.2) yields the stability estimate

$$|\tau - \widehat{\tau}| \leq \frac{1}{\lambda_{\min}} \big(\mathrm{e}^{L\tau} U^{-\delta} \, |x - \widehat{x}| + \mathrm{err}(\widehat{\tau}, h) \big) \, . \tag{5.3.7}$$

In this error bound the first term in the right hand side bounds the error due to perturbations in the initial value, whereas the second term corresponds to the error resulting from the numerical approximation of the next jump time.

5.3.2 The local error in phase space

Next we derive a bound of the error over one inter-jump interval in the phase space E. We denote the post-jump values of the exact PDMP and its solution, started at initial conditions x and $\widehat{x}$, after one inter-jump interval by $X(\tau)$ and $\widehat{X}(\widehat{\tau})$, respectively. Hence, given a realisation of a standard uniformly distributed random variable V the 'local' error in the phase space is given by

$$|X(\tau) - \widehat{X}(\widehat{\tau})| = |\phi(\tau, x) + \Theta(\phi(\tau, x), V) - \widehat{\phi}(\widehat{\tau}, \widehat{x}) - \Theta(\widehat{\phi}(\widehat{\tau}, \widehat{x}), V)| \, . \tag{5.3.8}$$

We recall that Θ is the measurable function used to define the post-jump values from realisations of standard uniform random variables, cf. (5.1.5) in Section 5.1.1. However, note that here we use $\Theta(\,\cdot\,,\,\cdot\,)$ as an abbreviation of the correct but cumbersome notation $(0, \Theta(\,\cdot\,,\,\cdot\,))$. In order to bound the error (5.3.8), we first estimate the difference in the continuous components in (5.3.8). Thus, as a first estimate the triangle inequality yields

$$|\phi(\tau, x) - \widehat{\phi}(\widehat{\tau}, \widehat{x})| \ \leq \ |\phi(\tau, x) - \phi(\widehat{\tau}, \widehat{x})| + |\phi(\widehat{\tau}, \widehat{x}) - \widehat{\phi}(\widehat{\tau}, \widehat{x})|. \tag{5.3.9}$$

The second term in the right hand side of (5.3.9) is bounded by $err(\widehat{\tau}, h)$ due to the stability of the ODE method (5.2.10). Hence it remains to consider the first term. A second application of the triangle inequality, this time to the first term in the right hand side of (5.3.9), yields

$$|\phi(\tau, x) - \phi(\widehat{\tau}, \widehat{x})| \leq |\phi(\tau, x) - \phi(\tau, \widehat{x})| + |\phi(\tau, \widehat{x}) - \phi(\widehat{\tau}, \widehat{x})|. \tag{5.3.10}$$

As the solution of the IVP (5.2.5) depends Lipschitz continuously on the initial condition the first term in the right hand side of (5.3.10) satisfies

$$|\phi(\tau, x) - \phi(\tau, \widehat{x})| \leq \mathrm{e}^{L\tau} \, |x - \widehat{x}|.$$

To estimate the second term in the right hand side of (5.3.10) we set $\tau_{\min} = \min\{\tau, \widehat{\tau}\}$

and obtain

$$\begin{aligned} |\phi(\tau,x)-\phi(\widehat{\tau},x)| &= |\phi(\tau_{\min},x)-\phi(\tau_{\min}+|\tau-\widehat{\tau}|,x)| \\ &= \Big|\int_0^{|\tau-\widehat{\tau}|}\begin{pmatrix} g(\phi(s,\phi(\tau_{\min},x))) \\ 0 \end{pmatrix}\mathrm{d}s\Big| \\ &\le \int_0^{|\tau-\widehat{\tau}|}|g(\phi(\tau_{\min}+s,x))|\mathrm{d}s \\ &\le M\,|\tau-\widehat{\tau}|\,. \end{aligned} \tag{5.3.11}$$

Hence, overall, we obtain for the left hand side of (5.3.10) the estimate

$$|\phi(\tau,x)-\phi(\widehat{\tau},\widehat{x})| \le \mathrm{e}^{L\tau}\,|x-\widehat{x}|+M\,|\tau-\widehat{\tau}|.$$

Thus employing estimate (5.3.7) we get an error bound for the left hand side of (5.3.9) as

$$|\phi(\tau,x)-\widehat{\phi}(\widehat{\tau},\widehat{x})| \le \mathrm{e}^{L\tau}\big(1+\lambda_{\min}^{-1}\,M\,U^{-\delta}\big)\,|x-\widehat{x}|+(1+\lambda_{\min}^{-1}\,M)\,\mathrm{err}(\widehat{\tau},h)\,. \tag{5.3.12}$$

That is (5.3.12) bounds the error of the approximation of the continuous part over one inter-jump interval. The error estimate in its right hand side is split into two parts. The first term estimates the error due to perturbations in the initial data and the second term estimates the error due to the numerical method employed for solving (5.2.5).

Finally, we bound the error introduced by the jump heights Θ for jumps at different points in phase space. In the following we do not only bound the error due to the jump heights, but show that for sufficiently small step size this error actually vanishes. Define

$$\mathrm{dist}(K) := \sup_{\theta,\vartheta\in K}|\theta-\vartheta|\,.$$

With this definition we have that for all $x,y\in E$

$$|\Theta(x,\cdot)-\Theta(y,\cdot)| \le 2\ \mathrm{dist}(K) < \infty,$$

as K is bounded by assumption. Then the considerations regarding the structure of the function Θ in Section 5.1.1 yield for the error the estimate

$$|\Theta(\phi(\tau,x),V)-\Theta(\widehat{\phi}(\widehat{\tau},\widehat{x}),V)| \le 2\,\mathrm{dist}(K)\ \chi(\phi(\tau,x),\widehat{\phi}(\widehat{\tau},\widehat{x}),V)\,,$$

where χ is a binary function

$$\chi(x,y,u) = \begin{cases} 0 & \text{if } |x-y| < \kappa(x,u), \\ \\ 1 & \text{otherwise,} \end{cases}$$

cf. Section 5.1.1. The variable V is again a realisation of a standard uniform distributed random variable. Note, as we have already proven that then $|\phi(\tau,x)-\widehat{\phi}(\widehat{\tau},\widehat{x})|$ converges to zero, cf. (5.3.12), we get $\chi(\phi(\tau,x),\widehat{\phi}(\widehat{\tau},\widehat{x}),V) = 0$ for small enough h and $|\widehat{x}-x|$.

To conclude, overall we obtain for the 'local' error (5.3.8) the estimate

$$\begin{aligned} |X(\tau)-\widehat{X}(\widehat{\tau})| \;&\leq\; \mathrm{e}^{L\tau}\big(1+\lambda_{\min}^{-1}\, M\, U^{-\delta}\big)\,|x-\widehat{x}| + (1+\lambda_{\min}^{-1}\, M)\,\mathrm{err}(\widehat{\tau},h) \\ &\qquad + 2\ \mathrm{dist}(K)\ \chi(\phi(\tau,x),\widehat{\phi}(\widehat{\tau},\widehat{x}),V)\,. \end{aligned} \tag{5.3.13}$$

5.3.3 The discrete global error

In order to prove convergence we have to be able to infer from the 'local' error to the global error. Thus we reintroduce the subindices n, hence $\tau_n,\ \widehat{\tau}_n,\ n=1,\ldots,N$, denote the nth jump times of the PDMP and its approximation, respectively, and $X(\tau_n)$, $\widehat{X}(\widehat{\tau}_n)$ denote the post-jump values of these processes at their jump times. We first show by induction over n that the global error at each jump time, i.e., $|X(\tau_n)-\widehat{X}(\widehat{\tau}_n)|$ for all $n=1,\ldots,N$, converges to zero for $h\to 0$. Afterwards we derive an upper bound for the global error uniform over all jump times to determine the asymptotic order of decrease of the global error. The splitting of the proof at this point into these two parts is necessary as in order to derive the asymptotic order of convergence we first need to show that the errors in the post jump values of θ vanish for small enough step sizes h. However, this is a necessary condition of convergence and we can assume it holds for small enough step sizes once the convergence is established.

For the induction basis we set $X(0)=\widehat{X}(0)$, then by (5.3.13)

$$|X(\tau_1)-\widehat{X}(\widehat{\tau}_1)| \;\leq\; (1+\lambda_{\min}^{-1}M)\,err(\widehat{\tau}_1,h) + 2\ \mathrm{dist}(K)\ \chi(\phi(\tau,x_0),\widehat{\phi}(\widehat{\tau},x_0),V_1)\,.$$

Obviously for h small enough the last term on the right hand side vanishes due to the convergence of the deterministic method. Then $h\to 0$ implies that $(1+\lambda_{\min}^{-1}M)\,err(\widehat{\tau}_1,h)\to 0$ and therefore $\lim_{h\to 0}|X(\tau_1)-\widehat{X}(\widehat{\tau}_1)|=0$.

Next, for the induction step we assume that $\lim_{h\to 0}|X(\tau_n)-\widehat{X}(\widehat{\tau}_n)| = 0$ holds for some $n \geq 0$. Then, for small enough h we have that

$$
\begin{aligned}
|X(\tau_{n+1}) - \widehat{X}(\widehat{\tau}_{n+1})| \;\leq\; & \mathrm{e}^{L(\tau_{n+1}-\tau_n)}\big(1 + \lambda_{\min}^{-1}\, M\, U_{n+1}^{-\delta}\big)\, |X(\tau_n) - \widehat{X}(\widehat{\tau}_n)| \\
& + (1 + \lambda_{\min}^{-1}\, M)\, \mathrm{err}(\widehat{\tau}_{n+1} - \widehat{\tau}_n, h)\,,
\end{aligned}
$$

and hence $\lim_{h\to 0} |X(\tau_{n+1}) - \widehat{X}(\widehat{\tau}_{n+1})| = 0$.

Therefore overall we find that the method is convergent in the sense that

$$
\lim_{h\to 0} \max_{n=1,\ldots,N} |X(t_n) - \widehat{X}(\widehat{t}_n)| = 0\,. \tag{5.3.14}
$$

Analogously we show that also the approximate jump times converge to the exact jump times. The error estimate (5.3.7) yields

$$
\begin{aligned}
&|\tau_{n+1} - \widehat{\tau}_{n+1}| \\
&\quad \leq\; |(\tau_{n+1} - \tau_n) - (\widehat{\tau}_{n+1} - \widehat{\tau}_n) + \tau_n - \widehat{\tau}_n| \\
&\quad \leq\; \mathrm{e}^{L(\tau_{n+1}-\tau_n)}\, \lambda_{\min}^{-1}\, U_{n+1}^{-\delta}\, |X(\tau_n) - \widehat{X}(\widehat{\tau}_n)| + \lambda_{\min}^{-1}\, \mathrm{err}(\widehat{\tau}_{n+1} - \widehat{\tau}_n, h) + |\tau_n - \widehat{\tau}_n|.
\end{aligned}
\tag{5.3.15}
$$

Then (5.3.14) and another inductive argument yield that

$$
\lim_{h\to 0} \max_{n=1,\ldots,N} |\tau_n - \widehat{\tau}_n| = 0\,. \tag{5.3.16}
$$

As these considerations leading to (5.3.14) and (5.3.16) are valid for almost all $\omega \in \Omega$ this completes the proof of (5.2.1) and (5.2.3). However, to satisfy the condition of pathwise convergence, Definition 5.2.1, it remains to show (5.2.2), i.e., the convergence at the interval endpoint which is deferred to Section 5.3.4. At this point we continue to derive estimates for the global errors (5.2.1) and (5.2.3) to obtain the rate of convergence for the error in phase space at the jump times, as well as the error in the jump times itself.

To this end we recursively apply (5.3.13) to its right hand side arriving at

$$
\begin{aligned}
|X(\tau_n) - \widehat{X}(\widehat{\tau}_n)| \;\leq\; & Z_n\, |x_0 - \widehat{x}_0| \\
& +\; Z_n \sum_{k=1}^{n} Z_k^{-1} \Big[(1 + \lambda_{\min}^{-1}\, M)\, \mathrm{err}(\widehat{\tau}_k - \widehat{\tau}_{k-1}, h) + \chi(X(\tau_{k-}), \widehat{X}(\widehat{\tau}_{k-}), V_k)\Big],
\end{aligned}
\tag{5.3.17}
$$

where for $k = 1\dots, N$ we set

$$Z_k = e^{L\tau_k} \prod_{i=1}^{k} (1 + \lambda_{\min}^{-1} M\, U_i^{-\delta}) .$$

As we are considering the asymptotic order of convergence for $h \to 0$ we can assume that h is small enough such that $\chi(X(\tau_{n-}), \widehat{X}(\widehat{\tau}_{n-}), V_n) = 0$ for all $n = 1, \dots, N$. Thus for $x_0 = \widehat{x}_0$ we obtain that for all $n = 1, \dots, N$

$$|X(\tau_n) - \widehat{X}(\widehat{\tau}_n)| \;\leq\; (1 + \lambda_{\min}^{-1} M)\, Z_n \sum_{k=1}^{n} Z_k^{-1}\, \mathrm{err}(\widehat{\tau}_k - \widehat{\tau}_{k-1}, h), \tag{5.3.18}$$

and therefore

$$\max_{n=1,\dots,N} |X(\tau_n) - \widehat{X}(\widehat{\tau}_n)| \;\leq\; (1 + \lambda_{\min}^{-1} M)\, Z_N \sum_{k=1}^{N} Z_k^{-1}\, \mathrm{err}(\widehat{\tau}_k - \widehat{\tau}_{k-1}, h) .$$

It follows that if $err(\tau, h) = \mathcal{O}(h^p)$ then

$$\max_{n=1,\dots,N} |X(\tau_n) - \widehat{X}(\widehat{\tau}_n)| = \mathcal{O}(h^p) . \tag{5.3.19}$$

Analogously we derive the order of convergence for the jump time approximations. Starting with $\tau_0 = \widehat{\tau}_0$ and recursively applying (5.3.15) to its right hand side yields

$$|\tau_n - \widehat{\tau}_n| \;\leq\; \lambda_{\min}^{-1} \sum_{k=1}^{n} \Big[\mathrm{err}(\widehat{\tau}_k - \widehat{\tau}_{k-1}, h) + e^{L(\tau_k - \tau_{k-1})}\, U_k^{-\delta}\; |X(\tau_{k-1}) - \widehat{X}(\widehat{\tau}_{k-1})| \Big] .$$

Employing the estimate (5.3.18) for the error in phase space yields for small enough h that for all $n = 1, \dots, N$

$$|\tau_n - \widehat{\tau}_n| \;\leq\; \sum_{k=1}^{n} \lambda_{\min}^{-1} \Big[1 + (1 + \lambda_{\min}^{-1} M)\, Z_k^{-1} \sum_{j=k}^{n-1} Z_j\, U_{j+1}^{-\delta}\, e^{L(\tau_{j+1} - \tau_j)} \Big]\, \mathrm{err}(\widehat{\tau}_k - \widehat{\tau}_{k-1}, h) .$$

Denoting the factor multiplying err in each summand by W_k^n yields that for all $n = 1, \dots, N$

$$|\tau_n - \widehat{\tau}_n| \leq \sum_{k=1}^{N} W_k^N\, \mathrm{err}(\widehat{\tau}_k - \widehat{\tau}_{k-1}, h)$$

and thus if $err(\tau, h) = \mathcal{O}(h^p)$, then

$$\max_{n=1,\dots,N} |\tau_n - \widehat{\tau}_n| = \mathcal{O}(h^p) . \tag{5.3.20}$$

Again, these considerations leading to (5.3.20) are valid for almost all $\omega \in \Omega$, and

(5.2.1) and (5.2.3) follow with the proposed order.

5.3.4 Convergence at the interval endpoint

Finally, for the method to satisfy the condition of pathwise convergence we establish convergence at the interval endpoint T. To this end we employ the estimate (5.3.15) on the error for the approximation of the $(N+1)$th jump time. This yields

$$\begin{aligned}
&|\tau_{N+1} - \widehat{\tau}_{N+1}| \\
&\qquad \le \mathrm{e}^{L(\tau_{N+1}-\tau_N)}\,\lambda_{\min}^{-1}\,U_{N+1}^{-\delta}\,|X(\tau_N) - \widehat{X}(\widehat{\tau}_N)| + \lambda_{\min}^{-1}\,\mathrm{err}(\widehat{\tau}_{N+1} - \widehat{\tau}_N, h) + |\tau_N - \widehat{\tau}_N|\,.
\end{aligned}$$

As all terms in the right hand side of this inequality converge to zero for $h \to 0$ it follows that there exists an $h^* > 0$, depending on U_{N+1}, such that $\widehat{\tau}_{N+1} > T$ for all $h < h^*$. In this case we then obtain for the error at the interval endpoint the estimate

$$\begin{aligned}
&|X(T) - \widehat{X}(T)| \\
&= |\phi(T-\tau_N, X(t_N)) - \widehat{\phi}(T-\widehat{\tau}_N, \widehat{X}(\widehat{\tau}_N))| \\
&\le |\phi(T-\tau_N, X(\tau_N)) - \phi(T-\widehat{\tau}_N, X(\tau_N))| + |\phi(T-\widehat{\tau}_N, X(\tau_N)) - \widehat{\phi}(T-\widehat{\tau}_N, \widehat{X}(\widehat{\tau}_N))| \\
&\le M\,|\tau_N - \widehat{\tau}_N| + \mathrm{e}^{L(T-\widehat{\tau}_N)}\,|X(\tau_N) - \widehat{X}(\widehat{\tau}_N)| + \mathrm{err}(T-\widehat{\tau}_N, h).
\end{aligned} \tag{5.3.21}$$

Here we have used in the last inequality on the first term an estimate analogous to (5.3.11) and on the second term the method's stability estimate (5.2.10). Due to the convergence results in the previous section leading to (5.3.19) and (5.3.20) and the assumptions on the numerical method the final estimate (5.3.21) converges to zero for $h \to 0$ for almost all $\omega \in \Omega$ and if the method is of order p then almost surely

$$|X(T) - \widehat{X}(T)| = \mathcal{O}(h^p)\,.$$

5.3.5 The continuous global error

We finish the proof by the extension of the statement (5.3.19) providing convergence at the jump times to its uniform version (5.2.18). Uniform convergence can only hold for the continuous component $Y(t)$ and cannot hold for the discontinuous component $\theta(t)$, cf. Fig. 5.3, which can be easily shown rigorously, cf. (5.3.23) below.

To prove the results in this section we first derive a simple auxiliary result on the asymptotic behaviour of the approximate jump times. As stated in the last section the convergence of the jump times (5.3.16) implies that $\widehat{\tau}_N > T$ for all $h < h^*$ for some h^*. Moreover there exists an $h^{**} < h^*$ such that for all $h < h^{**}$ it is valid that

$$(\tau_{n-1}, \tau_n) \cap (\widehat{\tau}_{n-1}, \widehat{\tau}_n) \neq \emptyset \quad \forall\, n = 1, \ldots, N. \tag{5.3.22}$$

That is, (5.3.22) says, that for small enough h the nth inter-jump interval of the exact PDMP and its approximation overlap for all n.
We prove (5.3.22) by contradiction. Assume that there exists an $n \leq N$ such that $(\tau_{n-1}, \tau_n) \cap (\widehat{\tau}_{n-1}, \widehat{\tau}_n) = \emptyset$ for all small h, then either

$$\tau_{n-1} < \tau_n \leq \widehat{\tau}_{n-1} \quad \text{or} \quad \widehat{\tau}_n \leq \tau_{n-1} < \tau_n \,,$$

which implies that either

$$0 < |\tau_n - \tau_{n-1}| \leq |\widehat{\tau}_{n-1} - \tau_{n-1}| \quad \text{or} \quad 0 < |\tau_{n-1} - \tau_n| \leq |\widehat{\tau}_n - \tau_n|.$$

Both are contradictions to $\lim_{h\to 0} \max_{n=0,\ldots,N} |\tau_n - \widehat{\tau}_n| = 0$ and thus (5.3.22) is proven.

For the remainder of this section we assume that $h < h^{**}$, hence $\widehat{t}_{N+1} > T$ and (5.3.22) holds. Thus a partition of the time interval $[0, T]$ is given by the following discretisation points

$$0 =: s_0 < s_1 := \max\{\tau_1, \widehat{\tau}_1\} < \ldots < s_N := \max\{\tau_N, \widehat{\tau}_N\} < s_{N+1} := T\,.$$

It now immediately follows that a uniform convergence result cannot hold for the component $\theta(t)$ as on every interval $[s_n, s_{n+1}]$ it is valid that

$$\sup_{t\in[s_n,s_{n+1}]} |\theta(t) - \widehat{\theta}(t)| = |\theta(t_n) - \theta(t_{n+1})| > 0 \quad \forall\, h \leq h^{**}\,. \tag{5.3.23}$$

This holds as the piecewise constant components $\theta(t)$ and $\widehat{\theta}(t)$ are equal at s_n and both jump exactly one time within the interval $[s_n, s_{n+1}]$ to the same post-jump value. Recall that we assume h small enough such that the errors in the post-jump values vanish. By definition, one jumps at s_{n+1} and the other strictly before. Therefore their maximal difference over the interval $[s_n, s_{n+1}]$ is exactly the jump height. Hence, the argument in Fig. 5.3 is made mathematically precise.

After these preliminary considerations we proceed to obtain the continuous convergence result (5.2.18) on the method's continuous component $\widehat{Y}(t)$. To this end we use the new time grid $s_0, \ldots, s_{N+1}$ and estimate the uniform error of the continuous component $\widehat{Y}(t)$ on each interval $[s_n, s_{n+1}]$. Note that on each of these intervals either

$$\text{i.)} \quad \widehat{\tau}_{n+1} < \tau_{n+1} \quad \text{or} \quad \text{ii.)} \quad \tau_{n+1} < \widehat{\tau}_{n+1}.$$

We first consider case i.). Recall that we use $y(t,x)$, $\widehat{y}(t,x)$ to denote the first component of the exact and numerical solution of the IVP (5.2.5) and we use the symbol $\vee$ to denote the maximum of two expressions. Then

$$\begin{aligned}\max_{t\in[s_n,s_{n+1}]}|Y(t)-\widehat{Y}(t)| &= \max_{t\in[s_n,\widehat{\tau}_{n+1}]}|Y(t)-\widehat{Y}(t)| \vee \max_{t\in[\widehat{\tau}_{n+1},\tau_{n+1}]}|Y(t)-\widehat{Y}(t)| \\ &= \max_{s\in[0,\widehat{\tau}_{n+1}-s_n]}|y(s,X(s_n))-\widehat{y}(s,\widehat{X}(s_n))| \\ &\quad \vee \max_{s\in[0,\tau_{n+1}-\widehat{\tau}_{n+1}]}|y(s,X(\widehat{\tau}_{n+1}))-\widehat{y}(s,\widehat{X}(\widehat{\tau}_{n+1}))|.\end{aligned}$$

Both error terms in the right hand side can be estimated using the stability of the continuous ODE method (5.2.10), which yields

$$\begin{aligned}\max_{t\in[s_n,s_{n+1}]}|Y(t)-\widehat{Y}(t)| \;\leq\; &\Big[\mathrm{e}^{L(s_{n+1}-s_n)}\,|X(s_n)-\widehat{X}(s_n)| \\ &\vee \mathrm{e}^{L(\tau_{n+1}-\widehat{\tau}_{n+1})}\,|X(\widehat{\tau}_{n+1})-\widehat{X}(\widehat{\tau}_{n+1})|\Big] + \mathrm{err}(s_{n+1}-s_n,h).\end{aligned}$$

As we assume that h is small enough such that $\theta(\tau_n)=\widehat{\theta}(\widehat{\tau}_n)$ and hence $\theta(s_n)=\widehat{\theta}(s_n)$, we obtain

$$\begin{aligned}&\max_{t\in[s_n,s_{n+1}]}|Y(t)-\widehat{Y}(t)| \\ &\leq \Big[\mathrm{e}^{L(s_{n+1}-s_n)}|Y(s_n)-\widehat{Y}(s_n)| \vee \\ &\qquad \mathrm{e}^{L(\tau_{n+1}-\widehat{\tau}_{n+1})}\big(|X(\widehat{\tau}_{n+1})-X(\tau_{n+1})|+|X(\tau_{n+1})-\widehat{X}(\widehat{\tau}_{n+1})|\big)\Big] + \mathrm{err}(s_{n+1}-s_n,h) \\ &\leq \Big[\mathrm{e}^{L(s_{n+1}-s_n)}|Y(s_n)-\widehat{Y}(s_n)| \vee \\ &\qquad \mathrm{e}^{L(\tau_{n+1}-\widehat{\tau}_{n+1})}\big(M\,|\widehat{\tau}_{n+1}-\tau_{n+1}|+|X(\tau_{n+1})-\widehat{X}(\widehat{\tau}_{n+1})|\big)\Big] + \mathrm{err}(s_{n+1}-s_n,h).\end{aligned}$$

Due to the results (5.3.19) and (5.3.20) it follows that

$$\mathrm{e}^{L(\tau_{n+1}-\widehat{\tau}_{n+1})}\big(M\,|\widehat{\tau}_{n+1}-\tau_{n+1}|+|X(\tau_{n+1})-\widehat{X}(\widehat{\tau}_{n+1})|\big)+\mathrm{err}(s_{n+1}-s_n,h)=\mathcal{O}(h^p)\,.$$

Hence we obtain the recursive relation

$$\max_{t\in[s_n,s_{n+1}]}|Y(t)-\widehat{Y}(t)| \leq \mathrm{e}^{L(s_{n+1}-s_n)}\,|Y(s_n)-\widehat{Y}(s_n)|+\mathcal{O}(h^p)\,. \tag{5.3.24}$$

Secondly, for the case ii.), i.e., $\tau_{n+1} < \widehat{\tau}_{n+1}$, we analogously obtain the estimate

$$\begin{aligned}
&\max_{t\in[s_n,s_{n+1}]} |Y(t) - \widehat{Y}(t)| \\
&\leq \Big[\mathrm{e}^{L(s_{n+1}-s_n)}\,|Y(s_n) - \widehat{Y}(s_n)| \vee \\
&\qquad \mathrm{e}^{L(\widehat{\tau}_{n+1}-\tau_{n+1})}\big(|X(\tau_{n+1}) - \widehat{X}(\widehat{\tau}_{n+1})| + |\widehat{X}(\widehat{\tau}_{n+1}) - \widehat{X}(\tau_{n+1})|\big)\Big] + \mathrm{err}(s_{n+1}-s_n, h).
\end{aligned}$$

To estimate the term $|\widehat{X}(\widehat{\tau}_{n+1}) - \widehat{X}(\tau_{n+1})|$ we employ the Lipschitz condition (5.2.13). Hence we obtain again a recursive relation of the form (5.3.24) . Thus such a relation holds for the interval $[s_n, s_{n+1}]$ in either cases.

Finally, recursively applying (5.3.24) to its right hand side yields the estimate

$$\max_{n=0,\ldots,N}\ \max_{t\in[s_n,s_{n+1}]} |Y(t) - \widehat{Y}(t)| \leq \mathrm{e}^{LT}|Y(0) - \widehat{Y}(0)| + \mathcal{O}(h^p)\,.$$

These considerations hold true for almost all $\omega \in \Omega$, hence the uniform convergence (5.2.18) follows for $Y(0) = \widehat{Y}(0)$ and the proof of Theorem 5.2.1 is completed.

5.4 Extensions of the convergence theorem

It is straightforward to extend Theorem 5.2.1 to general PDMPs with right continuous paths. That is, we consider PDMPs that do not have any qualitative differences in their components: All components allow for discontinuities, i.e., μ is a Markov kernel onto $\mathbb{R}^{d+m}$, and in between jumps the trajectories follow a deterministic motion given by an ODE

$$\begin{pmatrix}\dot{y}\\ \dot{\theta}\end{pmatrix} = \begin{pmatrix} g_1(y,\theta)\\ g_2(y,\theta)\end{pmatrix}\,,$$

i.e., the component θ is not piecewise constant anymore. This is the class of PDMPs as considered in [64]. Clearly, a continuous convergence result such as (5.2.18) cannot hold anymore. However, the statement about the asymptotic behaviour of the errors at the jump times, at the interval end point and the errors in the jump times is still valid.

Considerations as in Section 5.1.1 yield the existence of functions $\widetilde{\Theta}_n : E \times (0,1) \to \mathbb{R}^{d+m}$, $n \geq 1$, providing realisations of the random jump heights from realisations of standard uniformly distributed random variables. Instead of condition (5.1.7) we impose on these functions the following Lipschitz condition: for all $x, y \in \mathbb{R}^d$ and all $n \geq 1$ it is valid that almost surely

$$|\widetilde{\Theta}_n(x, U_{2n}) - \widetilde{\Theta}_n(y, U_{2n})| \leq C_2\,|x - y|\,, \tag{5.4.1}$$

where U_{2n}, $n \geq 1$ are the standard uniformly distributed random variables defining the post-jump values in the exact and approximate simulation algorithm A1 and A2.

Corollary 5.4.1. *Let* $(X(t))_{t\in[0,T]}$ *and* $(\widehat{X}(t))_{t\in[0,T]}$ *be a regular PDMP and its approximation, respectively, which have right-continuous paths. For the distribution of the post jump values condition* (5.4.1) *is satisfied. Then under the conditions of* Theorem 5.2.1 *the global error satisfies* (5.2.1)–(5.2.3), *i.e., the method converges pathwise in the sense of Definition 5.2.1. Moreover, if the continuous ODE method is of order* p *then the order of convergence is* p.

Proof. The proof of (5.2.1)–(5.2.3) in the case of this general class of PDMPs works analogously to the proof of Theorem 5.2.1 in Section 5.3. The arguments employed become even less technical. In particular, the last paragraph in Section 5.3.2 dealing with the error in the jump heights becomes redundant as does the induction argument in Section 5.3.3. We restrict the presentation to point out the main difference.

The proof proceeds as the proof in Section 5.3 with the main difference in estimating the local error in phase space, cf. Section 5.3.2. Using the Lipschitz condition (5.4.1) we first obtain from (5.3.8) the estimate

$$|X(\tau) - \widehat{X}(\widehat{\tau})| = (1 + C_2)\,|\phi(\tau, x) - \widehat{\phi}(\widehat{\tau}, \widehat{x})|\,.$$

Then, estimating the difference $|\phi(\tau, x) - \widehat{\phi}(\widehat{\tau}, \widehat{x})|$ as in Section 5.3.2 yields for the local error in phase space an estimate by

$$|X(\tau)-\widehat{X}(\widehat{\tau})| \;\leq\; \mathrm{e}^{L\tau}(1+C_2)\big(1+\lambda_{min}^{-1}\,M\,U^{-\delta}\big)\,|\widehat{x}-x|+(1+C_2)(1+\lambda_{min}^{-1}\,M)\,\mathrm{err}(\widehat{t},h)\,,$$

cf. the local error estimate (5.3.13). Starting with $X(0) = \widehat{X}(0)$ and a recursive application of this last inequality, cf. Section 5.3.3, yields a global error estimate

$$\max_{n=1,\dots,N} |X(\tau_n) - \widehat{X}(\widehat{\tau}_n)| \;\leq\; (1 + C_2)(1 + \lambda_{min}^{-1}\,M)\,Z_N \sum_{k=1}^{N} Z_k^{-1}\,err(\widehat{\tau}_k - \widehat{\tau}_{k-1}, h)\,.$$

As these calculations are valid for almost all $\omega \in \Omega$ the convergence (5.2.1) follows and is of order p when $\mathrm{err}(t, h) = \mathcal{O}(h^p)$. The proofs for the limits (5.2.3) and (5.2.2) work completely analogous to those for Theorem 5.2.1, cf. Sections 5.3.3 and 5.3.4, respectively. □

5.5 Numerical examples

To illustrate the theoretical findings regarding the order of convergence we have implemented simulation methods for PDMPs based on continuous Runge-Kutta methods

$$\begin{array}{c|c} c & A \\ \hline & \beta^T \end{array}$$

(a)

$$\begin{array}{c|c} 0 & 0 \\ \hline & 1 \end{array}$$

$$b_1(\xi) = \xi$$

(b)

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & \frac{1}{2} & \frac{1}{2} \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}$$

$$b_1(\xi) = \tfrac{1}{2}\xi(2-\xi)$$

$$b_2(\xi) = \tfrac{1}{2}\xi^2$$

(c)

$$\begin{array}{c|cc} \frac{1}{3} & \frac{5}{12} & -\frac{1}{12} \\ 1 & \frac{3}{4} & \frac{1}{4} \\ \hline & \frac{3}{4} & \frac{1}{4} \end{array}$$

$$b_1(\xi) = \tfrac{3}{4}\xi(2-\xi)$$

$$b_2(\xi) = \tfrac{3}{4}\xi(\xi - \tfrac{2}{3})$$

(d)

$$\begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{5}{24} & \frac{1}{3} & -\frac{1}{24} \\ 1 & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \\ \hline & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \end{array}$$

$$b_1(\xi) = 2\xi(\tfrac{1}{3}\xi^2 - \tfrac{3}{4}\xi + \tfrac{1}{2})$$

$$b_2(\xi) = 4\xi^2(\tfrac{1}{2} - \tfrac{1}{3}\xi)$$

$$b_3(\xi) = 2\xi^2(\tfrac{1}{3}\xi - \tfrac{1}{4})$$

(e)

Figure 5.4: Coefficients and interpolation formulae of the continuous Runge-Kutta methods implemented for the numerical examples: (a) general Butcher tableau, (b) forward Euler method, (c) trapezoidal rule, (d) RadauIIa method, (e) LobattoIIIa method.

of different order. For an IVP

$$\dot{y} = g(t, y), \quad y(0) = y_0$$

with $y(t)$, $y_0 \in \mathbb{R}^d$ and $t \in [0, T]$ an *s-stage continuous Runge-Kutta method* is a discretisation scheme of the form

$$\widehat{y}((n+1)h) \;=\; \widehat{y}(nh) + h \sum_{i=1}^{s} \beta_i \, g(nh + c_i h, k_i), \quad n = 0, \ldots, N-1, \; h = T/N, \tag{5.5.1}$$

with stage values k_i, $i = 1 \ldots, s$,

$$k_i = \widehat{y}(nh) + h \sum_{j=1}^{s} a_{ij} \, g(nh + c_j h, k_j) \tag{5.5.2}$$

combined with an interpolation formula

$$\widehat{y}(nh + \xi h) = \widehat{y}(nh) + h \sum_{i=1}^{s} b_i(\xi) \, g(nh + c_i h, k_i), \quad 0 \leq \xi \leq 1 \tag{5.5.3}$$

for the approximation on the intervals between the discretisation points nh. The coefficients $\beta = (\beta_1, \ldots, \beta_s)^T$, $A = (a_{ij})_{i,j=1,\ldots,s}$ and $c = (c_1, \ldots, c_s)^T$ are given by Butcher tableaus as in Fig. 5.4 (a). In particular, we implemented the *explicit Euler method* (order 1), the *trapezoidal rule* (order 2), the 2-*stage RadauIIa method* (order 3) and the 3-*stage LobattoIIIA method* (order 4). The coefficients and interpolation polynomials $b_i(\xi)$ for these methods are given in Fig. 5.4 (b)–(e) taken from [11].
Each experiment consists of the same trajectory simulated using Algorithm A2 based on the different continuous Runge-Kutta methods with decreasing step sizes h. We compared the approximations to a reference solution as an exact solution is not available. The reference solution is an approximation of the same trajectory simulated with very high accuracy. We stopped decreasing the step size h for an implementation of Algorithm A2 when the approximations entered the error range of the reference solution.

We have applied the methods to a stochastic hybrid version of the Hodgkin-Huxley model for a patch of neuronal membrane, cf. Section 3.1.2. We consider the model for two different sets of parameters, cf. Appendix A. The first set is the original Hodgkin-Huxley model for the squid giant axon. The second set is taken from [93] wherein the authors experimentally compare the performance of various pseudo-exact algorithms with respect to certain test statistics.
The hybrid version of a space-clamped Hodgkin-Huxley neuron model is a 14–dimensional PDMP, cf. Section 3.1.2, where a one-dimensional continuous variable

$Y(t) \in \mathbb{R}$ models the difference in the electrical potential. The remaining, piecewise constant variables $\theta(t) \in \mathbb{R}^{13}$ record the states of the ion channels immersed in the membrane. In the Hodgkin-Huxley model there are two different families of ion channels: sodium (Na) and potassium (K) channels with 8 and 5 distinct states, respectively. Their kinetic schemes are given in Fig. 1.1. Each component of $\theta(t)$ thus counts the number of channels in the specific state, i.e., $\theta_1, \dots, \theta_8$ correspond to the states of the Na–channels and $\theta_9, \dots, \theta_{13}$ to the states of the K–channels. The characteristics of the PDMP are as follows. Firstly, the family of ODEs (5.1.1) defining the inter jump evolution of the PDMP is given by the equation

$$C\dot{y} = -\overline{g}_{\mathrm{Na}}\theta_8(y - E_{\mathrm{Na}}) - \overline{g}_{\mathrm{K}}\theta_{13}(y - E_{\mathrm{K}}) - \overline{g}_{\mathrm{L}}(y - E_{\mathrm{L}}) + I(t)\,. \tag{5.5.4}$$

Secondly, the instantaneous jump rate λ is given by

$$\begin{aligned}\lambda((y,\theta)) \quad &= \begin{pmatrix} a_m(y) \\ b_m(y) \\ a_h(y) \\ b_h(y) \end{pmatrix}^T \begin{pmatrix} 3 & 2 & 1 & 0 & 3 & 2 & 1 & 0 \\ 0 & 1 & 2 & 3 & 0 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_8 \end{pmatrix} \\ &\qquad + \begin{pmatrix} a_n(y) \\ b_n(y) \end{pmatrix}^T \begin{pmatrix} 4 & 3 & 2 & 1 & 0 \\ 0 & 1 & 2 & 3 & 4 \end{pmatrix} \begin{pmatrix} \theta_9 \\ \vdots \\ \theta_{13} \end{pmatrix}.\end{aligned}$$

Thirdly, we specify the transition measure μ. The point probability of the event that one channel changes from state i to state j is given by the transition rate of state i to state j times the number of channels in state i divided by the total instantaneous rate λ. For example, the probability of the event of one channel changing from state 1 to 2 – conditional on the jump time being t – is given by

$$\frac{3\,a_m(Y(t))\,\theta_1(t-)}{\lambda((Y(t),\theta(t-)))} =: \mu((Y(t),\theta(t-)),\{\theta(t-) + (-1,1,0,\dots,0)^T\})\,.$$

All other events have zero probability, that is two channels do not change states simultaneously almost surely.

We remark that (5.5.4) differs from the general form (5.1.1) due to the added time dependent function $I(t)$ which denotes the external current input to the system. However, for the numerical experiments we present the input is piecewise constant of the form $I(t) = const. \cdot \mathbb{I}_{(t_1,t_2]}(t)$ (*monophasic input*) and hence one may think of the model as PDMPs 'glued' together at the times t_1, t_2 with the final state being the initial condition of the next. The reason for this type of input is that in numerical simulations of the neuron model one wants to sample first an initial condition from

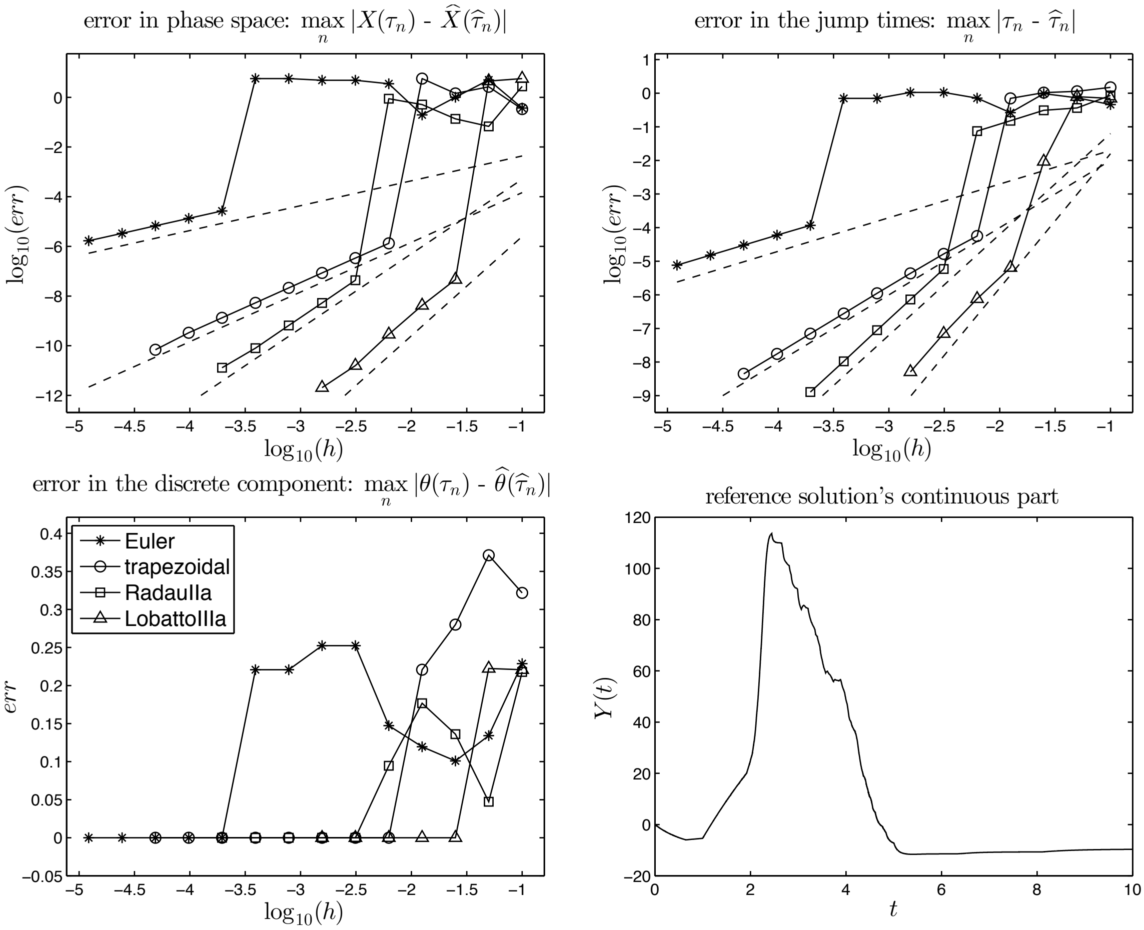


Figure 5.5: A first exemplary numerical experiment for a trajectory to the first parameter set. We have plotted the errors in phase space (top left), jump times (top right) and in the discrete component (bottom left). The path of the reference solution's continuous component calculated using MATLAB®'s `ode45` is shown in the bottom right panel.

the membrane at rest, i.e., with $I(t) \equiv 0$, for the start of the response to an input at time t_1.

We present two sample trajectories for the first parameter set. The reference solution for the trajectory in Fig. 5.5 was calculated using MATLAB®'s `ode45` implementation with `AbsTol` / `RelTol` $= 2.22045\text{e}^{-14}$. We used the built-in event detection for the calculation of the hitting times. For the trajectory in Fig. 5.6 the reference solution was calculated using the LobattoIIIa method with step size $h = 5\text{e}^{-6}$. On the one hand, the use of LobattoIIIA illustrates that high order methods yield good results in the absolute error range for reasonable equidistant step sizes. On the other hand, the use of `ode45` which employs automated step-size selection illustrates that methods employing automated step size selection yield good accuracy results and thus can in principle be used. However, we found that for the same level of accuracy simulation times for the equidistant LobattoIIIA method were considerably shorter than for `ode45`. Thus this equidistant implementation performs better than the simulation

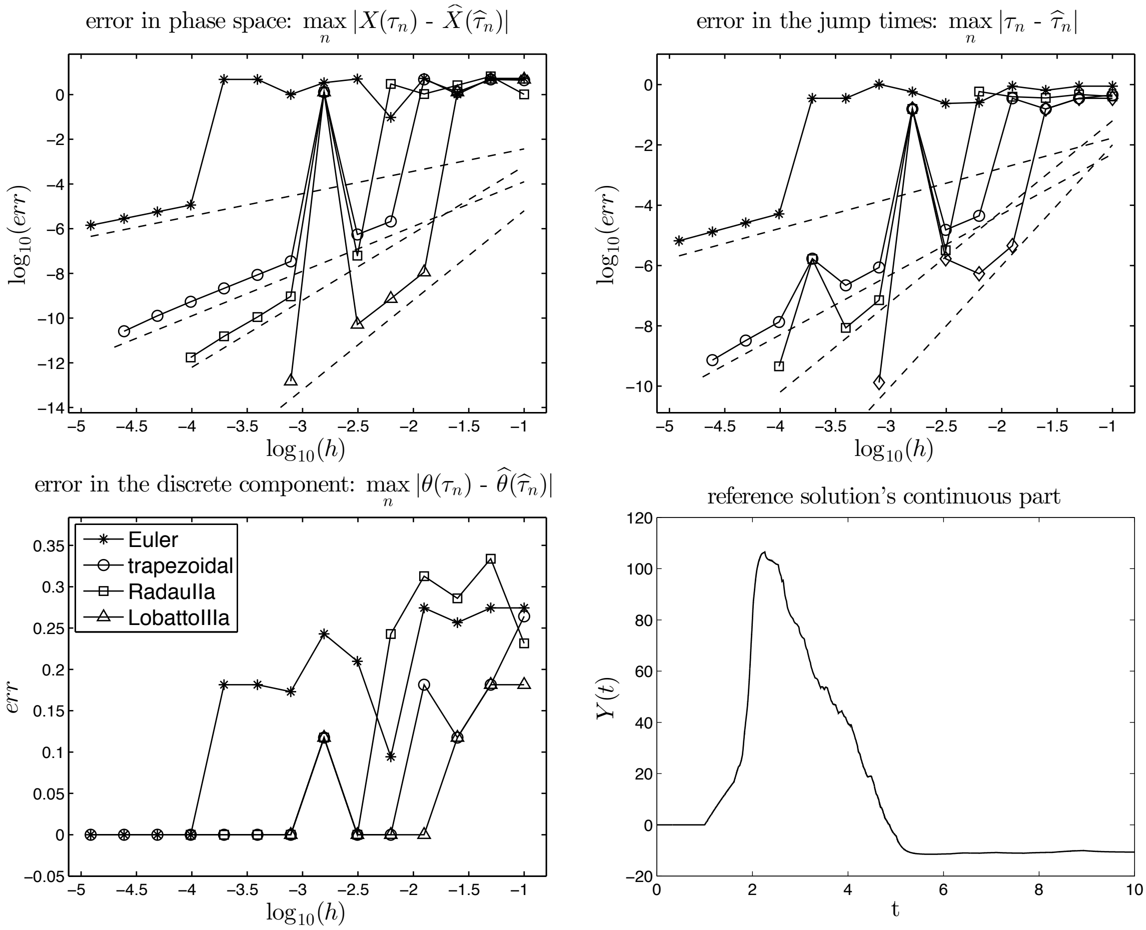


Figure 5.6: A second exemplary numerical experiment for a trajectory to the first parameter set. We have plotted the errors in phase space (top left), jump times (top right) and in the discrete component (bottom left). The path of the reference solution's continuous component calculated using the presented PDMP method based on the continuous LobattoIIIA is shown in the bottom right panel.

using MATLAB®'s time stepping algorithm contrary to the latter being specifically developed to speed up simulations. We note that these findings are not artefacts of the two specific paths presented but are persistent throughout all the simulations we conducted.

The error plots in the upper panels of Fig. 5.5, i.e., the error in phase space (5.2.1) (upper left panel) and the error in the jump times (5.2.3) (upper right panel) illustrate very clearly the theoretical result in Theorem 5.2.1: For step size h small enough such that the global error in the discrete component vanishes (lower left panel) the theoretical asymptotic order of convergence for the different methods is observed. For a guide of the eye we added lines of slope 1, 2, 3 and 4 beneath the errors of the methods with the respective orders. The plots of the second trajectory shown in Fig. 5.6 are a slightly less clear illustration of the asymptotic order of convergence. Here, the discrete error of the higher order methods has already vanished but reappears for a certain step size h. However, if we remove this outlier we observe the asymptotic order underlying the approximation as seen by comparison with the straight lines, that is, the convergence is non-monotonic.

As mentioned discussing the reference solution a naive use of automated step-size control with the intention of speeding up simulations and controlling the error can be misleading. However, it would be a task for an efficient and practicable pathwise step size detection and error control to detect such 'bad' step sizes as occur in the second trajectory and either avoid them or minimise their effect. These are crucial points for the implementation and performance of the algorithms. They demand a further thorough investigation in this direction, e.g., an analysis of the shortcomings of the step-size detection as employed in `ode45`, which we have not attempted to do in this study.

Finally, in a second example we consider a variant of the Hodgkin-Huxley model only dealing with currents due to Na–channels. This model reduces to 9 dimensions as $\theta_9(t), \ldots, \theta_{13}(t) \equiv 0$. The channel density is higher and the current per channel smaller as for the first parameter set which overall renders the trajectories 'less noisy', cf. the sample trajectory of the reference solution in Figs. 5.5 and 5.6 with the one in Fig. 5.7. The reference solution was again calculated using the LobattoIIIA method with step size $h = 5\mathrm{e}^{-6}$. Overall we find the same behaviour as in the first example. For small enough step sizes such that the errors in the discontinuous components vanish, the predicted order of the asymptotic error behaviour is observed. However, note that even for the smallest step size considered the Euler method lacks the accuracy to approximate the trajectory correctly. Its error is orders of magnitude larger than the error of higher order Runge-Kutta method with large step size, which strongly supports the use of higher order methods.

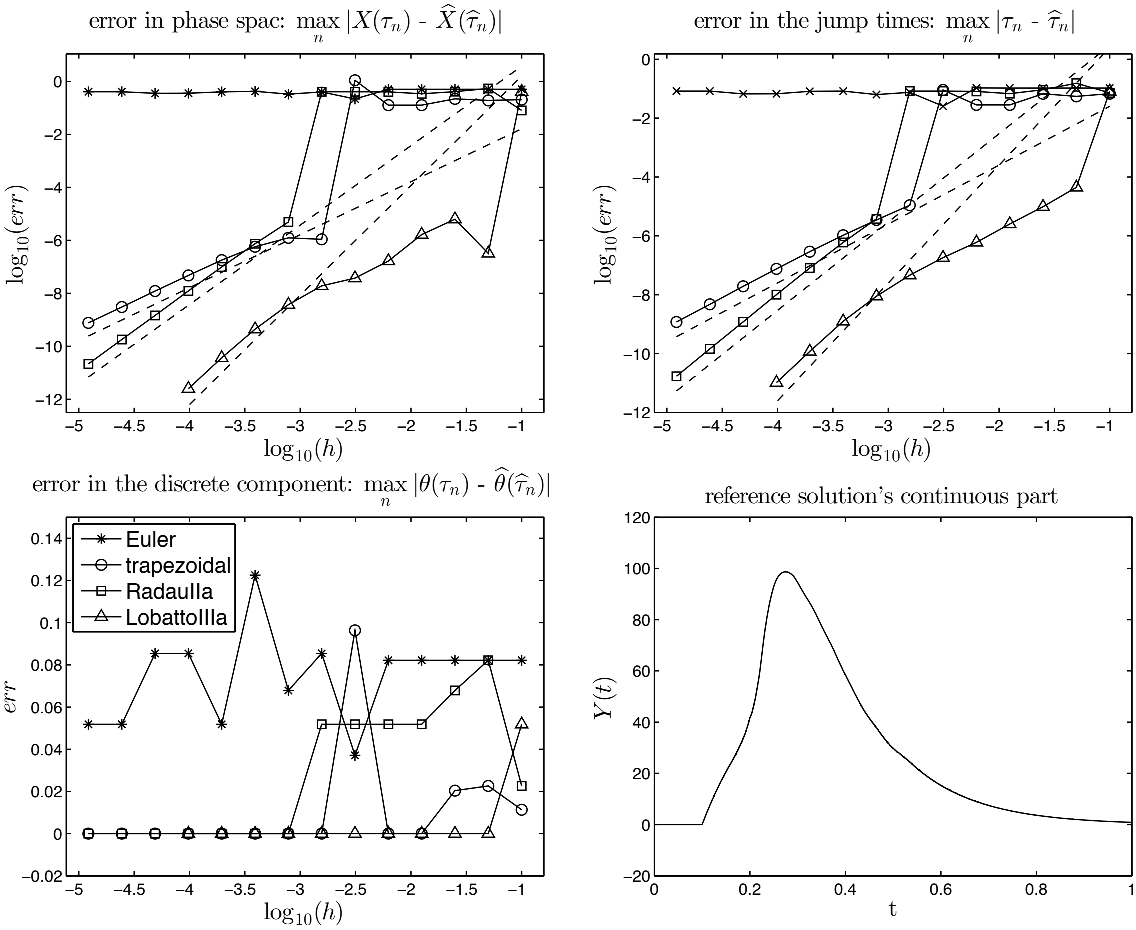


Figure 5.7: An exemplary numerical experiment for a trajectory to the second parameter set, cf. Appendix A. We have plotted the errors in phase space (top left), jump times (top right) and in the discrete component (bottom left). The path of the reference solution's continuous component calculated using the presented PDMP method based on the continuous LobattoIIIA is shown in the bottom right panel.

Chapter 6

Conclusion and further research

To conclude the thesis we briefly recapitulate the main results we have presented and comment on further research. As we have stated in the introduction the aims of the thesis are twofold. On the one hand, we are interested in a general advance of Piecewise Deterministic Markov Process theory and, on the other hand, the application of these new results to models of biological excitable membranes under the influence of channel noise. Broadly speaking we may identify a range of three topics within this twofold aims: firstly, deriving a mathematical framework for the stochastic hybrid modelling of spatio-temporal dynamics of excitable membranes and providing analytical tools for their analysis; secondly, as a reduction of the complexity of the models, deriving approximations to PDMPs by stochastic processes more accessible to an analysis; thirdly, the derivation and analysis of numerical simulation methods for PDMPs. The latter two aspects are tools that allow the study of more complex PMDP models which are hardly accessible to a theoretical analysis.

In the thesis we have achieved these aims step by step in the progression of the chapters. In Chapter 2 we presented the general class of Piecewise Deterministic Markov Processes. This class of processes allows a precise mathematical description of the internal noise structure of excitable membranes, with and without spatial dynamics, by hybrid stochastic models. As theoretical results we obtained the strong Markov property, the extended generator of these processes and an Itô-formula. Moreover, also an important contribution to general PDMP theory, we have shown how it is possible to cast spatio-temporal evolution into this framework. We believe that this approach is widely applicable to model systems where stochastic discrete events of individual, spatially distributed objects are coupled via the solution of partial differential equations and these two components mutually influence each other. Further, PDMPs are also closely related to *multi-scale modelling* on which we comment in Section 6.1 below. We have illustrated this general modelling framework for the example of the classical Hodgkin-Huxley model of a squid giant axon in Chapter 3. Generalisations of this modelling approach have already been discussed in Section 3.1.1. Further, it also immediately extends to a large class of biophysical models of excitable membranes, e.g., cardiac cells or models for calcium dynamics, which possess the same underlying physiological structure as neuronal membranes. The formulation of excitable membrane models by PDMPs allows a rigorous analytical and numerical investigation of these models based on the theory of PDMPs. As first general steps in these directions it enabled us

- to develop analytical approximations of reduced complexity by continuous stochastic processes such as solutions of SPDEs motivated by the approaches in [50, 96] in the finite-dimensional setting;
- to develop and analyse hybrid algorithms for finite-dimensional PDMP models.

We pursued these two aspects within the general PDMP theory in the subsequent two chapters of the thesis. First, as a general theoretical tool for PDMPs we derived a law of large numbers and martingale central limit theorems in Chapter 4. The former establishes a connection of stochastic hybrid models to deterministic models given, e.g., by systems of partial differential equations. Whereas the latter connects the stochastic fluctuations in the hybrid models to diffusion processes. As a prerequisite to these limit theorems we carried out a thorough discussion of Hilbert space valued martingales associated to the PDMPs. Furthermore, these limit theorems provide the basis for a general Langevin approximation to PDMPs, i.e., certain stochastic partial differential equations that are expected to be similar in their dynamics to PDMPs. We have applied these results to compartmental-type models of spatially extended neuronal membranes. Ultimately this yields a system of SPDEs which models the internal noise of a biological excitable membrane based on a theoretical derivation from exact stochastic hybrid models. Finally, in Chapter 5 we addressed the question of numerical simulation of PDMPs. We presented and analysed the convergence in the pathwise sense of a class of simulation methods for PDMPs in Euclidean space. We have illustrated our theoretical findings with numerical examples being stochastic hybrid models of space-clamped membranes. We are able to observe the proved order of convergence. Further, the methods produced sample trajectories that possess the qualitative behaviour expected of excitable membranes. The numerical examples also allowed us to draw conclusions on the comparative performance of different methods.

Finally, we now comment on further directions of research concerning the three aspects identified above building on the results derived in the present thesis.

6.1 Stochastic hybrid models of excitable media

We have repeatedly emphasised that PDMP models are the correct mathematical description of internal (channel) noise. However, a second important noise source

in biological systems, and excitable membranes in particular, is *external noise*. A discussion and modelling of external noise is not within the scope of this thesis and it is one starting point for further research and extension of the PDMP models which we briefly explain.
External influences to neuronal membranes directly affect the transmembrane potential and arise from various sources. In the case of deterministic external influences, which also includes, e.g., an externally applied macroscopic current at the membrane, the resulting model formulation readily fits in the presented PDMP framework, cf. Section 3.1.1. However, this is not what is usually understood as external *noise*. Normally, external noise is a fast fluctuating (and usually small) time-dependent input. It may arise as natural fluctuations in the transmembrane potential due to thermal noise or as small inputs into the membrane resulting from action potentials emitted by other neurons. This latter type of external noise is also called *shot noise*. Most importantly, external noise is generally modelled stochastically using white noise. Formally adding a white noise term to the membrane balance equation (1.2.2) or space-time white noise to the cable equation (1.2.1) we obtain equations which are, in a mathematical precise way, accurately treated within the framework of stochastic ordinary / partial differential equations (SDEs/SPDEs) driven by Wiener or Poisson noise. Coupling continuous stochastic dynamics given by SDEs to discontinuous jump dynamics analogously to the coupling in the case of PDMPs results in stochastic processes known as General Stochastic Hybrid Systems (GSHS) [28]. To the best of our knowledge coupling SPDE solutions to discontinuous dynamics, and thus extending the GSHS framework to include spatial dynamics, has not been considered in the literature so far. Further, extensive fundamental analytic studies of GSHS comparable to the monographs [37, 64] for PDMPs are not available. Therefore, in order to obtain a rigorous theory for stochastic hybrid membrane models incorporating spatial dynamics and external noise, first a rigorous analysis of the resulting processes has to be conducted and their well-definedness established.

We finish this section with a brief comment on the general idea of *multi-scale modelling* and its close connection to PDMPs and GSHS. This connection is particularly important for models of chemical reaction networks and the subsequent comments apply for spatially homogeneous chemical reaction systems. The multi-scale nature of such systems is already reflected in the different approaches to model dynamics in chemical reaction system using

(i) continuous time Markov chains, viz. the Stochastic Simulation Algorithm,

(ii) stochastic differential equations, i.e., the Langevin approximation to the Markov chain models,

(iii) or systems of ordinary differential equations, termed reaction rate equations in this application.

That is, the same system may be appropriately modelled by either of these approaches with respect to the aim of the later analysis. The appropriate choice depends mainly on the system size, i.e., number of reacting molecules, and / or the time horizon, i.e., whether we are interested in short-term or long-term behaviour. However, different scales may also be present in one model. That is given the system size and time horizon certain reactions may be most appropriately modelled deterministically and others by Markov chains. Whenever these two types of reactions affect the same molecules then an approximation of the fast reactions in the system by reaction rate equations results in a stochastic process which has the characteristic structure of a finite-dimensional PDMP.[1] By this approximation procedure a substantial model reduction in terms of complexity of the model (structural complexity as well as numerical complexity in attempts to simulate the stochastic model) can be achieved. In a recent study these approximations have been mathematically rigorously justified for spatially homogeneous chemical reaction systems in terms of weak limit theorems [34]. This result is complementary to the limit theorems derived in [79, 96]: Kurtz [79] presents limit theorems for jump processes to solution of ODEs, the authors in [34] present limit theorems of jump processes to PDMPs and limit theorems for PDMPs to solutions of ODEs are presented in [96].
A second possible method of model reduction is using the Langevin description to approximate the fast reactions instead of the reaction rate equation. The resulting stochastic process now follows the dynamics of an SDE system in between jumps of the slow reactions. Furthermore the rates of the slow reactions may depend on the paths of the SDEs. Such a method has been proposed in [107]. The mathematically correct treatment of these processes is within the framework of GSHS. This connection seems to have been unnoticed in the chemical reaction research community. To the best of our knowledge a theoretical analysis of these models and the approximation procedure is completely missing as is a numerical analysis of the simulation algorithm proposed in [107].

We believe one central question in the context of multi-scale approximations is a qualitative analysis of the long-time behaviour of the different modelling approaches and their comparison. For the individual model classes (i)–(iii) above, tools for the analysis of the long-time behaviour are widely available, e.g., classical stability analysis or

[1]There is a close conceptional connection between these multi-scale approximation to chemical reaction systems and hybrid models of excitable membranes. First of all the channel dynamics in models of excitable media are given in a kinetical description resembling chemical reactions. Here each 'reaction' corresponds to a possible change of channel states. Secondly, the flow of individual ions across the membranes occurs on a much faster time scale than the channel switchings and the membrane balance equation can be considered as a reaction rate approximation. Here ions can be considered as 'reacting' where the different states are 'being inside' and 'being outside' of the membrane.

dynamical systems theory for ODE systems, cf., e.g., [73, 101], or analysis of the invariant distributions of Markov chains and their ergodicity, or stability analysis for SDEs, e.g., mean-square linear stability when the system allows for a deterministic fixed point, cf. [59], or otherwise methods based on the theory Random Dynamical Systems, cf. a discussion by E. Buckwar, P. Kloeden and the present author in the context of SDEs with additive noise in [25]. This type of SDEs, for example, arises as a linear noise approximation to Markov chain models via the van Kampen system size expansion [118]. In the case of hybrid system, i.e., combining Markovian jump dynamics and deterministic evolution, either very few results on the long time behaviour are available, see [33, 40, 49] in the case of PMDPs, or are in the case of GSHSs completely missing. We believe the development of such tools is of major importance and should be a central aspects of future research effort.

6.2 Limit theorems

P. Bressloff characterises in [23] spatial models arising in neural systems in the following way: "*One of the distinct features of neural systems compared to chemical systems is that when spatial degrees of freedom are taken into account, the former involves non-local interactions rather than diffusive interactions.*" If we consider this characterisation of neural system as the defining feature of certain spatial systems we conjecture that the limit theorems presented in Chapter 4 find broad application for this type of models. Examples are PDMPs used for models of cell adhesion[2] and, in particular, stochastic neural field theory. Recently P. Bressloff provided an appropriate kinetic framework for the study of stochastic neural field equations and showed using the van Kampen system size expansion [118] and the continuum limit that the Wilson-Cowan equation is the corresponding deterministic limit [22]. Even more recently the same author presented in [23] a diffusion approximation via a functional Fokker-Planck equation also derived via the van Kampen system size expansion. In forthcoming work we apply the limit theorems derived in Chapter 4 to stochastic neural field equations, i.e., the deterministic limits are the Wilson-Cowan or the Amari equation, respectively. These theorems provide a more rigorous connection than the methods employed in [22, 23] and characterise the appropriate type of limit, i.e., limit in probability and limit in the mean in the law of large numbers and a limit in distribution for the diffusion approximation. With some additional work it is in this context possible to obtain estimates on the speed of convergence which is in general not possible via the van Kampen system size expansion.

Another topic for further research is motivated by corresponding results in finite-dimensions [80, 96] and for spatially inhomogeneous chemical reaction systems con-

[2]Personal communication with O. Bagdasar, University of Nottingham.

verging to reaction diffusion equations, cf. [75]. In these studies limit theorems are derived for the fluctuations around the deterministic limit identified by the law of large numbers. This corresponds to the linear noise approximation derived via the van Kampen system size expansion. Using the notation of Chapter 4 we conjecture that the sequence of processes, $\big(\sqrt{\alpha_n}\,(U_t^n - u(t), z^n(\theta_t^n) - p(t)\big)_{t\geq 0}$, $n \in \mathbb{N}$, converges in distribution to a suitable diffusion process. Moreover, we further conjecture that this limit is closely related to the asymptotic linearisation of the Langevin approximation around the solution of the deterministic limit, cf. [96] wherein this result is proven for finite-dimensional PDMPs.

Further, we believe that the Langevin approximation to spatio-temporal PDMP models of excitable membranes also poses an important object for further investigation. Its derivation was the initial motivation of the study of the limit theorems in Chapter 4 and it is the main application of these in the present thesis which enables to write down the system of SPDEs that constitute a Langevin approximation. This system now demands for further analysis, particularly, first of all the question of existence and uniqueness of the Langevin approximation has to be addressed. Subsequently, as SPDEs are analytically better accessible than jump processes a theoretical analysis of qualitative and quantitative properties of the models may be possible. A large deviation result may be particularly of interest to study analytically the propagation reliability and failure along an axon in addition to numerical investigations for compartmental-type hybrid models conducted in [45]. For finite-dimensional PDMPs a connection between Large Deviations of the PDMPs to the Langevin approximation has been considered in terms of a functional central limit theorem in [96] and a comparison of Large Deviations for birth-death processes and their diffusion approximation was carried out in [95]. It may be possible in future work to extend these results to PDMPs in infinite dimensions and their Langevin approximation by SPDEs.

Finally, we have not yet investigated how far the presented limit theorems in Chapter 4 also generalise the previously known results for stochastic spatio-temporal models of chemical reactions converging to reaction-diffusion equations, cf. the series of results on variations of the model in [75, 76, 77, 78] and [15, 16, 17, 18, 19]. This question necessitates a close inspection of the techniques employed in these studies and identify the analogons in our approach. An answer to this question would contribute to a more complete picture of limit-theorems for spatio-temporal stochastic models.

6.3 Numerical simulations

Regarding the numerical approximation of PDMPs future work may be pursued in various directions. We have already discussed one point in Section 5.5 regarding the implementations of the methods: From a practical point of view an important aspect is analysing step-size adaptation methods in connection with the random hitting time problem which needs to be solved at each step. This might yield an improvement in the efficiency of the methods and may avoid certain step-sizes producing a disproportional large error. Apart from this question we think of two further directions that may yield fruitful results.

Firstly, in contrast to deterministic numerical analysis there are different concepts to measure the error of a numerical approximation in stochastic numerical analysis. One of these concepts is *pathwise convergence* which we have analysed in Chapter 5. Different concepts are *convergence in the mean-square* and *weak convergence*, cf. [72, 92]. These concepts emphasise different aspects of the stochastic theory and the appropriateness of each depends on the aims the numerical simulations are used to achieve. Weak convergence, for example, is the appropriate convergence concept when distributional parameters of models are estimated via Monte-Carlo methods. In particular, in connection with multi-scale approximations to chemical reaction networks one is often interested in the evolution of the distribution of molecule numbers over time which is investigated using Monte-Carlo simulations. Therefore a task for further numerical analysis is to extend the presented pathwise convergence result to these other convergence concepts. We note that within the three aforementioned concepts the pathwise error is in some sense the strongest concept. We conjecture that our result can be used to establish mean-square and weak-convergence for the numerical methods we presented.

Secondly, we have considered convergence only for PDMPs taking values in Euclidean space. Thus a natural task is to develop numerical methods for the simulation of the spatio-temporal membrane models we discussed in the thesis. This amounts to consider numerical methods that in between jump times of the PDMP solve a partial differential equation instead of an ordinary differential equation. The development and analysis of such methods was not within the scope of this thesis, however, there is a connection to the methods we considered. We briefly comment on simulating spatio-temporal PDMPs.
For the numerical solution of PDEs, considered as abstract evolution equations, the infinite-dimensional state space of these necessarily has to be discretised to obtain an approximation on a finite-dimensional subspace which can be implemented in a computer. There exist various approaches for this task, e.g., finite-element methods or, simpler, finite-differences. For axon models where the PDE is the cable equation the latter is usually an appropriate choice. Then spatially discretising the cable equation using finite-differences we obtain a system of ODEs which, connected to the jump dynamics of the channels, defines a finite-dimensional PDMP. Now, this PDMP can be simulated using the methods we presented in Chapter 5.
An important question for axon models which can be investigated via numerical stud-

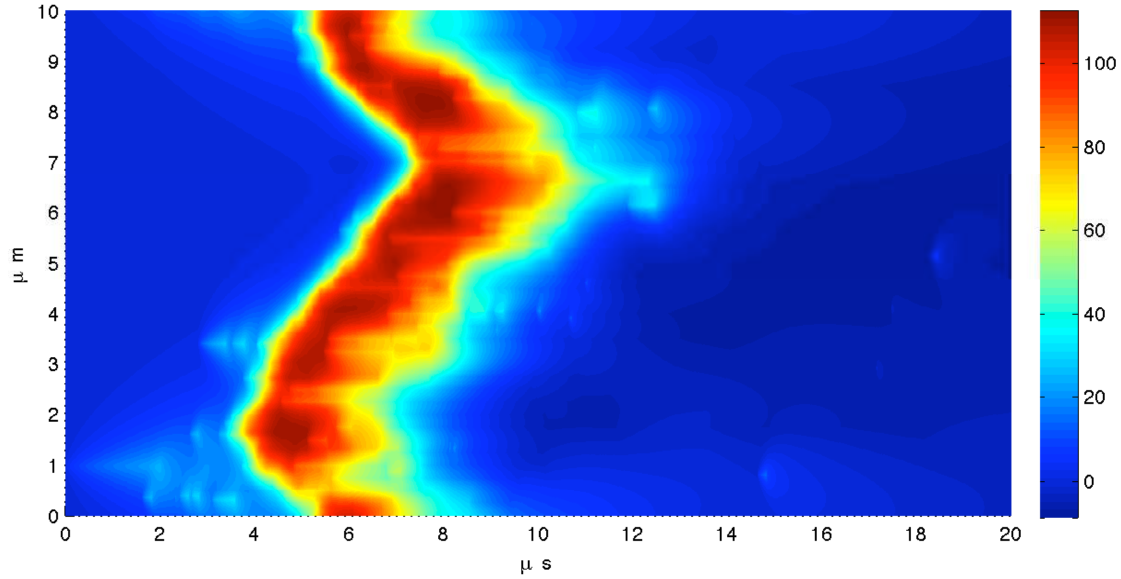


Figure 6.1: The spatial discretisation has step-size $\Delta x = 10/639$ and for solving the resulting finite-dimensional PDMP we used MATLAB®'s `ode45` solver. A channel is placed at each discretisation point with one K-channel for every 15 Na-channels. The parameters are chosen as in the classical Hodgkin-Huxley models, cf. Appendix A.

ies of spatio-temporal hybrid models of excitable membranes is, for example, the propagation speed of the action potential and propagation reliability. The latter has been studied in [45] using implementations that correspond to discretisations of the compartmental-type models presented in Section 3.2. However, such numerical studies are also possible using the exact hybrid models discussed in Section 3.1. In Fig. 6.1 we present one sample trajectory of the hybrid Hodgkin-Huxley model obtained by the aforementioned numerical method. That is, the cable equation equipped with periodic boundary conditions is spatially discretised using finite differences on an equidistant grid such that the gridpoints coincide with the channel locations. The trajectory in Fig. 6.1 illustrates the feasibility of this approach. Note that here the channel density does not reflect physiological reality nor does the current input strength due to one channel. However, in current research efforts[3] these methods are employed to experimentally investigate the wave speed of travelling waves in hybrid models where the model complexity is reduced via an averaging technique. For these simplified models simulation speeds for biophysical realistic parameters are acceptable for Monte Carlo Simulations. Furthermore, the example trajectory in Fig. 6.1 suggests that the hybrid stochastic models correctly reproduce the expected qualitative behaviour of action potentials: Action potentials move away from the point of excitation, which in Fig. 6.1 is at approximately $x = 2$ and $t = 4$, in the form of a standing wave, i.e., two complementary wave fronts following each other with the same speed. Moreover, we

[3] Personal communication with M. Thieullen and A. Genadot, Université Pierre at Marie Curie.

also observe the feature that two action potentials annihilate on collision.

To conclude, we conjecture that a convergence result can also be derived for numerical approximation of spatio-temporal PDMP models of excitable membranes. We think that a first natural step is an analysis of the method that arises discretising the cable equations using finite differences and numerically solving the resulting finite-dimensional PDMPs with methods derived in Chapter 5.

Appendix A

Rate functions and parameter values

The following are the rate functions and coefficients of the cable equation for the original Hodgkin-Huxley model of the squid giant axon taken from [74]. The rate functions are

$$a_n(u) = \frac{10-u}{100(\mathrm{e}^{(10-u)/10}-1)}, \quad a_m(u) = \frac{25-u}{10(\mathrm{e}^{(25-u)/10}-1)}, \quad a_h(u) = 0.07\mathrm{e}^{-u/20},$$

$$b_n(u) = 0.125\mathrm{e}^{-u/80}, \qquad b_m(u) = 4\mathrm{e}^{-u/18}, \qquad b_h(u) = \frac{1}{\mathrm{e}^{(30-u)/10}+1},$$

and the constants used are

$$\begin{aligned}
&E_{\mathrm{Na}} = 115 \text{ mV}, && \overline{g}_{\mathrm{Na}} = 4 \text{ pS}, && \eta_{\mathrm{Na}} = 300\ \mu\mathrm{m}^{-2},\\
&E_{\mathrm{K}} = -12 \text{ mV}, && \overline{g}_{\mathrm{K}} = 18 \text{ pS}, && \eta_{\mathrm{K}} = 20\ \mu\mathrm{m}^{-2},\\
&E_{\mathrm{L}} = 10.613 \text{ mV}, && \overline{g}_{\mathrm{L}} = 0.3 \text{ mS/cm}^2, && \\
&C = 1\ \mu\mathrm{F/cm}^2, && R = 35.4\ \Omega\cdot\mathrm{cm}, && 2r = 0.476 \text{ mm},
\end{aligned}$$

where η_{Na} and η_K denote the Na and K channel density, respectively. Note that the equilibrium potential for the leakage current E_{L} is chosen such that in the deterministic set of equations the membrane potential has a stable equilibrium at the origin. Usually the conductances are given independent of the channel density i.e., $\overline{g}_{\mathrm{Na}}\eta_{\mathrm{Na}} = 120\,\mathrm{mS/cm}^2$ and $\overline{w}_{\mathrm{K}}\eta_{\mathrm{K}} = 36\,\mathrm{mS/cm}^2$ which are the values used in the deterministic Hodgkin-Huxley model. Further, C is the membrane conductance, R the intracellular resistivity and $2r$ the axon diameter. Finally, input current I in the right hand side of membrane equations is in units of $\mu\mathrm{A/cm}^2$. The input that is considered for the numerical example in Section 5.5 is a monophasic current starting at $t = 1$ ms and lasting for 1 ms of strength 30 pA, i.e., $I(t) = 30 \cdot \mathbb{I}_{(1,2]}(t)$.

The rate functions and parameters for the neuron model taken from [93] are given by

$$a_m(u) = \frac{1.872(25.41 - u)}{\mathrm{e}^{(25.41-u)/6.06} - 1}, \qquad a_h(u) = \frac{-0.549(27.74 + u)}{1 - \mathrm{e}^{(u+27.74)/9.06}},$$

$$b_m(u) = \frac{3.973(21.001 - u)}{1 - \mathrm{e}^{(u-21.001)/9.41}}, \qquad b_h(u) = \frac{22.57}{1 + \mathrm{e}^{(56-u)/12.5}}\,.$$

and the constants used are

$$E_{\mathrm{Na}} = 144 \text{ mV}, \qquad \overline{g}_{\mathrm{Na}} = 2.569 \cdot 10^{-5} \text{ pS}, \qquad \eta_{\mathrm{Na}} = 1000\ \mu\mathrm{m}^{-2},$$

$$E_{\mathrm{K}} = 0 \text{ mV}, \qquad \overline{g}_{\mathrm{K}} = 0 \text{ pS}, \qquad \eta_{\mathrm{K}} = 0\ \mu\mathrm{m}^{-2},$$

$$E_{\mathrm{L}} = 0 \text{ mV}, \qquad g_{\mathrm{L}} = 1/(1953.49 \cdot 10^{3}) \text{ mS/cm}^2,$$

$$C = 0.0714 \cdot 10^{-6}\ \mu\mathrm{F/cm}^2\,.$$

Note that the numerical example in Section 5.5 considers Na–channels only. As The current input is a monophasic current starting at $t = 0.1$ ms and lasting for 0.1 ms of strength $35.1 \cdot 10^{-6}$ pA, i.e., $I(t) = 35.1 \cdot 10^{-6} \cdot \mathbb{I}_{(0.1,0.2]}(t)$. These are the parameters for the neuron considered in [93]. We note that the input current strength for the experiments in [93] is incorrectly reported and corrected in [24].

Appendix B

Random Counting Measures

In this part of the appendix we introduce some background on topics in probability theory which is necessary for the definition and analysis of PDMPs. The main purpose is to briefly recall the definitions and the most important properties used in the thesis and, especially, to harmonise the definitions in references on PDMPs [37, 64]. In Section B.1 we first briefly introduce the general class of Borel spaces which serve as state spaces for all stochastic processes considered in this thesis. The discussion of Borel spaces is to a large extent based on material presented in [13]. Then in Section B.2 we present the definition of random counting measures and collect results mainly from the monograph [64]. Random counting measures and their properties are the basis for the definition of PDMPs and a derivation of the most central results in PDMP theory in Chapter 2. Finally, in Section B.3 we introduce survivor functions which are essential to defining the distributions of inter-jump times of PDMPs.

B.1 Preliminaries

We recall that the Borel-σ-field of a topological space is the σ-field generated by the topology. In particular, if a separable metric space is equipped with the induced topology, i.e., the topology the base of which are the open balls with respect to the metric, the Borel-σ-field coincides with the σ-field generated by the open balls. We note, that trivially a metric space equipped with its induced topology is a metrisable space[1]. In this thesis, metric spaces are always normed spaces where the metric is induced by the norm.

Definition B.1.1. ([13, Sect. 7.3]) A topological space is called a *Borel space* if it is homeomorphic to a Borel subset of a complete, separable, metric space.

Every Borel space is separable and metrisable and every complete, separable, metrisable space is a Borel space. Further, every Borel subset of a Borel space is a Borel

[1]A *metrisable space* is a topological space X such that there exists a metric on X and the open balls of this metric are a base for the topology, cf. [13, Chap. 7].

space. Thus, in particular, every separable Banach space and a Borel subset thereof is a Borel space. A *Borel measurable space* is a Borel space equipped with its Borel-σ-field.

Definition B.1.1 for a Borel space is used in [37], however in [64] a Borel space is defined as a measurable space $(E, \mathcal{E})$ such that a bimeasurable bijection onto a Borel subset of $\mathbb{R}$ exists. Obviously, these definitions are closely related and it is easily shown that Definition B.1.1 implies the definition employed in [64]: Note that due to Urysohn's Theorem every Borel space is homeomorphic to a subset of the Hilbert cube and thus the cardinality of a Borel space is less than or equal to uncountable. Further, due to [13, Prop. 7.15] two Borel spaces are Borel-isomorphic if and only if they have the same cardinality. For uncountable Borel spaces this implies that there exists a bimeasurable bijection onto the unit interval $[0, 1]$, cf. [13, Prop. 7.16.1]. Obviously, if the Borel space is countable or finite it is Borel-isomorphic to a finite or countable subset of $[0, 1]$. As $[0, 1]$ as well as any finite or countable subsets thereof are Borel subsets of $\mathbb{R}$ it follows that for every Borel space in the sense of Definition B.1.1 there exists a bimeasurable bijection onto a Borel subset of $\mathbb{R}$.

Borel spaces have the following properties where (a)–(c) are reported in [64, p. 19] and (d) and (e) can be found in [13, Sect. 7.3].

Proposition B.1.1. *Let $(E, \mathcal{E})$ be a Borel measurable space.*

(a) *All singletons are measureable, i.e., $\{y\} \in \mathcal{E}$ for all $y \in E$. This implies that $\mathcal{E}$ separates points, i.e., for any $y \neq x$ there exists $A \in \mathcal{E}$ such that $y \in A$ and $x \in A^c$.*

(b) *The σ-field $\mathcal{E}$ is separable, i.e., $\mathcal{E}$ is countably generated.*

(c) *If X is an $(E, \mathcal{E})$-valued random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{F}' \subset \mathcal{F}$ is a sub-σ-field, then there always exists a regular conditional distribution of X given $\mathcal{F}'$.*

(d) *A finite or countable product of Borel spaces is a Borel space when equipped with the product topology. The Borel-σ-field generated from the product topology coincides with product-σ-field[2].*

(e) *Finally, finite and countable vectors of Borel-measurable functions are Borel-measurable and each component of a Borel-measurable vector function is Borel-measurable.*

Examples for Borel measurable spaces are $\mathbb{R}^d$, $d \geq 1$, with the usual Borel σ-field, the sequence space $\mathbb{R}^{\mathbb{N}}$ with the σ-field generated by the projections on the components

[2] The product topology is the coarsest topology such that all projections are continuous and the product σ-field is the smallest σ-field such that all projections are measurable.

and the space of $\mathbb{R}^d$-valued càdlàg processes with the σ-field generated by the coordinate projections [64]. Finally, also $\mathbb{R} \cup \{-\infty, \infty\}$ with the σ-field generated by the weakest topology containing the sets $[-\infty, a), (a, \infty], (a, b)$ for all $a, b \in \mathbb{R}$ is a Borel space [13]. In general, if $(E, \mathcal{E})$ is a Borel measurable space and ∇ is an isolated state then $E \cup \{\nabla\}$ is a Borel measurable space with its Borel σ-field given by $\mathcal{E} \vee \{\nabla\}$.[3]

B.2 Random counting measures and marked point processes

Taking aside the singular approach by Jacod and Skorokhod [68], the standard theory of piecewise deterministic processes (PDP) is based on the theory of random counting measures and a one-to-one identification between their events and the paths of a PDP. Therefore, we present subsequently a brief introduction to the topic and state the most important results which are regularly used in the main part of the thesis. The presentation follows [64] but to a large extent the results are analogous to the theory presented in [37, Appendix], nonetheless, the presentation in [64] is more general and, particularly, more detailed. All results which are proved in [64] are stated without proofs. We note that we use simplifications in notation in comparison to [64] in order to improve the readability of the section.

In the following sections the theory is first presented on canonical spaces, however, in Section B.2.7 we provide the necessary extensions to random counting measures / marked point processes on arbitrary probability spaces. In particular, the discussion is concluded with rigorously establishing a connection between the complementary approaches to PDMPs by [37, 64] and [68].

For the sake of completeness we also mention [21, 65, 66] wherein random counting measures are considered, called multivariate point processes in the terminology of the latter two references. The results presented below generally have counterparts in these presentations. We note that in [21] results are proved directly for marked point processes on arbitrary probability spaces (and filtrations of a special form (B.2.34)) without first considering canonical spaces. However, the presentation therein is restricted to compensators which allow for an intensity process, cf. Section B.2.6, which is a smaller class of marked point processes than considered in [37, 64] as well as in this thesis.

B.2.1 Definition and construction of random counting measures

Let $(E, \mathcal{E})$ be a Borel measurable space and ∇ is the *irrelevant mark* corresponding to a 'dead' state. We employ the notation $\overline{E} = E \cup \{\nabla\}$ and $\overline{\mathcal{E}} = \mathcal{E} \vee \{\nabla\}$ and note that $(\overline{E}, \overline{\mathcal{E}})$ is again a Borel measurable space. Further, $\mathcal{B}$ is used to denote the usual

[3]We note that in connection with σ-field the notation '$\vee$' always indicates the σ-field generated by the collection of sets on both sides of the symbol.

Borel sets on $[0,\infty)$ and $\overline{\mathcal{B}} := \mathcal{B} \vee \{\infty\}$. Finally, $\mathbb{N}_0$ is used to denote the set of natural numbers including zero and $\overline{\mathbb{N}}_0 := \mathbb{N} \cup \{\infty\}$ where, as usual, these spaces are equipped with the Borel sets with respect to the discrete topology.

Definition B.2.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be an arbitrary probability space. A *simple point process* on $(\Omega, \mathcal{F}, \mathbb{P})$ is a sequence $(\tau_n)_{n\geq 1}$ of $[0,\infty]$-valued random variables such that[4]

(i) $\mathbb{P}[0 < \tau_1 \leq \tau_2 \leq \ldots] = 1$,

(ii) $\mathbb{P}[\tau_n < \tau_{n+1}, \tau_n < \infty] = \mathbb{P}[\tau_n < \infty]$, $n \geq 1$,

(iii) $\mathbb{P}[\lim_{n\to\infty} \tau_n = \infty] = 1$.

A *marked point process* on $(\Omega, \mathcal{F}, \mathbb{P})$ is a sequence $(\tau_n, Y_n)_{n\geq 1}$ of paired random variables, where in addition to (i)-(iii) the $\overline{E}$-valued random variables Y_n satisfy[5]

(iv) $\mathbb{P}[Y_n \in E, \tau_n < \infty] = \mathbb{P}[\tau_n < \infty]$, $n \geq 1$,

(v) $\mathbb{P}[Y_n = \nabla, \tau_n = \infty] = \mathbb{P}[\tau_n = \infty]$, $n \geq 1$.

The set of paired sequences satisfying (i)-(v) is denoted by K, i.e.,

$$K := \Big\{ (t_n, y_n)_{n\geq 1} \in ([0,\infty] \times \overline{E})^{\mathbb{N}} \,\Big|\, 0 < t_1 \leq t_2 \leq \ldots ;\, t_n < t_{n+1}, y_n \in E \text{ iff } t_n < \infty \Big\}.$$

We define for all $n \geq 1$ the *coordinate projections*

$$T_n : K \to ([0,\infty], \overline{\mathcal{B}}) : k \mapsto T_n(k) := t_n, \qquad Y_n : K \to (\overline{E}, \overline{\mathcal{E}}) : k \mapsto Y_n(k) := y_n. \tag{B.2.1}$$

Then a σ-field $\mathcal{K}$ is generated on K by these projections, i.e., $\mathcal{K} = \sigma(T_n,\, Y_n,\, n \geq 1)$, which is the product σ-field on the product space K, and $(K, \mathcal{K})$ is a Borel measurable space. A probability on $(K, \mathcal{K})$ is called a *canonical marked point process*. Finally, we define the mappings

$$Z_n : K \to ([0,\infty] \times \overline{E})^n : k \mapsto ((T_1(k), Y_1(k)), \ldots, (T_n(k), Y_n(k))), \tag{B.2.2}$$

[4]In words: It is a sequence of (i) almost surely positive, non-decreasing random variables, which are (ii) almost surely strictly increasing if finite and (iii) almost surely 'non-explosive', i.e., almost surely there are only finitely many points in every finite interval.

[5]In words: The mark is (iv) almost surely in E if the event happens in finite time, otherwise (v) it is almost surely the irrelevant mark if the next event happens at infinity.

We briefly comment on the terminology of a process 'dying' and 'exploding'. To this end identify a marked point process with a piecewise constant stochastic process on $[0,\infty]$ taking values in $\overline{E}$ which for $\tau_n \leq t < \tau_{n+1}$ is at position Y_n and just assume an arbitrary initial condition for the process prior to the first jump. Then the difference between the process 'dying' and 'exploding' in finite time is the following: If a process dies it enters the dead state after finitely many jumps in finite time, i.e., a pair (τ_n, ∇) with $\tau_n < \infty$ is an event with non-zero probability, whereas if it explodes then it jumps infinitely often in finite time and after it had jumped infinitely often it is in the dead state. Both behaviours do not occur almost surely for a marked point process according to Definition B.2.1 as, on the one hand, $\mathbb{P}[t_n = \infty | Y_n = \nabla] = 1$ and hence a marked point process enters – and stays in – the irrelevant state $\{\nabla\}$ at infinity, meaning that a trajectory may stop jumping after finitely many jumps but is still 'alive'. On the other hand condition (iii) excludes explosion.

which are measurable with respect to $\mathcal{K}$ and the product σ-field on $([0,\infty] \times \overline{E})^n$.

Remark B.2.1. Given a marked point process on an arbitrary probability space $(\Omega, \mathcal{F}, \mathbb{P})$ this process can be seen as a K-valued random variable and we obtain a probability $\widehat{\mathbb{P}}$ on $(K, \mathcal{K})$ by the pushforward measure, i.e., for all $B \in \mathcal{K}$

$$\widehat{\mathbb{P}}(B) := \mathbb{P}\big(\{\omega \in \Omega : (\tau_n(\omega), Y_n(\omega))_{n\geq 1} \in B\}\big).$$

However, the aim is to construct a probability on $(K, \mathcal{K})$ directly.

Definition B.2.2. (a) We denote by H the space of *discrete counting measures* m on the measurable space $([0,\infty) \times E, \mathcal{B} \otimes \mathcal{E})$[6] which are of the form

$$m := \sum_{n=1}^{\infty} \delta_{t_n, y_n}, \tag{B.2.3}$$

where $(t_n, y_n)_{n\geq 1}$ is a sequence in K, i.e., the measure m is a sum of Dirac-measures putting unit mass at the points given by the elements of the sequence. In particular, such measures m are positive, σ-finite and satisfy $m(\,\cdot\,) \in \overline{\mathbb{N}}_0$, $m(\{0\} \times E) = 0$ and for $t > 0$ we have that $m(\{t\} \times E) \leq 1$, $m([0,t] \times E) < \infty$. To turn H into a measurable space we consider for all $C \in \mathcal{B} \otimes \mathcal{E}$ the mappings

$$\mu^C : H \to \overline{\mathbb{N}}_0 : m \mapsto m(C) \tag{B.2.4}$$

and define a σ-field on $\mathcal{H}$ on H by

$$\mathcal{H} := \sigma(\mu^C, C \in \mathcal{B} \otimes \mathcal{E})\,. \tag{B.2.5}$$

A probability on the measurable space $(H, \mathcal{H})$ is called a *canonical random counting measure.*

(b) Instead of mappings μ^C we can define for all $A \in \mathcal{E}$ a *counting process* $(N_t^A)_{t\geq 0}$ on H by

$$N_t^A : H \to \overline{\mathbb{N}}_0 : m \mapsto \mu^{[0,t]\times A}(m) = m([0,t] \times A)\,. \tag{B.2.6}$$

We frequently employ the shorter notation $N_t := N_t^E$, i.e., N_t counting all the jumps that occur up to time t. It turns out that

$$\sigma(N_t^A;\, t \geq 0, A \in \mathcal{E}) = \mathcal{H}\,. \tag{B.2.7}$$

That is, the σ-field on H generated by all counting processes (B.2.6) and the σ-field generated by all counting measures (B.2.4) coincide, hence the two families

[6] A further notational remark: In connection with σ-fields the symbol '$\otimes$' is always used to denote product σ-fields on cartesian product spaces indicated by the symbol '$\times$'.

of processes carry the same information.

(c) Finally, on the space $(H, \mathcal{H})$ the *canonical filtration* $(\mathcal{H}_t)_{t\geq 0}$ is given by the σ-fields

$$\mathcal{H}_t := \sigma(N_s^A;\, 0 \leq s \leq t, A \in \mathcal{E}) \tag{B.2.8}$$

where for $t = 0$ we get the trivial σ-field, i.e., $\mathcal{H}_0 = \{\emptyset, H\}$.

Remark B.2.2. Given a marked point process $(\tau_n, Y_n)_{n\geq 1}$ on an arbitrary probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we can, on the one hand, use it to define a random counting measure μ, i.e., an H-valued random variable on $(\Omega, \mathcal{F}, \mathbb{P})$, by

$$\mu(\omega) := \sum_{n=1}^{\infty} \mathbb{I}_{[\tau_n<\infty]}\, \delta_{\tau_n(\omega), Y_n(\omega)}, \tag{B.2.9}$$

where $\mu(\omega) \in H$ almost surely. Hence we obtain a probability on $(H, \mathcal{H})$ by the pushforward measure.
On the other hand, we can also define families of counting processes from a marked point process, i.e., for each $A \in \mathcal{E}$ let $(N_t^A)_{t\geq 0}$ be the process defined by

$$N_t^A(\omega) := \sum_{n=1}^{\infty} \mathbb{I}_{[\tau_n(\omega)\leq t, Y_n(\omega)\in A]} \;=\; \mu(\omega)([0, t] \times A) \tag{B.2.10}$$

which counts the number of events in $[0, t]$ with marks in A.

Proposition B.2.1. ([64, Sec. 3.2]) *The sequence elements T_n and Y_n can be recovered in a measurable fashion from the family of random measures $\{\mu^C,\, C \in \mathcal{B} \otimes \mathcal{E}\}$ or the family of counting processes $\{(N_t^A)_{t\geq 0},\, A \in \mathcal{E}\}$, respectively. Particularly, this implies that there exists a bimeasurable bijection $\Psi : (K, \mathcal{K}) \to (H, \mathcal{H})$.*

Note that Proposition B.2.1 implies that the coordinate projections (B.2.1) are measurable mappings on the space of counting measures $(H, \mathcal{H})$ in the sense that $T_n(m) = T_n \circ \Psi^{-1}(m)$ and $Y_n(m) = Y_n \circ \Psi^{-1}(m)$. Moreover, this turns $(H, \mathcal{H})$ into a Borel measurable space as defined in [64], i.e., there exists a bimeasurable bijection from $(H, \mathcal{H})$ to a Borel subset of the unit interval.

For the construction of a canonical marked point process we introduce for $n \geq 1$ the additional sets

$$K^n \;:=\; \Big\{(t_k, y_k)_{1\leq k\leq n} \in ([0, \infty] \times \overline{E})^n \,\Big|\, 0 < t_1 \leq \ldots \leq t_n \leq \infty; \quad \text{with } t_k < t_{k+1}, y_k \in E \text{ iff } t_k < \infty\Big\},$$

$$J^n \;:=\; \Big\{(z_n, t) \in K^n \times [0, \infty] \,\Big|\, t_n \leq t, \text{ with } t_n < t \text{ iff } t_n < \infty\Big\},$$

and the set $J^0 := [0, \infty]$. That is $K^n = Z_n(K)$ is the projection of the sequence space K to its first n entries and J^n is this projection plus an additional time element which is compatible for being the $(n+1)$th time element of a sequence in K. Generated by the corresponding coordinate projections (B.2.1) we obtain σ-fields $\mathcal{K}^n$, $\mathcal{J}^n$ on K^n and J^n, respectively, and thus obtain measurable spaces. As before these spaces are Borel measurable spaces and $\mathcal{K}^n$ and $\mathcal{J}^n$ are the respective product σ-fields.

Then, a canonical marked point process is successively constructed defining the marginal distributions of the time τ_{n+1} given $(\tau_1, \dots, \tau_n; Y_1, \dots, Y_n) \in K^n$ and of the mark Y_{n+1} given $(\tau_1, \dots, \tau_{n+1}; Y_1, \dots, Y_n) \in J^n$. We assume

(i) there is a probability P^0 on $[0, \infty]$, which defines the first jump time,

(ii) for every $n \geq 1$ there is a Markov kernel P^n from $(K^n, \mathcal{K}^n)$ to $([0, \infty], \mathcal{B})$ and

(iii) for every $n \geq 0$ there is a Markov kernel π^n from $(J^n, \mathcal{J}^n)$ to $(\overline{E}, \overline{\mathcal{E}})$.

The following theorem provides the existence of a canonical random counting measure.

Theorem B.2.1. ([64, Thm. 3.2.1]) *Given a probability P^0 and Markov kernels $(P^n)_{n\geq 1}$, $(\pi^n)_{n\geq 0}$ satisfying for all $z_n \in K^n$ and all $(z_n, t) \in J^n$*

$$\begin{aligned} &P^n(z_n, (t_n, \infty]) = 1 \quad \text{if } t_n < \infty, \quad P^n(z_n, \{\infty\}) = 1 \qquad \text{if } t_n = \infty, \\ &\pi^n((z_n, t), E) = 1 \qquad \text{if } t < \infty, \quad \pi^n((z_n, t), \{\nabla\}) = 1 \quad \text{if } t = \infty, \end{aligned} \tag{B.2.11}$$

then there exists a unique probability $\widehat{\mathbb{P}}$ on the paired sequence space containing K but also all explosive sequences, such that

(i) *$P^0(B) = \widehat{\mathbb{P}}[\,T_1^{-1}(B)\,]$ for all $B \in \mathcal{B}$, i.e., P^0 is the pushforward measure of the first time entry in the sequences,*

(ii) *$P^n(z_n, \cdot\,)$ is a regular conditional distribution of T_{n+1} given $[Z_n = z_n]$ for all $z_n \in K^n, n \geq 1$,*

(iii) *$\pi^n((z_n, t), \cdot\,)$ is a regular conditional distribution of Y_{n+1} given $[Z_n = z_n,\ T_{n+1} = t]$ for all $(z_n, t) \in J^n, n \geq 0$.*

Moreover, if and only if

$$\widehat{\mathbb{P}}\Big[\lim_{n\to\infty} T_n = \infty\Big] = 1 \tag{B.2.12}$$

then $\widehat{\mathbb{P}}$ defines a canonical marked point process $\mathbb{P}$, i.e., $\widehat{\mathbb{P}}(K) = 1$, and the probability $\mathbb{P}$ is the restriction of the probability $\widehat{\mathbb{P}}$ to K.

Due to the bimeasurable Ψ map between $(K, \mathcal{K})$ and $(H, \mathcal{H})$ we immediately obtain under Theorem B.2.1 a probability measure on the latter space, i.e., a canonical random counting measure, by the pushforward measure. Further, it is clear that

different choices of Markov kernels can lead to the same probability on $(H, \mathcal{H})$. In particular, Markov kernels generating the same probability may differ on a set which is attained by Z_n or (Z_n, T_{n+1}), respectively, only with probability zero.
It is cumbersome and difficult, cf. [37] and [64, p. 18], to give general conditions on the Markov kernel defining a probability on the paired sequence space such that it is a canonical marked point process. Davis discusses simple conditions [37, Prop. 24.6] for special cases and Jacobsen gives conditions in connection with compensators and intensity processes, cf. [64, p. 62,68]. These simple conditions are usually easy to verify for special cases of interest as, e.g., the hybrid models in Chapter 3.

We continue defining conditional distributions for canonical random counting measures with respect to their past. In order to define conditional probabilities with respect to the 'past' of processes we define the following operators.

Definition B.2.3. We define *shift operators* θ_s, $s \geq 0$, for counting processes and random counting measures by

$$\theta_s N_t^A := \begin{cases} 0 & \text{if } t < s, \\ N_t^A - N_s^A & \text{if } s \leq t, \end{cases}$$

$$\theta_s \mu := \mu(\,\cdot \cap ((s, \infty) \times E)) = \sum_{n \in \mathbb{N}: s < \tau_n < \infty} \delta_{\tau_n, Y_n}\,.$$

That is, the shifted counting process $(\theta_s N_t^A)_{t \geq 0}$ only counts points that occur strictly after s and the shifted random measure $\theta_s \mu$ is supported on (s, ∞). Writing $(\tau_{n,s})_{n \geq 1}$, $(Y_{n,s})_{n \geq 1}$ for the sequence of time points and marks determining $\theta_s \mu$ we have, e.g., on the set $[N_s = k]$ that

$$\tau_{n,s} = \theta_s \tau_n = \tau_n(\theta_s \mu) = \tau_{k+n}, \quad Y_{n,s} = \theta_s Y_n = Y_n(\theta_s \mu) = Y_{k+n},$$

which is, in more general terms,

$$\theta_s \tau_n = \tau_{N_s + n}, \quad \theta_s Y_n = Y_{N_s + n}.$$

This means that $\tau_{n,s}$ and $Y_{n,s}$ are the nth jump time and mark strictly after time s and the shifted random counting measure weights $A \in \mathcal{E}$ by the marks therein corresponding only to time points strictly after s.

Further, we also define the *translated shift* θ_s^*, $s \geq 0$ by

$$\theta_s^* \mu = \sum_{n \in \mathbb{N}: s < \tau_n < \infty} \delta_{\tau_n - s, Y_n} \tag{B.2.13}$$

and hence for $C \in \mathcal{B} \otimes \mathcal{E}$

$$\theta_s^* \mu(C) = \theta_s \mu(C + s)$$

where $C + s = \{(t + s, y) | \, (t, y) \in C\}$. We note that the translated shift operators form a semigroup on H, i.e.,

$$\theta_0^* \mu = \mu \quad \text{and} \quad \theta_{t+s}^* \mu = \theta_s^* \circ \theta_t^* \mu \, .$$

Further, the connection between shifted random measures is given for $B \in \mathcal{E}$ by

$$\theta_s^* \mu((t_1, t_2] \times B) = \theta_s \mu((t_1 + s, t_2 + s] \times B) = \mu((t_1 + s, t_2 + s] \cap (s, \infty) \times B).$$

The translated shift operator is of particular importance when considering a homogeneous Markov property for random counting measures and is used in Chapter 2, however, it is not discussed further in this appendix.
It is a straightforward extension to define shift operators with respect to random times. Of particular interest are shift operators θ_{T_k}, $\theta_{T_k}^*$, $k \in \mathbb{N}$, which restrict the counting processes and random measures to events after the kth jump time.

The following theorem provides the connection between the transition kernels for probabilities conditioned on the 'past' of a point process. To this end we define the *concatenation operation* between *acceptable* elements of two sets K^k and K^n, $k, n \in \mathbb{N}$, i.e.,

$$\sqcup : \, K^k \times K^n \mapsto K^{k+n} : z_k \sqcup \widetilde{z}_n = (t_1, \ldots, t_k, \widetilde{t}_1, \ldots, \widetilde{t}_n; y_1, \ldots, y_k, \widetilde{y}_1, \ldots, \widetilde{y}_n) \, .$$

Here 'acceptable' means that the two events $z_k = (t_1, \ldots, t_k; y_1, \ldots, y_n) \in K^k$ and $\widetilde{z}_n = (\widetilde{t}_1, \ldots, \widetilde{t}_n; \widetilde{y}_1, \ldots, \widetilde{y}_n) \in K^n$ are such that $t_k < \widetilde{t}_1$ iff $t_k < \infty$ and $\widetilde{t}_1 = \infty$ iff $t_k = \infty$, i.e., the resulting concatenation is an element of K^{k+n}.

Theorem B.2.2. ([64, Thm. 4.3.3]) *For any $s \in [0, \infty), k \in \mathbb{N}_0$ and all $z_k \in K^n$ the conditional distribution of $\theta_s \mu$ given $[N_s = k, Z_k = z_k]$ is the probability $\mathbb{P}|_{k,z_k}$ on $(H, \mathcal{H})$ generated by the Markov kernels $(P^n_{|k,z_k})_{n \geq 0}$ from K^n to $[0, \infty]$ and $(\pi^n_{|k,z_k})_{n \geq 0}$ from J^n to $\overline{E}$ given by*

$$P^0_{|k,z_k} \quad := \quad P^k(z_k, \, \cdot \, | (s, \infty]),$$

$$P^n_{|k,z_k}(\widetilde{z}_n, \, \cdot \,) \quad := \quad P^{k+n}(z_k \sqcup \widetilde{z}_n, \, \cdot \,), \quad \widetilde{z}_n \in K^n, n \geq 1,$$

$$\pi^n_{|k,z_k}((\widetilde{z}_n, t), \, \cdot \,) \quad := \quad \pi^{k+n}((z_k \sqcup \widetilde{z}_n, t), \, \cdot \,), \quad \widetilde{z}_n \in K^n, n \geq 0, \, t > \widetilde{t}_n.$$

For later, very practical use we obtain the following version of a total conditional expectation.

Lemma B.2.1. ([64, Corol. 4.2.2]) *Let* $\mathbb{P}$ *be a probability on* $(H, \mathcal{H})$ *and* U *be a real-valued, integrable random variable, then*

$$\mathbb{E}[U|\mathcal{H}_t] = \sum_{n=1}^{\infty} \mathbb{I}_{[N_t=n]}\, \mathbb{E}\big[\, U \,\big|Z_n, \mathbb{I}_{[\tau_{n+1}>t]}\big] \, .$$

We note that on $[\tau_{n+1} > t]$

$$\mathbb{E}\big[U\big|Z_n, \mathbb{I}_{[\tau_{n+1}>t]}\big] = \frac{1}{S^n_{Z_n}(t)} \mathbb{E}\big[U\, \mathbb{I}_{[\tau_{n+1}>t]}\big|Z_n\big] \, .$$

Thus, Lemma B.2.1 implies that in order to calculate conditional expectations with respect to the elements $\mathcal{H}_t$ of the canonical filtration it is sufficient to consider the conditional expectations on the sets $[Z_n = z_n, N_t = n]$. That means the conditional probabilities in Theorem B.2.2 simply determine the conditional distribution of $\theta_s\mu$ given $\mathcal{H}_s$ on the set $[N_s = k, Z_k = z_k]$ and hence the conditional probability

$$\mathbb{P}[\theta_s\mu \in \cdot\, |\mathcal{H}_s] \quad =: \mathbb{P}^{N_s, Z_{N_s}}$$

is the probability on $(H, \mathcal{H})$ generated by the families of Markov kernels $(P^n_{|N_s, Z_{N_s}})_{n\geq 0}$ and $(\pi^n_{|N_s, Z_{N_s}})_{n\geq 0}$.

B.2.2 Stochastic processes on canonical spaces

Definition B.2.4. A *stochastic process* $(X_t)_{t\geq 0}$ with *state space* $(G, \mathcal{G})$ defined on $(H, \mathcal{H})$ is a family of G-valued random variables X_t. We say such a process is

(i) *measurable* if the map $(t, m) \mapsto X_t(m)$ is $\mathcal{B} \otimes \mathcal{H}$-measurable,

(ii) *adapted* to the filtration $(\mathcal{H}_t)_{t\geq 0}$ if it is measurable and each X_t is $\mathcal{H}_t$-measurable or

(iii) $\mathcal{H}_t$*-predictable* if X_0 is $\mathcal{H}_0$-measurable and if the map $(t, m) \mapsto X_t(m)$ restricted to $((0, \infty) \times H, \mathcal{B} \otimes \mathcal{H})$ is measurable with respect to the *predictable* σ*-field*, which is the σ-field generated by the sets

$$(t, \infty) \times B, \quad t \geq 0,\; B \in \mathcal{H}_t.$$

In particular, left continuous, $\mathcal{H}_t$-adapted processes are $\mathcal{H}_t$-predictable.

The following theorem characterises the elements of the σ-fields of the canonical filtration and the class of processes which are adapted or predictable with respect to the canonical filtration $(\mathcal{H}_t)_{t\geq 0}$. Particularly, note that the filtration $(\mathcal{H}_t)_{t\geq 0}$ is NOT completed.

Theorem B.2.3. ([64, Prop. 4.2.1(b)])

(a) *A set $B \subseteq H$ belongs to $\mathcal{H}_t$ if and only if for every $n \geq 0$ there exists a set $B_n \in (\mathcal{B} \otimes \mathcal{E})^n \cap ((0,t] \times E)^n$ such that*

$$B \cap [N_t = n] = \big\{ m \in H : Z_n(m) \in B_n,\, T_{n+1}(m) > t \big\}\,. \tag{B.2.14}$$

(b) *The canonical filtration is right-continuous, i.e., $\mathcal{H}_t = \bigcap_{\epsilon>0} \mathcal{H}_{t+\epsilon}$ for all $t \geq 0$.*

(c) *A real valued process $(X_t)_{t\geq 0}$ is adapted or predictable, respectively , if and only if for all $n \geq 0$ there exist measurable functions*

$$f^n : ((0,\infty) \times E)^n \times [0,\infty) \to \mathbb{R} : (z_n, t) \mapsto f^n_{z_n}(t)$$

such that for all $t \in [0,\infty)$ identically on H

$$X_t = f^{N_t}_{Z_{N_t}}(t) = \sum_{n\in\mathbb{N}:T_n<\infty} f^n_{Z_n}(t)\, \mathbb{I}_{[T_n,T_{n+1})}(t) \tag{B.2.15}$$

or

$$X_t = f^{N_{t-}}_{Z_{N_{t-}}}(t) = \sum_{n\in\mathbb{N}:T_n<\infty} f^n_{Z_n}(t)\, \mathbb{I}_{(T_n,T_{n+1}]}(t), \tag{B.2.16}$$

respectively. Note that we set $T_0 := 0$ and always $X_0 = f^0(0)$, where the condition that X_0 is $\mathcal{H}_0$-measurable implies that X_0 is constant.

Recall that $N_t \in \mathbb{N}_0$ is the random variable that counts the number of points up to and including time t, cf. (B.2.6). Hence, the left limit N_{t-} counts the number of events strictly before time t. Thus the theorem states that the value of an adapted process at time t can be computed from the number of events in $[0,t]$, their time points and their marks. On the other hand, for a process to be predictable it suffices to know the points on $[0,t)$, their location and marks. Further, if a process $(X_t)_{t\geq 0}$ is adapted this implies clearly that $\mathcal{H}^X_t \subset \mathcal{H}_t$. Moreover, the law of an adapted or predictable process is defined by the probability $\mathbb{P}$ as each X_t is a measurable map with respect to m and thus its distribution on $(G,\mathcal{G})$ is defined by the pushforward measure.

A $[0,\infty]$-valued random variable τ is a *stopping time* with respect to the σ-field $(\mathcal{H}_t)_{t\geq 0}$ if $[\tau \leq t] \in \mathcal{H}_t$ for all $t \geq 0$. Due to the right continuity of the filtration this is equivalent to $[\tau < t] \in \mathcal{H}_t$, i.e., every stopping time is an optional time and vice versa. We define the stopped σ-field as

$$\mathcal{H}_\tau = \{B \in \mathcal{H} : B \cap [\tau \leq t] \in \mathcal{H}_t \;\forall\, t \geq 0\}.$$

In particular, all T_n are stopping times and $\mathcal{H}_{T_n}$ is the σ-field generated by Z_n [64, p. 69].

Part (a) of Theorem B.2.3 also provides the connection to a complementary approach to piecewise deterministic Markov processes. The approach is based on the notion of a *jumping filtration* defined below. The authors in [68] then define *jumping Markov processes* as strong Markov processes that are adapted to a jumping filtration.

Definition B.2.5. [68, 67] A right-continuous filtration $(\mathcal{F}_t)_{t\geq 1}$ on an arbitrary probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called an *a.s. jumping filtration* if there exists a sequence of increasing stopping times $(\tau_n)_{n\geq 0}$ with $\tau_0 = 0$ and $\lim_{n\to\infty} \tau_n = \infty$ a.s., called a *jumping sequence*, such that for all $n \geq 0$ and $t \geq 0$

$$\mathcal{F}_t \cap [\tau_n \leq t < \tau_{n+1}] = \mathcal{F}_{\tau_n} \cap [\tau_n \leq t < \tau_{n+1}] \tag{B.2.17}$$

up to $\mathbb{P}$-null sets. A filtration is called *jumping filtration* if (B.2.17) holds for all sets. Note that for a jumping filtration the jumping sequence is by no means unique and the completion of an a.s. jumping sequence is a jumping sequence.

The property (B.2.17) heuristically means that the filtration is 'constant' between jump times. That is in probabilistic terms the paths of the process contain no additional information in between jumps compared to the information contained in the process just after the jump. The connection to marked point processes is now the following. On the one hand, every jumping sequence is generated by a marked point process, cf. [67], and on the other hand a canonical marked point process generates a jumping filtration, which is proved in the following proposition. That is, every $\mathcal{H}_t$-adapted process which in addition is strong Markov is a jumping Markov process in the sense of [68].

Proposition B.2.2. *The canonical filtration $(\mathcal{H}_t)_{t\geq 0}$ is a jumping filtration on $(H, \mathcal{H})$ with jumping sequence $(T_n)_{n\geq 0}$.*

Proof. First, we note that by Theorem B.2.3(b) we have that the canonical filtration is right-continuous. On the one hand, assume $B \in \mathcal{H}_t$, then by Theorem B.2.3(a)

$$\begin{aligned} B \cap [T_n \leq t < T_{n+1}] &= \{z_n \in C_n, T_{n+1} > t\} \\ &= \{z_n \in C_n\} \cap [T_n \leq t < T_{n+1}] \ \in \mathcal{H}_{T_n} \cap [T_n \leq t < T_{n+1}] \end{aligned}$$

as $\mathcal{H}_{T_n} = \sigma(Z_n)$. On the other hand, assume $B \in \mathcal{H}_{T_n}$ then

$$B \cap [T_n \leq t < T_{n+1}] = B \cap [T_n \leq t] \cap [T_n \leq t < T_{n+1}] \in \mathcal{H}_t \cap [T_n \leq t < T_{n+1}]$$

by the definition of the stopped σ-field. As the jump times of a canonical marked point process are increasing and satisfy $\lim_{n\to\infty} T_n = \infty$ by definition they constitute a valid jumping sequence. □

B.2.3 Doob-Meyer Decomposition

Let $\mathbb{P}$ be the probability on $(H, \mathcal{H})$ defined by the initial probability P^0 and the Markov kernels $(P^n)_{n\geq 1}$ and $(\pi^n)_{n\geq 0}$. If $t_n < \infty$ let $\nu^n_{z_n}$ denote the hazard measure, cf. Section B.3, for the probability $\mathbb{P}^n_{z_n}$ on $[0, \infty]$ defined by the Markov kernel $P^n(z_n, \cdot)$. Since the probability $\mathbb{P}^n_{z_n}$ is concentrated on $(t_n, \infty]$ so is $\nu^n_{z_n}$.

Definition B.2.6. (a) The *total compensator for* $\mathbb{P}$ is the process $(\widetilde{N}_t)_{t\geq 0}$ on $(H, \mathcal{H})$ given by

$$\widetilde{N}_t := \sum_{n=0}^{N_t} \nu^n_{Z_n}\big((\tau_n, \tau_{n+1} \wedge t]\big) . \tag{B.2.18}$$

(b) The $\mathbb{P}$-*compensator for the counting process* $(N^C_t)_{t\geq 0}$ for every $C \in \mathcal{E}$ is the process $(\widetilde{N}^C_t)_{t\geq 0}$ on $(H, \mathcal{H})$ given by

$$\widetilde{N}^C_t := \int_{[0,t]} \pi^{N_{s-}}\big((Z_{N_{s-}}, s), C\big)\, \mathrm{d}\widetilde{N}_s . \tag{B.2.19}$$

Obviously, we have that $\widetilde{N}^E_t = \widetilde{N}_t$.

(c) The *compensating measure for* $\mathbb{P}$ is the random, non-negative, almost surely σ-finite measure $\widetilde{N}$ on $[0, \infty) \times E$ given for all $A \in \mathcal{B} \otimes \mathcal{E}$ by

$$\widetilde{N}(A) := \int_{[0,\infty)} \int_E \mathbb{I}_A(s, y)\ \pi^{N_{s-}}\big((Z_{N_{s-}}, s), \mathrm{d}y\big)\, \mathrm{d}\widetilde{N}_s \tag{B.2.20}$$

Obviously we have that $\widetilde{N}([0, t] \times C) = \widetilde{N}^C_t$.

The compensators $(\widetilde{N}_t)_{t\geq 0}$ and $(\widetilde{N}^C_t)_{t\geq 0}$, $C \in \mathcal{E}$, are non-negative, increasing, right-continuous processes starting at 0 for $t = 0$. They are finite almost surely and thus can be associated with a positive σ-finite random measure on $([0, \infty), \mathcal{B})$. We note that the compensators are defined from the Markov kernels generating the probability $\mathbb{P}$ on $(H, \mathcal{H})$ via the hazard measures ν^n. The labelling '$\mathbb{P}$-compensator' is, however, justified as the compensators obtained from two sets of Markov kernels generating the same probability are indistinguishable [64, p. 52].

Proposition B.2.3. ([64, Prop. 4.3.1(b)]) *The compensators* $(\widetilde{N}^C_t)_{t\geq 0}$ *for the counting processes* $(N^C_t)_{t\geq 0}$ *are predictable for all* $C \in \mathcal{E}$.

Finally, the compensators characterise probabilities on $(H, \mathcal{H})$ uniquely, that is, two different probabilities cannot have the same compensators.

Theorem B.2.4. ([64, Thm. 4.3.2]) *Suppose* $\widetilde{N}$ *is the compensating measure for some probability* $\mathbb{P}$ *on* $(H, \mathcal{H})$*, then* $\mathbb{P}$ *is uniquely defined.*

The main result in this section is the Doob-Meyer-Decomposition of the counting processes which, obviously, are adapted, locally integrable increasing processes. Hence, due to the Doob-Meyer Decomposition (cf. [66, Sect. I.3b]) there exists a unique increasing process, predictable and zero at time zero, such that the counting processes can be represented by a sum of a local martingale and this increasing process. In particular, it is shown that this unique increasing process is precisely the compensator defined in Definition B.2.6.

Recall that an adapted, real-valued process $(M_t)_{t\geq 0}$ is called a *local martingale* if there exists an increasing sequence of stopping times ρ_n, called *localising sequence*, with $\lim_{n\to\infty} \rho_n = \infty$ a.s. such that for each $n \geq 1$ the stopped process $(M_{t\wedge\rho_n})_{t\geq 0}$ is a martingale. Note that (local) martingales are defined with respect to a given filtration. In this section this filtration is always the canonical filtration $(\mathcal{H}_t)$ and usually the localising sequence is given by the random times T_n. In subsequent sections and the main body of the thesis, when processes are defined on arbitrary probability spaces, we include the filtration in the specification of a (local) martingale if ambiguities are possible.

Theorem B.2.5. (Doob-Meyer-Decomposition [64, Thm. 4.5.2]) *Let $\mathbb{P}$ be a probability on $(H,\mathcal{H})$ with compensating measure $\widetilde{N}$ and compensators $(\widetilde{N}_t^C)_{t\geq 0}$. Then for every $C \in \mathcal{E}$ the process $(M_t^C)_{t\geq 0}$ given by*

$$M_t^C := N_t^C - \widetilde{N}_t^C$$

is a local martingale with respect to the sequence of stopping times $(T_n)_{n\geq 1}$. Moreover, the compensator $(\widetilde{N}_t^C)_{t\geq 0}$ is up to indistinguishability with respect to the probability $\mathbb{P}$ the unique right-continuous $\mathcal{H}_t$-predictable process $(V_t)_{t\geq 0}$ satisfying $V_0 = 0$ a.s., such that $M_t := N_t^C - V_t$ is a local martingale.

We note that a sufficient condition for the local martingale $(M_t^C)_{t\geq 0}$ to be a martingale is that $\mathbb{E}N_t^C < \infty$ for all $t \geq 0$. The following proposition collects some further properties of the processes $(M_t^C)_{t\geq 0}$. The statements are restricted to random counting measures that have continuous compensators which is sufficient for applications to PDMPs that arise in this thesis. A generalisation to discontinuous compensators is possible, see [64, p.74].

Proposition B.2.4. ([64, Prop. 4.5.3]) *Let $\mathbb{P}$ be a probability on $(H,\mathcal{H})$ with compensators $(\widetilde{N}_t^C)_{t\geq 0}$ and continuous total compensator $(\widetilde{N}_t)_{t\geq 0}$, then it holds that*

(a) *for every $C \in \mathcal{E}$ the process*

$$t \mapsto (M_t^C)^2 - \widetilde{N}_t^C \tag{B.2.21}$$

is a local martingale and it is a martingale in case that $\mathbb{E}N_t^C < \infty$. *Moreover, for all* $n \geq 1$, $t \geq 0$ *it holds that*

$$\mathbb{E}(M_{t\wedge T_n}^C)^2 < \infty\,,$$

while for a given $t \geq 0$ *we a have that* $\mathbb{E}(M_t^C)^2 < \infty$ *if* $\mathbb{E}N_t^C < \infty$;

(b) *for every* $C, D \in \mathcal{E}$ *with* $C \cap D = \emptyset$ *the process*

$$t \mapsto M_t^C\, M_t^D \tag{B.2.22}$$

is a local martingale and it is a martingale if $\mathbb{E}N_t^C, \mathbb{E}N_t^D < \infty$ *for all* $t \geq 0$. *Moreover, for all* $n \geq 1$, $t \geq 0$ *it holds that*

$$\mathbb{E}|M_{t\wedge T_n}^C\, M_{t\wedge T_n}^D| < \infty\,,$$

while for a given $t \geq 0$ *we have that* $\mathbb{E}|M_t^C\, M_t^D| < \infty$ *if* $\mathbb{E}N_t^C, \mathbb{E}N_t^D < \infty$.

B.2.4 Stochastic integrals and the Martingale Representation Theorem

In this section we define stochastic integrals with respect to random counting measures and their compensators. As integrands we choose real-valued, $\mathcal{B}\otimes\mathcal{E}\otimes\mathcal{H}$-measurable functions

$$g:\,[0,\infty)\times E\times H \to \mathbb{R}:\,(t,y,m)\mapsto g(t,y,m), \tag{B.2.23}$$

where most of the time we omit the dependence of g on the events $m \in H$ and just write $g(t,y)$. We say an integrand is adapted or predictable if for every $y \in E$ the process $(g(t,y))_{t\geq 0}$ is adapted or predictable, respectively.

Definition B.2.7. (a) For all integrands g the *stochastic integral with respect to the canonical random counting measure* m is a well defined real valued process $(N_t^g)_{t\geq 0}$ with

$$\begin{aligned} N_t^g \;&:=\; \int_{[0,t]\times E} g(s,y)\, m(\mathrm{d}s,\mathrm{d}y) \qquad\qquad (B.2.24)\\ &=\; \sum_{n\in\mathbb{N}:\tau_n\leq t} g(T_n(m),Y_n(m),m) \;=\; \sum_{n=1}^{N_t(m)} g(T_n(m),Y_n(m),m), \end{aligned}$$

which is almost surly a finite sum. If g is adapted then the stochastic integral is an adapted process for t. In (B.2.24) N_t is the counting process that counts all events up to and includeing time t, cf. (B.2.6).

(b) The *stochastic integral with respect to the compensating measure* is the process

$(\widetilde{N}_t^g)_{t\geq 0}$ given by

$$\widetilde{N}_t^g := \int_{[0,t]\times E} g(s,y)\,\widetilde{N}(\mathrm{d}s \times \mathrm{d}y) = \int_{(0,t]}\int_E g(s,y)\,\pi^{N_{s-}}(Z_{N_{s-}},\mathrm{d}y)\,\mathrm{d}\widetilde{N}_s\,. \quad \text{(B.2.25)}$$

This stochastic integral is adapted or predictable if the integrand g is adapted or predictable, respectively.

(c) The *compensated stochastic integral* is the process $(M_t^g)_{t\geq 0}$ defined by

$$M_t^g := N_t^g - \widetilde{N}_t^g = \int_{[0,t]} g(s,y)M(\mathrm{d}s,\mathrm{d}y), \quad \text{(B.2.26)}$$

where the measure $M := m - \widetilde{N}$ is called the *compensated random measure* or *associated martingale measure.*

We note that in definitions (B.2.24) – (B.2.26) it makes no difference in choosing the integration interval in the time domain as $(0,t]$ or $[0,t]$ as neither random counting measure m nor the compensator $\widetilde{N}$ have positive mass at the 'point' $\{0\}\times E$. Further, as $\widetilde{N}_t$ is for each $m \in \mathcal{H}$ an increasing function the process (B.2.25) is well defined as a $[0,\infty]$-valued Riemann-Stieltjes integral for integrands $g \geq 0$, as well as for integrands $g \leq 0$ when it is $[-\infty,0]$-valued.
Next, we present simple conditions such that the stochastic integrals with respect to the compensating measures are almost surely finite with respect to the measure $\mathbb{P}$. On the one hand, if $g \geq 0$ then it is sufficient for $\mathbb{P}[\widetilde{N}_t^g < \infty\ \forall\, t \geq 0] = 1$ that

$$\mathbb{P}\Big[\sup\nolimits_{s\leq t, y\in E} g(s,y) < \infty\ \forall\, t \geq 0\Big] = 1$$

and, on the other hand, if $g \leq 0$ it is sufficient for $\mathbb{P}[\widetilde{N}_t^g > \infty\ \forall\, t \geq 0] = 1$ that

$$\mathbb{P}\Big[\sup\nolimits_{s\leq t, y\in E} -g(s,y) < \infty\ \forall\, t \geq 0\Big] = 1.$$

Finally, we can define stochastic integrals for arbitrary g by $\widetilde{N}_t^g := \widetilde{N}_t^{g^+} - \widetilde{N}_t^{g^-}$, if the stochastic integrals of the positive and negative part of g, i.e., $\widetilde{N}_t^{g^+}$ and $\widetilde{N}_t^{g^-}$, are almost surely finite. Obviously, a sufficient condition that the stochastic integral for arbitrary g is almost surely finite is

$$\mathbb{P}\Big[\sup\nolimits_{s\leq t, y\in E} |g(s,y)| < \infty\ \forall\, t \geq 0\Big] = 1\,. \quad \text{(B.2.27)}$$

For the compensated stochastic integral we note that the martingale measure M is a random signed measure on $[0,\infty)\times E$ but it need not be defined on all sets of $\mathcal{B}\otimes\mathcal{E}$ as $m([0,\infty)\times E) = \widetilde{N}([0,\infty)\times E) = \infty$ almost surely is possible. However, the restriction of M to $[0,t]\times E$ for any $t \geq 0$ is well-defined and almost surely finite.

In accordance with the notation in [37, Appendix] we denote by $L^1(\mathbb{P})$ the set of all predictable, integrable processes g, i.e., all predictable g such that

$$\|g\|_{L^1(\mathbb{P})} := \mathbb{E}\int_{[0,\infty)\times E} |g(s,y,m)|\, m(\mathrm{d}s,\mathrm{d}y) = \mathbb{E}\sum_{n:\,T_n<\infty} |g(T_n,Y_n,m)| \,<\, \infty\,. \tag{B.2.28}$$

Further, we say g is locally integrable, i.e., $g \in L^1_{\mathrm{loc}}(\mathbb{P})$, if there exists a localising sequence of stopping times $(\tau_n)_{n\geq 1}$ such that $g\,\mathbb{I}_{[t<\tau_n]} \in L^1(\mathbb{P})$. Obviously, for a predictable integrand g condition (B.2.27) implies that $g \in L^1_{\mathrm{loc}}(\mathbb{P})$. Analogously, we define $L^1(\widetilde{N})$ and $L^1_{\mathrm{loc}}(\widetilde{N})$ where

$$\|g\|_{L^1(\widetilde{N})} := \mathbb{E}\int_{[0,\infty)\times E} |g(s,y)|\, \widetilde{N}(\mathrm{d}s\times\mathrm{d}y)\,.$$

In [37, Prop. A.3] it is proven that it holds $L^1(\mathbb{P}) = L^1(\widetilde{N})$, $L^1_{\mathrm{loc}}(\mathbb{P}) = L^1_{\mathrm{loc}}(\widetilde{N})$ and $\|\cdot\|_{L^1(\mathbb{P})} = \|\cdot\|_{L^1(\widetilde{N})}$. In particular, the stochastic integrals (B.2.24) – (B.2.26) are well-defined for $g \in L^1_{\mathrm{loc}}(\mathbb{P})$ and almost surely finite.

Theorem B.2.6. (Martingale Representation Theorem, [64, Thm. 4.6.1]) *Let $\mathbb{P}$ be a probability on $(H,\mathcal{H})$.*

(a) *Let $(M_t)_{t\geq 0}$ be a right-continuous, local martingale. Then there exists a predictable integrand g such that*[7]

$$M_t = M_0 + M_t^g\,.$$

(b) *Let $g \geq 0$ be a predictable integrand, then for all $t \geq 0$, $n \geq 1$*

$$\mathbb{E}N^g_{t\wedge T_n} = \mathbb{E}\widetilde{N}^g_{t\wedge T_n}\,.$$

If these expectations are finite for all $t \geq 0$, $n \geq 1$, then $(M_t^g)_{t\geq 0}$ is a local martingale with respect to the localising sequence $(T_n)_{n\geq 1}$.

(c) *Let $g \geq 0$ be a predictable integrand, then for all $t \geq 0$*

$$\mathbb{E}N^g_t = \mathbb{E}\widetilde{N}^g_t\,.$$

If these expectations are finite for all $t \geq 0$ then $(M_t^g)_{t\geq 0}$ is a martingale.

(d) *Let g be a predictable integrand then $(M_t^g)_{t\geq 0}$ is a local martingale with respect to the localising sequence $(T_n)_{n\geq 1}$ if $\mathbb{E}N^{|g|}_{t\wedge T_n} < \infty$ for all $t \geq 0$, $n \geq 1$, i.e., if $g \in L^1_{\mathrm{loc}}(\mathbb{P})$.*

[7] This implies piecewise continuity of a local martingale $(M_t)_{t\geq 0}$ (continuity on each $[\tau_n,\tau_{n+1})$) if the total compensator is continuous, cf. [64, Remark 4.6.2].

(e) *Let g be a predictable integrand then $(M_t^g)_{t\geq 0}$ is a martingale if $\mathbb{E}N_t^{|g|} < \infty$ for all $t \geq 0, n \geq 1$, i.e., if $g \in L^1(\mathbb{P})$.*

The section is concluded with results on the *quadratic variation* for local martingales. As for a corresponding preliminary result in Proposition B.2.4 we assume in the following that the total compensator $\widetilde{N}_t$ is continuous. Let g, h be predictable integrands such that the stochastic integrals $\widetilde{N}_t^{g^2}$ and $\widetilde{N}_t^{h^2}$ are almost surely finite. We define

$$\begin{aligned} \left\langle M^g \right\rangle_t &:= \widetilde{N}_t^{g^2}, \\ \left\langle M^g, M^h \right\rangle_t &:= \widetilde{N}_t^{gh}. \end{aligned} \tag{B.2.29}$$

Proposition B.2.5. ([64, Prop. 4.6.2]) *Under the above conditions it holds that the processes defined by*

$$(M_t^g)^2 - \left\langle M^g \right\rangle_t,$$

$$M_t^g M_t^h - \left\langle M^g, M^h \right\rangle_t$$

are local martingales if $\mathbb{E}N_{t\wedge T_n}^{g^2}, \mathbb{E}N_{t\wedge T_n}^{h^2} < \infty$ for all $t \geq 0$. If in addition $\mathbb{E}N_t^{g^2}$, $\mathbb{E}N_t^{h^2} < \infty$ for all $t \geq 0$ then these processes are martingales.

B.2.5 Itô formula

We next present a general Itô formula for piecewise-continuous, real-valued processes, which states that such a process can be written as a sum of a predictable process and a stochastic integral with a predictable integrand. Again we assume that the total compensator is continuous. The function f below results from the representation of real-valued adapted processes in Theorem B.2.3(c). As the process is piecewise continuous, f is continuous.

Theorem B.2.7. (Itô formula, [64, Thm. 4.7.1]) *Assume that the compensator $\widetilde{N}_t$ is continuous and $(X_t)_{t\geq 0}$ is an adapted, real-valued process which is piecewise continuous. Then, provided that*

$$\int_{[0,t]\times E} \left| f_{Z_{s-}\sqcup(s,y)}^{N_{s-}+1}(t) - f_{Z_{s-}}^{N_{s-}}(s) \right| \widetilde{N}(\mathrm{d}s, \mathrm{d}y) < \infty \quad \text{a.s.}$$

for all $t \geq 0$ (note that the integrand is predictable), it holds that for all $t \geq 0$

$$X_t = X_0 + U_t + \int_{[0,t]\times E} f_{Z_{s-}\sqcup(s,y)}^{N_{s-}+1}(t) - f_{Z_{s-}}^{N_{s-}}(s)\, M(\mathrm{d}s, \mathrm{d}y)\,. \tag{B.2.30}$$

The process $(U_t)_{t\geq 0}$ defined via (B.2.30) *is continuous and predictable. Furthermore,* *it holds that* $U_0 \equiv 0$ *and*

$$\int_{[0,0]\times E} f^{N_{s-}+1}_{Z_{s-}\sqcup(s,y)}(t) - f^{N_{s-}}_{Z_{s-}}(s)\, M(\mathrm{d}s,\mathrm{d}y) = 0\,.$$

Provided that the integral in (B.2.30) *is a local martingale, the decomposition* (B.2.30) *of* $(X_t)_{t\geq 0}$ *into the sum of its initial value, a predictable process starting at zero and a local martingale is unique up to indistinguishability with respect to the probability* $\mathbb{P}$*.*

Note that the integrand in the stochastic integral in (B.2.30) can also be written as $\widehat{X}_t(y) - X_{t-}$ where $\widehat{X}_t(y)$ is the value of X_t obtained by retaining the behaviour of μ on $[0,t)$ and pretending that a jump with mark y occurs at time t. In special cases for Markov processes the form of U_t can be explicitly given, cf. Section 2.2.2. In particular, then we can also drop the assumption that the compensator is continuous.

B.2.6 Intensity measures

In certain special cases it is possible to state the compensators using ordinary Lebesgue integrals. Particularly, this is possible in the case that the Markov kernels $(P^n)_{n\geq 0}$ are absolutely continuous with respect to the Lebesgue measure.

Definition B.2.8. Let $\mathbb{P}$ be a probability on $(H,\mathcal{H})$ with compensating measure $\widetilde{N}$ and let $C\in\mathcal{E}$. A predictable process $(\Lambda^C_t)_{t\geq 0}$ with $\Lambda^C_t \geq 0$ for all $t\geq 0$ is an *intensity process* for the counting measure $(N^C_t)_{t\geq 0}$ under $\mathbb{P}$ if it holds for all $t\geq 0$ that

$$\widetilde{N}^C_t = \int_{[0,t]} \Lambda^C_s \,\mathrm{d}s \quad \text{a.s.}$$

The collection $\{(\Lambda^C_t)_{t\geq 0}, C\in\mathcal{E}\}$ is an *intensity measure* for $\mathbb{P}$ if each $(\Lambda^C_t)_{t\geq 0}$ is an intensity process and the mapping $C\mapsto \Lambda^C_t$ is almost surely a non-negative measure on $(E,\mathcal{E})$ for all t.

Definition B.2.9. Let κ be a non-negative, σ-finite measure on $(E,\mathcal{E})$ and the map $(m,t,y)\mapsto \Lambda_t(y,m)$ be measurable and such that $\{(\Lambda_t(y))_{t\geq 0},\, y\in E\}$ is a collection of nonnegative, predictable processes. Here, as always, we omit the dependency of a random function on the argument m. Then, Λ is called a κ*-intensity process* for $\mathbb{P}$, if it holds for all $t\geq 0$ and all $C\in\mathcal{E}$ that

$$\widetilde{N}^C_t = \int_{[0,t]}\int_C \Lambda_s(y)\,\kappa(\mathrm{d}y)\,\mathrm{d}s \quad \text{a.s.}$$

or, equivalently, for all $A\in\mathcal{B}\otimes\mathcal{E}$ that

$$\widetilde{N}(A) = \int_{[0,\infty)}\int_E \mathbb{I}_A(s,y)\ \Lambda_s(y)\,\kappa(\mathrm{d}y)\,\mathrm{d}s \quad \text{a.s.}$$

Note that a κ-intensity process is predictable. Further, the existence of a κ-intensity process stipulates that the measure $\widetilde{N}$ has a density with respect to the product measure, i.e., $\widetilde{N}(\mathrm{d}t, \mathrm{d}y) = \Lambda_t(y)\,\kappa(\mathrm{d}y)\,\mathrm{d}t$ and clearly

$$\Lambda_t^C = \int_C \Lambda_t(y)\,\kappa(\mathrm{d}y)$$

is an intensity process if Λ is a κ-intensity process. Note that a κ-intensity process defines the compensator $\widetilde{N}$ uniquely which in turn defines the measure $\mathbb{P}$ uniquely. Further, regarding the stochastic integrals, if Λ is a κ-intensity process then

$$\widetilde{N}_t^g = \int_{(0,t]} \int_E g(s,y)\,\Lambda_t(y)\,\kappa(\mathrm{d}y)\,\mathrm{d}t\,.$$

Proposition B.2.6. ([64, Prop. 4.4.1(b)]) *Let $\mathbb{P}$ be a measure on $(H, \mathcal{H})$ which is generated by the Markov kernels $(P^n)_{n\geq 0}$ and $(\pi^n)_{n\geq 1}$. Further, let the measure $\mathbb{P}^n_{z_n} := P^n(z_n, \cdot)$ be absolutely continuous with respect to the Lebesgue measure with hazard rate $u^n_{z_n}(t)$ almost surely for every $n \geq 0$, cf. Example B.3.1 in Section B.3. Then $\{(\Lambda_t^C)_{t\geq 0}, C \in \mathcal{E}\}$ is an intensity measure for $\mathbb{P}$ where*

$$\Lambda_t := u^{N_{t-}}_{Z_{t-}}(t)\,\pi^{N_{t-}}_{Z_{t-},t}(C)\,.$$

If in addition it holds that there exists a non-negative, σ-finite measure κ on $(E, \mathcal{E})$ such that almost surely for every n and for Lebesgue almost every t the measures $\pi^n_{Z_n,t}$ are absolutely continuous with respect to κ with a density $p^n_{Z_n}$ such that the map $(z_n, t, y) \mapsto p^n_{z_n}(t)$ is measurable, then Λ is a κ-intensity for $\mathbb{P}$ where

$$\Lambda_t(y) := u^{N_{t-}}_{Z_{t-}}(t)\,p^{N_{t-}}_{Z_{t-}}(t,y)\,. \tag{B.2.31}$$

In the literature another concept of intensities for counting processes is widespread, the *instantaneous jump rate*, which is in general not necessarily equivalent to the intensity process Λ. Let $\mathbb{P}$ be a measure on $(H, \mathcal{H})$, let $A \in \mathcal{E}$ and each N_t^C has the intensity process Λ_t^C. Moreover we assume for the intensity process that almost surely all limits from the right exist and we define the *right intensity process (instantaneous jump rate)* by

$$\Lambda_{t+}^C := \lim_{h\downarrow 0} \Lambda_{t+h}^C\,.$$

Note that by definition intensity processes Λ^C are predictable, however, defined as right limits, instantaneous jump rates are right-continuous and in this sense not necessarily predictable. In case of the intensity process being continuous it coincides with its right intensity. Instantaneous jump rates are commonly used in chemical reaction kinetics, in which case intensities are piecewise constant, and, e.g., in [121] in the context of numerical methods for General Stochastic Hybrid Systems where intensities are continuously changing. The terminology instantenous jump rate for right intensities is motivated by the following observation.

Proposition B.2.7. ([64, Prop. 4.4.2(b)]) *Under the above assumption it holds for all $t \geq 0$ almost surely that*

$$\Lambda^C_{t+} = \lim_{h \downarrow 0} \frac{1}{h} \mathbb{P}\big[N^C_{t+h} - N^C_t \geq 1\,, Y_{1,t} \in C \,\big|\, \mathcal{H}_t\big]$$

where $Y_{1,t}$ denotes the mark of the first event after time t.

B.2.7 Random counting measures on arbitrary probability spaces

So far definitions and results have been presented for random counting measures or, equivalently, marked point processes on the canonical space $(H, \mathcal{H})$ or $(K, \mathcal{K})$. However, as with random variables in general, one usually prefers to define random counting measures on arbitrary filtered probability spaces taking values in the canonical space. Therefore an essential step is to provide the connection of results derived on canonical spaces to random measures defined on arbitrary probability spaces with respect to an arbitrary right-continuous filtration.
Thus, in the following let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ be an arbitrary filtered probability space with a (for now) unspecified right-continuous filtration and μ be a random counting measure defined thereon with marks in a Borel space $(E, \mathcal{E})$. That is, μ is an $(H, \mathcal{H})$-valued random variable defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Note that in contrast to the notation in the preceding sections of this chapter $\mathbb{P}$ now denotes a probability on the measurable space $(\Omega, \mathcal{F})$ and NOT on the measurable space $(H, \mathcal{H})$ of counting measures. (However, $\mathbb{P}$ also defines a probability on $(H, \mathcal{H})$ by the pushfoward measure with respect to the random variable μ, see below.) Then we define for every $C \in \mathcal{E}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ the counting process $N^C_t := \mu([0, t] \times C)$, i.e., each N^C_t is a integer-valued random variable in the sense that $N^C_t = \mu^{[0,T]\times C} \circ \mu$ with μ^A as in (B.2.4). We always assume that the filtration $(\mathcal{F}_t)_{t\geq 0}$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is such that these processes are adapted, i.e., for all $C \in \mathcal{E}$ and all $t \geq 0$ the random variable N^C_t is $\mathcal{F}_t$-measurable. Further, a marked point process $((\tau_n, Y_n))_{n\geq 1}$ is defined on $(\Omega, \mathcal{F}, \mathbb{P})$ by the usual identification of random counting measures and marked point processes. It holds that each τ_n is an $\mathcal{F}_t$-stopping time and each Y_n is $\mathcal{F}_{\tau_n}$-measurable.

A certain special role is assigned to the filtration on $(\Omega, \mathcal{F}, \mathbb{P})$ generated by the random counting measure which we denote by $(\mathcal{F}^\mu_t)_{t\geq 0}$. That is, in analogy to the definition of the canonical filtration on $(H, \mathcal{H})$, cf. (B.2.8), $(\mathcal{F}^\mu_t)_{t\geq 0}$ is generated by all counting processes $(N_t^C)_{t\geq 0}$, $C \in \mathcal{E}$, i.e.,

$$\mathcal{F}_t^\mu := \sigma(N_s^C; s \leq t, C \in \mathcal{E}). \tag{B.2.32}$$

By assumption we have $\mathcal{F}_t^\mu \subset \mathcal{F}_t$ for all $t \geq 0$ and, most importantly, it holds that for all $t \geq 0$, cf. [64, p. 95],

$$\mathcal{F}_t^\mu = \mu^{-1}(\mathcal{H}_t), \tag{B.2.33}$$

i.e., $(\mathcal{F}_t^\mu)_{t\geq 0}$ is the preimage of the canonical filtration $(\mathcal{H}_t)_{t\geq 0}$ with respect to the measurable function μ. Furthermore, it holds that $\mathcal{F}_\infty^\mu := \bigvee_{t\geq 0} \mathcal{F}_t^\mu = \mu^{-1}(\mathcal{H})$. On the space $(H, \mathcal{H})$ a probability $\overline{\mathbb{P}}$ is defined by the pushforward measure, that is, $\overline{\mathbb{P}}(A) := \mathbb{P}[\mu \in A]$ for all $A \in \mathcal{H}$, and we write $\widetilde{N}^*$ for the compensating measure of $\overline{\mathbb{P}}$ in the sense of Definition B.2.20. Then, all the results in the preceding sections for canonical random counting measures carry over to results about the random counting measure μ, the filtration $(\mathcal{F}_t^\mu)_{t\geq 0}$ and the positive random compensating measure $\widetilde{N} := \widetilde{N}^* \circ \mu$ as we are in the exact same situation as for the canonical space. According to Jacobsen [64] the proofs in this case differ from their corresponding results on the canonical space only by an application of the transformation theorem for integrals.

Nevertheless, the aim of this section is to define the compensator directly on an arbitrary, filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ supporting an adapted random measure μ, where the filtration is NOT necessarily given by the filtration generated by μ. The reason, why it is insatisfactory for PDMP theory to deal with stochastic processes adapted to the filtration $(\mathcal{F}_t^\mu)$ is the following: The initial σ-field $\mathcal{F}_0^\mu$ of the filtration $(\mathcal{F}_t^\mu)_{t\geq 0}$ is the trival σ-field $\{\emptyset, \Omega\}$. Therefore, any process $(X_t)_{t\geq 0}$ adapted to $(\mathcal{F}_t^\mu)_{t\geq 0}$ satisfies necessarily $X_0 \equiv$ *constant*. Hence, if we want to account for processes with non-constant initial condition the filtration has to be enlarged.

Definition B.2.10. The $\mathcal{F}_t$*-compensating measure for* μ is the positive random measure $\widetilde{N}$ with $\widetilde{N}(\{0\} \times E) = 0$ almost surely such that $\widetilde{N}_t^C := \widetilde{N}([0,t] \times C)$ defines for all $C \in \mathcal{E}$ an $\mathcal{F}_t$-predictable, right continuous process, which is necessarily increasing, such that the process

$$M_t^C := N_t^C - \widetilde{N}_t^C$$

is an $\mathcal{F}_t$-local martingale.

Clearly, a special case of Definition B.2.10 is the $\mathcal{F}_t^\mu$-compensator which corresponds to the compensator on the canonical space, i.e., $\widetilde{N} = \widetilde{N}^* \circ \mu$. Further, we note that the existence of the compensators $\widetilde{N}_t^C$ in Definition B.2.10 is guaranteed for all $C \in \mathcal{E}$ by the Doob-Meyer decomposition, see, e.g., [66, Sect. I.3b]. Each compensator $(\widetilde{N}_t^C)_{t\geq 0}$ is unique up to indistinguishability and as $(E, \mathcal{E})$ is a Borel space also the $\mathcal{F}_t$-compensating measure $\widetilde{N}$ always exists, cf. [64, p. 96].

In general compensators as defined in Definition B.2.10 depend on the filtration and, in particular, may differ for different filtrations [64]. From the considerations on the canonical space we know that the $\mathcal{F}_t^\mu$-generator determines the distribution of μ, cf. Theorem B.2.4. However, it is not generally true that the $\mathcal{F}_t$-compensator determines the distribution of μ, much less does it determine $\mathbb{P}$ as the distribution of μ is only a marginal distribution of the probability $\mathbb{P}$ on $(\Omega, \mathcal{F})$, cf [64, p. 96]. However, in a special case, presented in the following theorem, the $\mathcal{F}_t$-compensator defines the distribution of μ uniquely.

Theorem B.2.8. ([64, Thm. 4.8.1]) *Let μ be a random counting measure on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ with $\mathcal{F}_t$-compensating measure $\widetilde{N}$. If $\widetilde{N}$ is $\mathcal{F}_t^\mu$-predictable, then $\widetilde{N} = \widetilde{N}^* \circ \mu$ up to indistinguishability where $\widetilde{N}^*$ is the compensator of the pushforward measure in the sense of* (B.2.20).

To conclude the section, we elaborate a bit further on processes and random counting measures where we consider a special structure of the σ-field $\mathcal{F}$ and the filtration $(\mathcal{F}_t)_{t\geq 0}$. In particular, it corresponds to the structure for filtrations assumed in connection with the definition of PDMPs in Section 2.1. Assume that $\mathcal{F}^0$ is a σ-field on Ω. Then we impose for the filtration $(\mathcal{F}_t)_{t\geq 0}$ the structure

$$\mathcal{F}_t := \mathcal{F}^0 \vee \mathcal{F}_t^\mu. \tag{B.2.34}$$

This special structure of a filtration is regularly discussed in connection with stochastic processes, e.g., PDMPs, defined via random counting measures, see, Jacobsen [64, p. 100], Brémaud [21, Chapt. VIII] and Jacod [65]. As discussed before in order to consider sufficiently rich processes the filtration $(\mathcal{F}_t)_{t\geq 0}$ needs to be finer than the filtration generated by μ. The above definition (B.2.34) on the one hand accomplishes this feat but also, on the other hand, keeps the filtration as coarse as possible. Thus, the main idea in considering filtrations of the form $\mathcal{F}^0 \vee \mathcal{F}_t^\mu$ is that the σ-field $\mathcal{F}^0$ which equals the initial σ-field in the filtration accommodates the randomness of the initial condition, i.e., X_0 is $\mathcal{F}^0$-measurable, see Section 2.1. Without loss of generality we may further assume that $\mathcal{F} = \mathcal{F}^0 \vee \bigvee_{t\geq 0} \mathcal{F}_t^\mu$.
Then the $\mathcal{F}_t$-compensating measure is given by the compensating measure with respect to the conditional probability $\mathbb{P}[\,\cdot\,|\mathcal{F}^0]$ [64, p. 100]. Thus, if we assume that[8] $\mathcal{F}^0 = \sigma(X_0)$ – which is altogether sufficient for the purpose of this thesis – the conditional probability given the event $[X_0 = x_0]$ defines a push-forward measure $\mathbb{P}^{0,X_0}$ on $(H, \mathcal{H})$. Thus, assume that for each 'initial condition' $x_0 \in E$ a family Markov kernels $(P_{x_0}^n)_{n\geq 0}$ and $(\pi_{x_0}^n)_{n\geq 0}$ is given such that the measures $\mathbb{P}^{0,x_0}$ on $(H, \mathcal{H})$ defined due to

[8] The same holds for $\mathcal{F}^0 = \sigma(X_0) \vee \mathcal{F}^*$ where $\mathcal{F}^*$ is a σ-field which is independent of the initial condition X_0 and the random counting measure μ. In this way we can accommodate further independent random variables on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Theorem B.2.1 equal the conditional push-forward measures of the random counting measures μ given the event $[X_0 = x_0]$. Then we obtain that the $\mathcal{F}_t$-compensator conditional on $[X_0 = x_0]$ is given by the $\mathcal{F}_t^{\mu_{x_0}}$-compensator where μ_{x_0} is the random counting measure corresponding to the Markov kernels $(P^n_{x_0})_{n\geq 0}$ and $(\pi^n_{x_0})_{n\geq 0}$. Hence, the results of the preceding sections on the canonical space do not only extend to arbitrary spaces and the filtration $(\mathcal{F}_t^{\mu})_{t\geq 0}$ but also to filtrations with the special structure (B.2.34).

Finally, the following proposition further elucidates the structure of the special filtration $(\mathcal{F}_t)_{t\geq 0}$ and moreover the connection to the approach to PDMPs in [68] which we discussed in Section B.2.2 remains valid for random counting measures defined on arbitrary probability spaces.

Proposition B.2.8. *The filtration $(\mathcal{F}_t)_{t\geq 0}$ is a jumping filtration with jumping sequence $(\tau_n)_{n\geq 0}$ and it is the minimal filtration such that for all $n \geq 1$ the jump time τ_n is a stopping time and the stopped σ-field contains $\mathcal{F}^0 \vee \sigma(Z_n)$.*

We note that it can be shown analogously to [65, Prop. 3.40] that it actually holds $\mathcal{F}_{\tau_n} = \mathcal{F}^0 \vee \sigma(Z_n)$. Further, Proposition B.2.8 also holds for the filtration $(\mathcal{F}_t^{\mu})_{t\geq 0}$ which is seen by choosing $\mathcal{F}^0 = \{\emptyset, \Omega\}$.

Proof. Throughout the proof we employ the notation $\mathcal{Z}_n := \mathcal{F}^0 \vee \sigma(Z_n) = \mathcal{F}^0 \vee \sigma(\tau_n, Y_m; m \leq n)$, $\mathcal{Z}_\infty := \bigvee_{n\geq 1} \mathcal{Z}_n$ and $\tau_\infty := \lim_{n\to\infty} \tau_n$. The proof is structured as follows. We, first propose (a) a family of σ-fields $\mathcal{K}_t$ on Ω and show that it constitutes a filtration. Moreover, (b) this filtration is the smallest filtration such that for all $n \geq 1$ the jump time τ_n is a stopping time and the stopped σ-field contains $\mathcal{Z}_n$. Furthermore, we show that $(\mathcal{K}_t)_{t\geq 0}$ satisfies the condition for it to be a jumping filtration (B.2.17). The proof is completed showing (c) that the filtration $(\mathcal{K}_t)_{t\geq 0}$ coincides with the filtration $(\mathcal{F}_t)_{t\geq 0}$ which we show (d) to be right-continuous. We note that parts (a) and (b) of the proof are based on the proof of an analogous result in [65, Prop. 3.38, 3.40] extended and adapted to the presentation in this thesis.

(a) Let $\mathcal{K}_t$ denote the collection of subsets of Ω such that $A \in \mathcal{K}_t$ if it has the form

$$A = \Big(\bigcup_{n\geq 1} \big(A_n \cap [\tau_n \leq t < \tau_{n+1}]\big)\Big) \cup \big(A_\infty \cap [\tau_\infty \leq t]\big) \tag{B.2.35}$$

for sets $A_n \in \mathcal{Z}_n$ for all $n \geq 1$ and $n = \infty$. Note that the right hand side of (B.2.35) is a countable disjoint union. We provide a second equivalent characterisation of the elements of $\mathcal{K}_t$. $\mathcal{K}_t$ consists of those sets $A \in \mathcal{Z}_\infty$ such that there exists for each $n \geq 1$ a set $A_n \in \mathcal{Z}_n$ and

$$A_n \cap [t < \tau_{n+1}] = A \cap [t < \tau_{n+1}]\,. \tag{B.2.36}$$

On the one hand, it is clear that $\mathcal{K}_t \subset \mathcal{Z}_\infty$ as the right hand side (B.2.35) is a countable union of elements of $\mathcal{Z}_\infty$ and it is obvious that each $A \in \mathcal{K}_t$ satisfies (B.2.36). On the other hand, let $A \in \mathcal{Z}_\infty$ satisfy (B.2.36), then by an intersection of both sides of equality (B.2.36) with the set $[\tau_n \leq t]$ we obtain

$$A_n \cap [\tau_n \leq t < \tau_{n+1}] = A \cap [\tau_t \leq t < \tau_{n+1}] \,.$$

Thus taking the union over all $n \geq 1$ yields

$$\begin{aligned}\bigcup_{n\geq 1}\big(A_n \cap [\tau_t \leq t < \tau_{n+1}]\big) &= \bigcup_{n\geq 1}\big(A \cap [\tau_n \leq t < \tau_{n+1}]\big) \\ &= A \cap \bigcup_{n\geq 1}[\tau_n \leq t < \tau_{n+1}] \;=\; A \cap [\tau_\infty \leq t]^c \,.\end{aligned}$$

Finally, we take the union of both sides with $A \cap [\tau_\infty \leq t]$ and obtain

$$\bigcup_{n\geq 1}\big(A_n \cap [\tau_t \leq t < \tau_{n+1}]\big) \cup \big(A \cap [\tau_\infty \leq t]\big) = A \,.$$

Thus A is of the form (B.2.35) as $A \in \mathcal{Z}_\infty$.

Using the characterisation (B.2.36) it is easy to see that each $\mathcal{K}_t$ is a σ-field. Firstly, it is non-empty as $A = \Omega$ satisfies (B.2.36) with $A_n = \Omega$ for all $n \geq 1$ and similarly we find that $\mathcal{F}^0 \subset \mathcal{K}_t$ for all $t \geq 0$. Secondly, for A^c with A satisfying (B.2.36) we obtain

$$\begin{aligned}A^c \cap [t < \tau_{n+1}] &= [t < \tau_{n+1}]\backslash A \\ &= [t < \tau_{n+1}]\backslash(A \cap [t < \tau_{n+1}]) \;=\; [t < \tau_{n+1}]\backslash(A_n \cap [t < \tau_{n+1}]) \\ &= \big([t < \tau_{n+1}]\backslash A_n\big) \cup \big([t < \tau_{n+1}]\backslash[t < \tau_{n+1}]\big) \;=\; A_n^c \cap [t < \tau_{n+1}] \,.\end{aligned}$$

Thus the complement A^c also satisfies (B.2.36). Thirdly, let A^i, $i \geq 1$ satisfy (B.2.36) then we obtain

$$\begin{aligned}\Big(\bigcup_{i\geq 1} A^i\Big) \cap [t < \tau_{n+1}] &= \Big(\bigcup_{i\geq 1} A^i \cap [t < \tau_{n+1}]\Big) \\ &= \Big(\bigcup_{i\geq 1} A_n^i \cap [t < \tau_{n+1}]\Big) \;=\; \Big(\bigcup_{i\geq 1} A_n^i\Big) \cap [t < \tau_{n+1}]\end{aligned}$$

where each $\bigcup_{i\geq 1} A_n^i$ is in $\mathcal{Z}_n$. Therefore we have shown that $\mathcal{K}_t$ is a σ-field for each $t \geq 0$.

It remains to show that the family of σ-fields $(\mathcal{K}_t)_{t\geq 0}$ is a filtration. For a set $A \in \mathcal{K}_t$ taking the intersection on both sides of equality (B.2.36) with $[s \leq \tau_{n+1}]$ yields

$$A \cap [s < \tau_{n+1}] = A_n \cap [s < \tau_{n+1}] \quad \text{for } s \geq t\,.$$

Thus $A \in \mathcal{K}_s$ and hence $\mathcal{K}_t \subset \mathcal{K}_s$ for all $t \leq s$.

(b) Next we show that each τ_n is a stopping time with respect to the filtration $(\mathcal{K}_t)_{t\geq 0}$ and that the stopped σ-fields $\mathcal{K}_{\tau_n}$ contain $\mathcal{Z}_n$. For the former we note that for all $n \geq 0$

$$[\tau_n \leq t] \cap [t < \tau_m] = \begin{cases} \emptyset & \text{for } m \leq n \\ [\tau_n \leq t] \cap [t < \tau_m] & \text{for } n < m\,. \end{cases}$$

As each $\mathcal{Z}_m$ contains the empty set and $[\tau_n \leq t] \in \mathcal{Z}_n \subset \mathcal{Z}_m$, $m > n$, it follows that $[\tau_n \leq t] \in \mathcal{K}_t$ for all $t \geq 0$ and thus all jumping times τ_n are $\mathcal{K}_t$-stopping times. Next we consider the set $[Y_n \in B] \cap [\tau_n \leq t]$ for some $B \in \mathcal{E}$ which yields

$$[Y_n \in B] \cap [\tau_n \leq t] \cap [t < \tau_m] = \begin{cases} \emptyset & \text{for } m \leq n \\ [Y_n \in B] \cap [\tau_n \leq t] \cap [t < \tau_m] & \text{for } n < m\,. \end{cases}$$

As before this implies that $[Y_n \in B] \cap [\tau_n \leq t] \in \mathcal{K}_t$ for all $t \geq 0$. Hence by definition of the stopped σ-field we obtain that $[Y_n \in B] \in \mathcal{K}_{\tau_n}$. Finally, it is clear that $\mathcal{K}_{\tau_n}$ contains $\mathcal{F}^0$ as a sub-σ-field for all $n \geq 1$. Thus, as $\mathcal{Z}_n$ is the minimal σ-field containing $\mathcal{F}^0$, such that τ_m and Y_m, $m \leq n$, are measurable, it follows that $\mathcal{Z}_n \subset \mathcal{K}_{\tau_n}$.

It is now easy to see that $(\mathcal{K}_t)_{t\geq 0}$ satisfies condition (B.2.17) for it to be a jumping filtration. Assume, on the one hand, that $A \in \mathcal{K}_t$ then by (B.2.35)

$$A \cap [\tau_n \leq t < \tau_{n+1}] = A_n \cap [\tau_n \leq t < \tau_{n+1}]$$

for some $A_n \in \mathcal{Z}_n \subset \mathcal{K}_{\tau_n}$. On the other hand, assume $A \in \mathcal{K}_{\tau_n}$, then

$$A \cap [\tau_n \leq t < \tau_{n+1}] = A \cap [\tau_n \leq t] \cap [\tau_n \leq t < \tau_{n+1}]$$

where $A \cap [\tau_n \leq t] \in \mathcal{K}_t$ by definition of the stopped σ-field.

Next, we assume that $(\mathcal{K}_t^{\min})_{t\geq 0}$ is the smallest filtration such that each τ_n is a stopping time and the stopped σ-field contains $\mathcal{Z}_n$, $n \geq 1$. We show that this filtration coincides with $(\mathcal{K}_t)_{t\geq 0}$. To this end we consider the characterisation of the elements of $\mathcal{K}_t$ given by (B.2.35), i.e., for now $A_n \in \mathcal{Z}_n$. As τ_n is a stopping time with respect to the minimal filtration and the stopped σ-field contains $\mathcal{Z}_n$ it holds that

$$A_n \cap [\tau \le t] \in \mathcal{K}_t^{\min} \quad \text{and} \quad [t < \tau_{n+1}] = [\tau_{n+1} \le t]^c \in \mathcal{K}_t^{\min}.$$

Hence, $A_n \cap [\tau_n \le t < \tau_{n+1}] = A_n \cap [\tau_n \le t] \cap [t < \tau_{n+1}] \in \mathcal{K}_t^{\min}$. Further, also $A_\infty \cap [\tau_\infty \le t] \in \mathcal{K}_t^{\min}$ as τ_∞ is a $\mathcal{K}_t^{\min}$-stopping time by definition as the pointwise limit of stopping times and $\mathcal{Z}_\infty$ is contained in its stopped σ-field[9]. Thus by the characterisation (B.2.35) it follows that $A \in \mathcal{K}_t$ implies $A \in \mathcal{K}_t^{\min}$. As the filtration $(\mathcal{K}_t)_{t\ge 0}$ has the same properties for which $(\mathcal{K}_t^{\min})_{t\ge 0}$ is assumed to be minimal it follows that

$$\mathcal{K}_t^{\min} = \mathcal{K}_t \quad \forall\, t \ge 0\,.$$

(c) Finally we need to connect the filtration $(\mathcal{K}_t)_{t\ge 0}$ to the filtration $(\mathcal{F}_t)_{t\ge 0}$ as defined by (B.2.34). Due to [64, p. 94] we have that the jump times τ_n are $\mathcal{F}_t$-stopping times and the stopped σ-fields $\mathcal{F}_{\tau_n}$ contain $\mathcal{Z}_n$. Hence, the filtration $(\mathcal{F}_t)_{t\ge 0}$ satisfies the conditions $(\mathcal{K}_t)_{t\ge 0}$ is minimal for. Therefore to conclude that the two filtrations coincide it remains to show that the counting processes generating the filtration $(\mathcal{F}_t^{\mu})_{t\ge 0}$ are adapted to the filtration $(\mathcal{K}_t)_{t\ge 0}$, i.e., $(\mathcal{K}_t)_{t\ge 0}$ satisfies the conditions $(\mathcal{F}_t)_{t\ge 0}$ is assumed to be minimal for.

Recall that each random variable N_t^C, $t \ge 0$ and $C \in \mathcal{E}$ is given by

$$N_t^C = \sum_{n=1}^{\infty} \mathbb{I}_{[Y_n \in C]}\, \mathbb{I}_{[\tau_n \le t]}\,. \tag{B.2.37}$$

First note that each summand is a random variable

$$\omega \mapsto \mathbb{I}_C(Y_n(\omega)\, \mathbb{I}_{(0,t]}(\tau_n(\omega)),$$

which is measurable with respect to some σ-field on Ω if it contains the set

$$\{\omega \in \Omega | Y_n(\omega) \in C, \tau(\omega) \le t\} = [Y_n \in C] \cap [\tau_n \le t]\,.$$

We have already shown that this set is contained in $\mathcal{K}_t$ for all $t \ge 0$.
For the $\mathbb{N}_0 \cup \{\infty\}$-valued random variable (B.2.37) we obtain for each $n \ge 0$ the preimage

$$(N_t^C)^{-1}(\{n\}) = [\tau_n \le t] \cap \Bigg(\bigcup_{m \ge n} \Big[[N_t = m] \cap \Big(\big[n \text{ out of } m \text{ marks } Y_i \text{ are in } C \big] \Big) \Big] \Bigg)\,. \tag{B.2.38}$$

This set is an intersection of $[\tau_n \le t] \in \mathcal{K}_t$ with a countable union. Hence, the

[9] It holds that $\tau_\infty \ge \tau_n$ for all $n \ge 1$ and thus $\mathcal{K}_{\tau_n}^{\min} \subset \mathcal{K}_{\tau_\infty}^{\min}$. This in turn implies that $\mathcal{K}_{\tau_\infty}^{\min}$ contains all $\mathcal{Z}_n$, $n \ge 1$. As $\mathcal{Z}_\infty$ is the smallest σ-field containing all $\mathcal{Z}_n$ it follows that $\mathcal{Z}_\infty \subset \mathcal{K}_{\tau_\infty}^{\min}$.

preimage is an element of $\mathcal{K}_t$ if each set in the union is in $\mathcal{K}_t$. Obviously $[N_t = m] = [\tau_m \le t] \cap [t \le \tau_{m+1}] \in \mathcal{K}_t$ as τ_m are $\mathcal{K}_t$-stopping times. Thus it remains to show that $\mathcal{K}_t$ also contains its intersection with the set '$\big[n$ out of m marks Y_i are in $C\big]$'. This set in turn is a finite union of finite intersections of sets of the form $[Y_i \in C]$ and $[Y_i \in C]^c$ with $i \le m$ which are, if intersected with sets $[\tau_m \le t]$ elements of $\mathcal{K}_t$. Finally, note that also the preimage of $\{\infty\}$ is contained in $\mathcal{K}_t$ as $(N_t^C)^{-1}(\{\infty\}) = (N_t^C)^{-1}(\mathbb{N}_0^c) = ((N_t^C)^{-1}(\mathbb{N}_0))^c$ which is the complement of a countable union of elements of $\mathcal{K}_t$, hence an element of $\mathcal{K}_t$ itself.
Hence, we obtain that each N_s^C is $\mathcal{K}_t$ measurable for $s \le t$, $C \in \mathcal{E}$. As $\mathcal{F}_t$ is defined as the minimal σ-field containing $\mathcal{F}^0$ such that this holds it follows that

$$\mathcal{F}_t = \mathcal{K}_t \quad \forall\, t \ge 0\,.$$

(d) We conclude the proof showing that the filtration $(\mathcal{F}_t^\mu)_{t\ge 0}$ is right-continuous which immediately implies that $(\mathcal{F}_t)_{t\ge 0}$ is right-continuous. Assume $A \in \bigcap_{\epsilon>0} \mathcal{F}_{t+\epsilon}^\mu$ then there exists for all $\epsilon > 0$ a $B_\epsilon \in \mathcal{H}_{t+\epsilon}$ such that $A = \mu^{-1}(B_\epsilon)$. In particular, we can choose $B = B_\epsilon$ for all $\epsilon > 0$. Thus, $B \in \mathcal{H}_t$ as the canonical filtration on $(H, \mathcal{H})$ is right-continuous, see Theorem B.2.3(b). This implies that $A = \mu^{-1}(B) \in \mathcal{F}_t^\mu$ due to (B.2.33). Hence, $\mathcal{F}_t^\mu \subseteq \bigcap_{\epsilon>0} \mathcal{F}_{t+\epsilon}^\mu \subseteq \mathcal{F}_t^\mu$ and therefore the filtration is right-continuous.

In particular, it now follows that $(\mathcal{F}_t)_{t\ge 0}$ is a jumping filtration in the sense of Definition B.2.5.

□

B.3 Survivor functions and hazard measures

In this Section $\mathbb{P}$ always denotes a probability distribution on the closed positive reals $\mathbb{R}_+ \cup \{\infty\}$ with cumulative distribution function F, i.e., $F(t) = \mathbb{P}([0,t])$ for all $0 \le t \le \infty$. Hence, F is a non-decreasing, right-continuous function.

Definition B.3.1. The *survivor function* $S : \mathbb{R}_+ \cup \{\infty\} \to [0,1]$ for the probability $\mathbb{P}$ is the function defined by

$$S(t) \;:=\; 1 - F(t) \;=\; \mathbb{P}((t,\infty]) \tag{B.3.1}$$

and thus necessarily non-increasing and right-continuous. In this thesis we always assume that $S(0) = 1$.

The terminology 'survivor function' stems from reliability theory. That is, if the probability $\mathbb{P}$ is the distribution of the time of some event happening, e.g., some machine or part thereof breaking down ('dies'), then $S(t)$ denotes the probability that this happens after time t, i.e., the machine 'survives' until time t. Clearly there

is a one-to-one correspondence between a probability $\mathbb{P}$, its cumulative distribution function F and its survivor function S. Further, the assumption $S(0) = 1$ is equivalent to $F(0) = \mathbb{P}(\{0\}) = 0$, that is, the event does not happen instantaneously almost surely but only after some positive time.
We denote the left limit of the survivor function at time t by $S(t-)$, i.e.,

$$\lim_{\epsilon \downarrow 0} S(t-\epsilon) =: S(t-) = \mathbb{P}([t,\infty]) = S(t) + \mathbb{P}(\{t\}) .$$

Further, we define the *termination time* $t^\dagger$ for $\mathbb{P}$ by

$$t^\dagger := \inf\{t > 0 : S(t) = 0\} = \inf\{t > 0 : F(t) = 1\}. \tag{B.3.2}$$

Thus the probability $\mathbb{P}$ has no mass after the termination time, i.e., if $\mathbb{P}$ defines the distribution of the time some event is happening, then this event occurs almost surely before or at time $t^\dagger$. As we always assume that $S(0) = 1$ it follows that $t^\dagger > 0$ due to the right-continuity of the survivor function. Moreover, one of the following two cases holds:

$$\text{(I)} \quad t^\dagger = \infty \quad \Leftrightarrow \quad S(t) > 0 \quad \forall\, t > 0$$

$$\text{(II)} \quad t^\dagger < \infty \quad \Rightarrow \quad S(t^\dagger) = 0$$

with either

$$\text{(IIa)} \qquad \mathbb{P}(\{t^\dagger\}) > 0 \text{ and } S(t^\dagger-) > 0,$$

$$\text{(IIb)} \qquad \mathbb{P}(\{t^\dagger\}) = 0 \text{ and } S(t^\dagger - \epsilon) > 0 \quad \forall\, 0 < \epsilon < t^\dagger.$$

Case (I) corresponds to a distribution which has probability mass on all of $(0,\infty]$ with point probabilities possible (also at $\{\infty\}$). Case (II) corresponds to distributions that are concentrated on $(0,t^\dagger]$ and have no mass afterwards, i.e., $S(t^\dagger) = \mathbb{P}((t^\dagger,\infty]) = 0$. The sublasses (IIa) and (IIb) differ in the behaviour of the probability $\mathbb{P}$ at the terminal point. In case (IIa) there is a discrete, positive mass at the terminal point, whereas in case (IIb) there is not and there is mass between $t^\dagger$ and every point before, i.e., $S(t^\dagger - \epsilon) = \mathbb{P}((t^\dagger - \epsilon,\infty]) = \mathbb{P}((t^\dagger - \epsilon, t^\dagger)) > 0$ for all $\epsilon \in (0,t^\dagger)$.

Definition B.3.2. The *hazard measure* for $\mathbb{P}$ is the $[0,\infty]$-valued (not necessarily σ-finite) measure ν on $(0,\infty)$ which is absolutely continuous with respect to $\mathbb{P}$ and for $t \geq 0$ defined by the Radon-Nikodym derivative

$$\frac{\mathrm{d}\nu}{\mathrm{d}\mathbb{P}}(t) = \begin{cases} \dfrac{1}{S(t-)} & \text{if } S(t-) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Due to the absolute continuity with respect to $\mathbb{P}$ it holds that $\nu((t^\dagger, \infty)) = 0$ and thus the Radon-Nikodym derivative is well defined as $\mathbb{P}(\{t \geq 0 : S(t-) = 0\}) = 0$ because in case (I) this probability corresponds to $\mathbb{P}(\emptyset)$, in case (IIa) this probability corresponds to $\mathbb{P}((t^\dagger, \infty)) = 0$ and in case (IIb) this probability corresponds to $\mathbb{P}([t, \infty]) = \mathbb{P}(\{t^\dagger\}) + \mathbb{P}((t^\dagger, \infty)) = 0$. Therefore the hazard measure is well defined for all Borel sets B on $(0, \infty)$ and by definition

$$\nu(B) = \int_B \frac{1}{S(t-)}\, \mathbb{P}(\mathrm{d}t) = \int_B \frac{1}{S(t-)}\, \mathrm{d}F(t) = -\int_B \frac{1}{S(t-)}\, \mathrm{d}S(t).$$

In the following theorem we collect general properties of hazard measures.

Theorem B.3.1. ([64, Thm. 4.1.1])

(a) *If $t < t^\dagger$, then $\nu((0, t]) < \infty$.*

(b) *For all $t \in (0, \infty)$ it holds that $\nu(\{t\}) \leq 1$ and $\nu(\{t\}) < 1$ for all $t < t^\dagger$.*

(c) *If $t^\dagger < \infty$, then $\nu(\{t^\dagger\}) = 1$ if and only if $\mathbb{P}(\{t^\dagger\}) > 0$.*

(d) *If $\mathbb{P}(\{t^\dagger\}) > 0$, then $\nu((0, t^\dagger]) < \infty$ if $t^\dagger < \infty$ and $\nu((0, \infty)) < \infty$ if $t^\dagger = \infty$.*

(e) *If $\mathbb{P}(\{t^\dagger\}) = 0$, then $\nu((0, t^\dagger)) = \infty$ (wether or not $t^\dagger$ is finite).*

We conclude the section with the discussion of the special cases of classes of survivor functions and hazard measures of importance for this thesis.

Example B.3.1. Suppose that $\mathbb{P}$ has a density f with respect to the Lebesgue measure, which in particular implies that there is no atom except, possibly, $\{\infty\}$ with positive probability. Then ν has the density $u(t) = f(t)/S(t-)$ if $t < t^\dagger$ and this function u is called the *hazard rate* of $\mathbb{P}$. Assume, in addition, that f and thus also u is continuous on $(0, t^\dagger)$, then for $t < t^\dagger$

$$u(t) = -\frac{\mathrm{d}}{\mathrm{d}t} \log S(t)$$

and conversely since $S(0) = 1$

$$S(t) = \exp\Big(-\int_0^t u(s)\, \mathrm{d}s\Big).$$

Example B.3.2. A second example is a hazard measure for a probability which is absolutely continuous with respect to the Lebesgue measure on $[0, t^\dagger)$ with a finite termination point, i.e., $t^\dagger < \infty$ with positive probability. Thus the cumulative distribution function is continuous and non-decreasing on the interval $[0, t^\dagger)$ with a jump of positive height $\mathbb{P}(\{t^\dagger\}) = 1 - F(t^\dagger-)$ at $t^\dagger$ and $F(t) = 1$ on $[t^\dagger, \infty)$. Therefore the hazard measure is given by

$$\nu(B) = \int_B \frac{1}{S(t-)}\,\mathrm{d}F(t) = \int_{B\cap(0,t^\dagger)} \frac{1}{S(t-)}\,\mathrm{d}F(t) + \frac{1}{S(t^\dagger-)}\mathbb{P}(\{t^\dagger\})\,\mathbb{I}_B(t^\dagger).$$

As $S(t^\dagger-) = S(t^\dagger) + \mathbb{P}(\{t^\dagger\}) = \mathbb{P}(\{t^\dagger\})$ we obtain

$$\nu(B) = \int_{B\cap(0,t^\dagger)} \frac{1}{S(t-)}\,\mathrm{d}F(t) + \mathbb{I}_B(t^\dagger) = \int_{B\cap(0,t^\dagger)} \frac{f(t)}{S(t-)}\,\mathrm{d}t + \mathbb{I}_B(t^\dagger)$$

where $f(t)$ is the density of $\mathbb{P}$ with respect to the Lebesgue measure on $(0, t^\dagger)$, cf. Theorem B.3.1(c). Hence, finally this yields for the hazard rate $u(t)$

$$\nu(B) = \int_{B\cap(0,t^\dagger)} u(t)\,\mathrm{d}t + \mathbb{I}_B(t^\dagger).$$

with

$$S(t) = \mathbb{I}_{[0,t^\dagger)}(t)\,\exp\Big(-\int_0^t u(s)\,\mathrm{d}s\Big)$$

where $\int_0^{t^\dagger} u(t)\mathrm{d}t < \infty$.

Appendix C

Linear Parabolic Partial Differential Equations

The examples of partial differential equations considered in this thesis arise as models of excitable media which are always parabolic equations. On the one hand they generate the dynamical system that gives the evolution in between jumps of the PDMP model and, on the other hand, act as the macroscopic limits for the stochastic models. Therefore, we briefly present the relevant existence theory for those. Moreover, the specific equations we encounter in this thesis are all linear and thus linear existence theory is sufficient. When we consider the nonlinear system of an excitable media equation in Section 3.3 we prove existence using a fixed point argument building on linear theory.

Let D be a bounded domain in $\mathbb{R}^d$, $d \geq 1$. The problem we have in mind to solve is of the following type: We look for a function $u : (0,T) \times D \to \mathbb{R}$ which satisfies the initial-boundary value problem

$$\begin{aligned} u_t &= A(t)u + f(t) \quad \text{on } D \times [0,T] \\ u(0) &= u_0 \quad \text{on } D \\ u(t) &= 0 \quad \text{on } \partial D\,. \end{aligned} \tag{C.0.1}$$

Here, the inhomogeneous term is $f \in L^2((0,t), H^{-1}(D))$ and the linear operators $A(t)$ are second order elliptic operators, which, in divergence form, are given by

$$A(t)u := \sum_{i,j=1}^{d} (a_{ij}(t,x)u_{x_i})_{x_j} + \sum_{i=1}^{d} b_i(t,x)u_{x_i} + c(t,x)u\,. \tag{C.0.2}$$

The measurable coefficient functions a_{ij}, b_i, c are assumed to be essentially bounded in the space-time domain, i.e., elements of $L^\infty((0,T) \times D)$. We say a function

$u \in L^2((0,T), H_0^1(D)) \cap H^1((0,T), H_0^{-1}(D))$ is a *weak solution* to the Dirichlet problem (C.0.1) if for almost all $t \in (0,T)$ it holds that

$$\langle \dot{u}, v \rangle_{H^1} = \langle A(t)u, v \rangle_{H^1} + \langle f(t), v \rangle_{H^1} \qquad \forall\, v \in H_0^1(D)$$

and the initial value satisfies $u(0) = u_0$ in $L^2(D)$. As the solution space implies that $u \in C^0([0,T], L^2(D))$, the initial condition is well-defined.

Existence theory for equations of the type (C.0.1) is based on the general theory of linear first-order evolution equations. For a detailed exposition we refer to Zeidler [122, Chap. 23] but elements of the theory are also contained in [43, 103, 104, 116]. To align with the abstract theory we choose the evolution triple $H_0^1(D) \subset L^2(D) \subset H^{-1}(D)$. We note that the particular choice of $H_0^1(D)$ reflects the boundary condition in (C.0.1) which requires solutions that vanish at the boundary, i.e., *Dirichlet boundary conditions*. A bilinear form $a(u,v,t)$ associated with the operators given in (C.0.2) is defined by

$$a(u,v,t) := \int_D \sum_{i,j=1}^{d} a_{ij}(t,x) u_{x_i} v_{x_j}\, \mathrm{d}x - \int_D \sum_{i=1}^{d} b_i(t,x) u_{x_i} v - c(t,x) uv\, \mathrm{d}x\,, \qquad \text{(C.0.3)}$$

which is obviously well-defined for all $u, v \in H_0^1(D)$. Note that the domain of the operator $A(t)$ in $L^2(D)$ is given by $D(A) := H_0^1(D) \cap H^2(D)$ for any $t \in [0,T]$. As we assume that the coefficients are essentially bounded the bilinear form (C.0.3) is bounded in $H_0^1(D)$ in u, v for all $t \geq 0$ due to the Cauchy-Schwarz inequality. A central condition for the existence of a solution is the coercivity of the bilinear form a, i.e., for all $u \in H_0^1(D)$ and all $t \geq 0$

$$a(u,u,t) \;\; \geq \;\; \gamma_1 \|u\|^2_{H^1(D)} - \gamma_2 \|u\|^2_{L^2(D)} \qquad \text{(C.0.4)}$$

for appropriate constants γ_1, $\gamma_2 > 0$. In order for (C.0.4) to hold it is, in addition to the boundedness of the coefficients of the operator $A(t)$, sufficient to assume *uniform ellipticity*[1] for the linear operators $A(t)$. That is, the coefficient functions a_{ij} are such that there exists a constant $\xi > 0$ such that for all $(t,x) \in (0,T) \times D$ it holds that

$$\sum_{i,j=1}^{d} a_{ij}(x,t)\, y_i y_j \geq \xi\, |y|^2 \quad \forall\, y \in \mathbb{R}^d. \qquad \text{(C.0.5)}$$

The following theorem provides an initial result on the well-posedness of the problem and is an immediate consequence of the corresponding result on abstract equations,

[1] To be more precise, according to the terminology in [52], property (C.0.5) with $\xi > 0$ states the operator is *strictly elliptic*. However, as we assume $a(t,x)$ to be bounded there exists an estimate from above analogous to (C.0.5). This implies the operator is uniformly elliptic.

cf. [116, 122]. For a direct proof of the following theorem we refer to [43, Chap. 7, Thms. 3, 4].

Theorem C.0.2. *Assume that the coefficients of the operator* (C.0.2) *are almost everywhere bounded and* (C.0.5) *holds. Then there exists for every initial condition* $u_0 \in L^2(D)$ *and data* $f \in L^2((0,T), H_0^{-1}(D))$ *a unique weak solution*

$$u \in L^2((0,T), H_0^1(D)) \cap H^1((0,T), H_0^{-1}(D)) .$$

This implies that $u \in C^0([0,T], H_0^1(D))$*. Furthermore, the solution satisfies the estimates*

$$\|u\|_{L^2((0,T),H^1)} + \|\dot{u}\|_{L^2((0,T),H^{-1})} \leq C_1\big(\|f\|_{L^2((0,T),H^{-1})} + \|u_0\|_{L^2}\big) ,$$
$$\|u\|_{C^0([0,T],L^2)} \leq C_2\Big(\|u\|_{L^2((0,T),H^1)} + \|\dot{u}\|_{L^2((0,T),H^{-1})}\Big)$$

where the constant C_1 *depends only on the time horizon* T*, the domain* D *and the* $L^\infty((0,T)\times D)$*-norm of the coefficients of* A *and* C_2 *depends only on* T *and* D*. Hence, the solution depends continuously on the initial condition.*

Further, analogously to the case of abstract evolution equations it is possible to obtain improved regularity results under additional assumptions on the operators $A(t)$, the inhomogeneous term b and the initial condition u_0, cf. [43, Chap. 7]. Thus we assume for the following that the coefficients of $A(t)$ are independent of time and twice continuously differentiable on $\overline{D}$. Further, the domain D satisfies $\partial D \in C^2$. The inclusion of solutions into continuous spaces in the following theorem and the appropriate estimates are due to [43, Chap. 5.9, Thm. 3, 4].

Theorem C.0.3. (a) *For initial conditions* $u_0 \in H_0^1(D)$ *and an inhomogeneous term* $f \in L^2((0,T), L^2(D))$ *the solution satisfies in addition*

$$u \in L^2((0,T), H^2(D)) \cap L^\infty((0,T), H_0^1(D)),$$
$$\dot{u} \in L^2((0,T), L^2(D)),$$

which implies $u \in C^0([0,T], H_0^1(D))$*. Furthermore, the solution satisfies the estimate*

$$\|u\|_{L^2((0,T),H^2)} + \|u\|_{L^\infty((0,T),H_0^1)} + \|\dot{u}\|_{L^2((0,T),L^2)} \leq C_3\Big(\|f\|_{L^2((0,T),L^2)} + \|u_0\|_{H_0^1}\Big),$$
$$\|u\|_{C^0([0,T],H^1)} \leq C_4\Big(\|u\|_{L^2((0,T),H^2)} + \|\dot{u}\|_{L^2((0,T),L^2)}\Big), \tag{C.0.6}$$

where the constant C_3 *depends only on* T*,* D*, the coefficients* a_{ij} *and the* $L^\infty((0,T)\times D)$*-norm of the coefficients* b_i*,* c*, and* C_4 *depends only on* T *and* D*.*

(b) *For initial conditions* $u_0 \in H_0^1(D) \cap H^2(D)$ *and an inhomogeneous term* $f \in H^1((0,T), L^2(D))$ *the solution satisfies*

$$u \in L^\infty((0,T), H^2(D)),$$

$$\dot{u} \in L^2((0,T), H_0^1(D)),\ \ddot{u} \in L^2((0,T), H^{-1}(D)),$$

which implies $\dot{u} \in C^0([0,T], L^2(D))$ *and thus* $u \in C^1([0,T], L^2)$. *Furthermore, the solution satisfies the estimate*

$$\begin{aligned} \|u\|_{L^2((0,T),H^2)} + \|u\|_{L^\infty((0,T),H^2)} + \|\dot{u}\|_{L^2((0,T),H^1)} + \|\ddot{u}\|_{L^2((0,T),H^{-1})} \\ \leq C_5\Big(\|f\|_{H^1((0,T),L^2)} + \|u_0\|_{H^2}\Big), \\ \|\dot{u}\|_{C^0([0,T],L^2)} \leq C_6\Big(\|\dot{u}\|_{L^2((0,T),H^1)} + \|\ddot{u}\|_{L^2((0,T),H^{-1})}\Big), \end{aligned} \tag{C.0.7}$$

where the constant C_5 *depends only on* T, D, *the coefficients* a_{ij} *and the* $L^\infty((0,T) \times D)$*-norm of the coefficients* b_i, c, *and* C_6 *depends only on* T *and* D.

Remark C.0.1. For the application in the proof of the existence of a solution to general excitable media equations, cf. Section 3.3.1, it is necessary to extend the results in Theorem C.0.3 to time-dependent zeroth order coefficients $c = c(t)$. We note that the result in (a) remains valid for time-dependent coefficients b_i, c. For the results in part (b) we restrict the spatial dimension to $d \leq 3$ and assume that the zeroth order term in (C.0.3) is time dependent, i.e., $c = c(t)$. In addition to the assumptions above $c \in H^1((0,T), L^2)$ and $\dot{c} \in L^4((0,T), L^4)$. Then the first extension is that the assertion of Theorem C.0.3 (b) remains valid where the right hand side in estimate (C.0.7) is given by

$$C_5^*\Big(\|f\|_{H^1((0,T),L^2)} + \|u_0\|_{H^2} + C_3^2\big(\|f\|_{L^2((0,T),L^2)} + \|u_0\|_{H_0^1}\big)^2\Big).$$

Here the constant C_5^* depends in addition to the dependencies of the constant C_5 also on $\|\dot{c}\|_{L^4((0,T),L^4)}$. If in addition $\dot{c} \in L^\infty((0,T) \times D)$ then the estimate (C.0.7) is also satisfied for the right hand side

$$C_5^{**}\Big(\|f\|_{H^1((0,T),L^2)} + \|u_0\|_{H^2}\Big)$$

where the constant C_5^{**} depends in addition to the dependencies of the constant C_5 also on $\|\dot{c}\|_{L^\infty((0,T)\times D)}$. These extensions follow immediately using the same method of proof as employed in [43, Chap. 7,Thm 5]. By a close inspection of the proof we find that under the above assumptions the crucial estimates therein can still be derived with only slight modifications of the estimation procedures incorporating the changed assumptions.

Secondly, if in addition ∂D is C^3, $c \in L^2((0,T),H^2) \cap L^\infty((0,T),H^1)$, whereas the remaining coefficients a_{ij}, b_i, which are independent of time, are in $C^2(\overline{D})$ and $f \in L^2((0,T),H^1)$, then it follows that

$$u \in L^2((0,T),H^3)\,. \tag{C.0.8}$$

Thus in combination with the result in Theorem C.0.3 (b) it holds due to [43, Chap. 5.9, Thm. 4] that $u \in C^0([0,T],H^2)$. Particularly, it holds that

$$\|u\|_{L^2((0,T),H^3)} \leq C_7\Big(\|f\|_{L^2((0,T),H^1)} + \|f\|_{H^1((0,T),L^2)} + \|u_0\|_{H^2}\Big)\,, \tag{C.0.9}$$

and

$$\|u\|_{C^0([0,T],H^2)} \leq C_8\Big(\|u\|_{L^2((0,T),H^3)} + \|\dot{u}\|_{L^2((0,T),H^1)}\Big) \tag{C.0.10}$$

where the constant C_7 depends on T, D and the $L^\infty((0,T)\times D)$-norm of the coefficients A, the $L^\infty((0,T)\times D)$-norm of the first order spatial derivatives of a_{ij}, b_j and on $\|c\|_{L^\infty((0,T),H^1)}$ and $\|\dot{c}\|_{L^\infty((0,T),\times D)}$. The constant C_8 depends on T and D only.

In order to establish (C.0.8) note that a close inspection of the arguments used to derive the elliptic regularity results [43, Chap. 6,Thms. 1,2,4,5] shows that $u(t)$, which solves for almost all $t \in [0,T]$ the elliptic equation $-A(t)u(t) = f(t) - \dot{u}(t)$, satisfies for almost all $t \in [0,T]$

$$\|u(t)\|_{H^3} \leq C'\big(\|f(t) - \dot{u}(t)\|_{H^1} + \|u(t)\|_{L^2} + C_5\|D_{x_i}c(t)\|_{L^2}\|u\|_{L^\infty((0,T),H^2)}\big)\,,$$

where the constant C' depends on the $L^\infty((0,T)\times D)$-norm of the coefficients of a_{ij}, b_j and their first order spatial derivatives as well as the $L^\infty((0,T)\times D)$-norm of c. Finally, integration over $(0,T)$ and the estimates (C.0.7) yield (C.0.8).

To conclude we emphasise that for the constants C_1–C_8 in the above two theorems and this remark the dependency on the time horizon T is monotonically increasing for increasing T, but, most importantly, the constants are finite for any finite T.

Bibliography

[1] G. Acosta and R. G. Durán. An optimal Poincaré inequality in L_1 for convex domains. *Proc. Amer. Math. Soc.*, 132(1):195–202, 2004.

[2] R. A. Adams and J. F. J. Fournier. *Sobolev Spaces, 2nd Ed.* Academic Press, New York, 2003.

[3] A. Alfonsi, E. Cancés, G. Turinci, B. Di Ventura, and W. Huisinga. Adaptive hybrid simulation of hybrid stochastic and deterministic models for biochemical reactions. *ESAIM Proc.*, 14:1–13, 2005.

[4] C. D. Aliprantis and K. C. Border. *Infinite Dimensional Analysis, A Hitchhikers Guide, 3rd Ed.* Springer, Berlin, 2006.

[5] L. Arnold. *Stochastische Differentialgleichungen.* Oldenbourg, München, 1973.

[6] L. Arnold and M. Theodosopulu. Deterministic limit of the stochastic model of chemical reactions with diffusion. *Adv. Appl. Prob.*, 12(2):367–379, 1980.

[7] T. D. Austin. The emergence of the deterministic Hodgkin-Huxley equations as a limit from the underlying stochastic ion-channel mechanism. *Ann. Appl. Prob.*, 18(4):1279–1325, 2006.

[8] H. Bauer. *Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie, 3. Aufl.* De Gruyter, Berlin, 1978.

[9] J. Bect. *Processus de Markov diffusif par morceaux: outils analytiques et numériques.* PhD thesis, Université Paris Sud XI, 2007.

[10] J. Bect. A unifying formulation of the Fokker-Planck-Kolmogorov equation for general stochastic hybrid systems. *Nonlinear Analysis: Hybrid Systems*, 4(2):357–370, 2010.

[11] A. Bellen and M. Zennaro. *Numerical Methods for Delay Differential Equations.* Oxford University Press, Oxford, 2003.

[12] M. R. Bennett, W. Gibson, and J. Robinson. Probabilistic secretion of quanta and the synaptosecretosome hypothesis: Evoked release at active zones of varicosities, boutons and endplates. *Biophys. J.*, 73(4):1815–1829, 1997.

[13] D. P. Bertsekas and S. E. Shreve. *Stochastic Optimal Control: The Discrete Time Case.* Academic Press, New York, 1978.

[14] P. Billingsley. *Convergence of Probability Measures, 2nd Ed.* Wiley, New York, 1999.

[15] D. Blount. Comparison of stochastic and deterministic models of a linear chemical reaction with diffusion. *Ann. Probab.*, 19(4):1440–1462, 1991.

[16] D. Blount. Law of large numbers in the supremum norm for a chemical reaction with diffusion. *Ann. Appl. Probab.*, 2(1):131–141, 1992.

[17] D. Blount. Limit theorems for a sequence of nonlinear reaction-diffusion systems. *Stoch. Proc. Appl.*, 45(2):193–207, 1993.

[18] D. Blount. Density-dependent limits for a nonlinear reaction-diffusion model. *Ann. Probab.*, 22(4):2040–2070, 1994.

[19] D. Blount. Diffusion limits for a nonlinear density dependent space-time population model. *Ann. Probab.*, 24(2):639–659, 1996.

[20] R. M. Blumenthal and R. K. Getoor. *Markov Processes and Potential Theory.* Academic Press, New York, 1968.

[21] P. Brémaud. *Point processes and queues: Martingale Dynamics*. Springer, New York, 1981.

[22] P. C. Bressloff. Stochastic neural field theory and the system-size expansion. *SIAM J. Appl. Math.*, 70:1488–1521, 2009.

[23] P. C. Bressloff. Metastable states and quasicycles in a stochastic Wilson-Cowan model of neuronal population dynamics. OCCAM Pre-print 10/61, 2010.

[24] I. C. Bruce. Implementation issues in approximate methods for stochastic Hodgkin-Huxley models. *Ann. Biomed. Eng.*, 35(2):315–318, 2007.

[25] E. Buckwar, P. E. Kloeden, and M. G. Riedler. The numerical stability of stochastic differential equations with additive noise. *Stochastics and Dynamics*, 11(2-3):265–281, 2011.

[26] E. Buckwar and M. G. Riedler. Stochastic Runge-Kutta methods for Itô stochastic differential equations of jump type. *J. Comp. Appl. Math.*, in press, 2011.

[27] E. Buckwar and M. G. Riedler. Exact modelling of neuronal membranes including spatio-temportal evolution. *J. Math. Bio.*, published online, Jan. 18, 2011.

[28] M. L. Bujorianu and J. Lygeros. Toward a general theory of stochastic hybrid systems. In H. A. P. Blom and J. Lygeros, editors, *Stochastic Hybrid Systems*, pages 3–30. Springer, Berlin, 2006.

[29] H. Cai. Piecewise Deterministic Markov Processes. *Stoch. Anal. Appl.*, 11(3):255–274, 1993.

[30] D. Chafaï, F. Malrieu, and K. Paroux. On the long time behavior of the tcp window size process. *Stoch. Proc. Appl.*, 120(8):1518–1534, 2010.

[31] C. C. Chow and White J. A. Spontaneous action potentials due to channel fluctuations. *Biophys. J.*, 71(6):3013–3021, 1996.

[32] J. R. Clay and L. J. DeFelice. Relationship between membrane excitability and single channel open-close kinetics. *Biophys. J.*, 42(2):151–157, 1983.

[33] O. L. V. Costa. Stationary distributions for Piecewise-Deterministic Markov Processes. *J. Appl. Prob.*, 27(60-73), 1990.

[34] A. Crudu, A. Debussche, A. Muller, and O. Radulescu. Convergence of stochastic gene networks to hybrid piecewise deterministic processes. arXiv:1101.1431v1, 2011.

[35] G. Da Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimensions*. Cambridge University Press, Cambridge, 1992.

[36] M. H. A. Davis. Piecewise-deterministic Markov processes: a general class of non-diffusion stochastic models (with discussion). *J. R. Stat. Soc., Ser. B*, 46:353–388, 1984.

[37] M. H. A. Davis. *Markov Models and Optimisation*. Chapman and Hall, London, 1993.

[38] L. J. DeFelice and A. Isaac. Chaotic states in a random world: Relationship between the nonlinear differential equations of excitability and the stochastic properties of ion channels. *J. Stat. Phys.*, 70(1-2):339–354, 1993.

[39] A. D. Dorval and J. A. White. Channel noise is essential for perithreshold oscillations in entorhinal stellate neurons. *J. Neuroscience*, 26(43):10025–10028, 2005.

[40] C. Dufour and O. L. V. Costa. Stability of Piecewise Deterministic Markov Processes. *SIAM J. Control Optim.*, 37(5):1483–1502, 1999.

[41] S. N. Ethier and T. G. Kurtz. *Markov Processes: Characterization and Convergence*. Wiley, New York, 1986.

[42] J. Evans and N. Shenk. Solutions to axon equations. *Biophys. J.*, 10:1090–1101, 1970.

[43] L. C. Evans. *Partial Differential Equations, 2nd Ed.* AMS, Rhode Island, 2010.

[44] A. A. Faisal. Stochastic simulation of neurons, axons and action potentials. In G. Lord and C. Laing, editors, *Stochastic Methods in Neuroscience*, pages 297–343. Oxford University Press, Oxford, 2010.

[45] A. A. Faisal and S. B. Laughlin. Stochastic simulations on the reliability of action potential propagation in thin axons. *PLoS Comp. Biol.*, 3(5):e79, 2007.

[46] A. A. Faisal, L. P. J. Selen, and D. M. Wolpert. Noise in the nervous system. *Nature Reviews Neuroscience*, 9(4):292–303, 2008.

[47] M. Falcke. On the role of stochastic channel behaviour in intracellular Ca^{2+} dynamics. *Biophys. J.*, 84:42–56, 2002.

[48] F. H. Fenton and E. M. Cherry. Models of cardiac cell. *Scholarpedia*, 8(3):1868, 2008.

[49] F. Fontbona, H. Guérin, and F Malrieu. Quantitative estimates for the long time behavior of a PDMP describing the movement of bacteria. arXiv:1006.0982v1, 2010.

[50] R. F. Fox and Y. Lu. Emergent collective behaviour in large numbers of globally coupled independently stochastic ion channels. *Phys. Rev. E*, 49(4):3421–3431, 1994.

[51] J. E. Gentle. *Random Number Generation and Monte-Carlo Methods, 2nd Ed.* Springer, New York, 2003.

[52] N. Gilbarg and N. S. Trudinger. *Elliptic Partial Differential Equations of Second Order, 2nd Ed.* Springer, Berlin, 1983.

[53] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.*, 22:403–434, 1976.

[54] J. R. Groff, H. DeRemigio, and G. D. Smith. Markov chain models of ion channels and calcium release sites. In G. Lord and C. Laing, editors, *Stochastic Methods in Neuroscience*, pages 29–64. Oxford University Press, Oxford, 2010.

[55] L. Grüne and O. Junge. *Gewöhnliche Differentialgleichungen*. Vieweg Teubner, Wiesbaden, 2009.

[56] M. R. Guevara and T. J. Lewis. A minimal single-channel model for the regulatory of beating in the sinoatrial node. *Chaos*, 5(1):174–183, 1995.

[57] E. Hairer, S. P. Nørsett, and G. Wanner. *Solving Ordinary Differential Equations I, 2nd Ed.* Springer, Berlin, 2000.

[58] E. L. Haseltine and J. B. Rawlings. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *J. Chem. Phys.*, 117(15):6959–6969, 2002.

[59] R. Z. Hasḿinskiĭ. *Stochastic Stability of Differential Equations*. Sijthoff & Noordhoff, Alphen aan den Rijn, 1980.

[60] J. P. Hespanha. A model for stochastic hybrid systems with application to communication networks. *Nonlinear Anal.*, 62(8):1353–1383, 2005.

[61] H. Heuser. *Funktionalanalysis, 2nd Ed.* Teubner, Stuttgart, 1986.

[62] A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.*, 117:500–544, 1952.

[63] N. S. Imennov and J. T. Rubinstein. Effects of electrode-to-fiber distance on temporal neural response with electrical stimulation. *IEEE Trans. Biomed. Eng.*, 56(10):2493–2501, 2009.

[64] M. Jacobsen. *Point Process Theory and Applications: Marked Point and Piecewise Deterministic Processes*. Birkhäuser, Boston, 2006.

[65] J. Jacod. *Calcul Stochastique et Problèmes de Martingales*. Springer, New York, 1979.

[66] J. Jacod and A. N. Shiryaev. *Limit Theorems for Stochastic Processes*. Springer, Berlin, 1987.

[67] J. Jacod and A. V. Skorokhod. Jumping filtrations and martingales with finite variation. In *Sem. Proba. 28*, pages 21–35. Springer, Berlin, 1994.

[68] J. Jacod and A. V. Skorokhod. Jumping Markov processes. *Ann. Inst. Henri Poincaré, Probab. Stat.*, 32(1):11–67, 1996.

[69] A. Joffe and M. Métivier. Weak convergence of sequences of semimartingales with applications to multitype branching processes. *Adv. Appl. Prob.*, 18:20–65, 1986.

[70] G. Kalantzis. Hybrid stochastic simulations of intracellular reaction-diffusion systems. *Comp. Bio. Chem*, 33:205–215, 2009.

[71] G. Kallianpur and J. Xiong. Diffusion approximation of nuclear space-valued stochastic differential equations driven by Poisson random measures. *Ann. Appl. Prob.*, 5(2):493–517, 1995.

[72] P. E. Kloeden and E. Platen. *The Numerical Solution of Stochastic Differential Equations*. Springer, Berlin, 1992.

[73] P. E. Kloeden and M. Rasmussen. *Nonautonomous Dynamical Systems*. AMS, Rhode Island, forthcoming, 2011.

[74] C. Koch. *Biophysics of Computation*. Oxford University Press, New York, 1999.

[75] P. Kotelenez. Law of large numbers and central limit theorem for linear chemical reactions with diffusion. *Ann. Prob.*, 14(1):173–193, 1986.

[76] P. Kotelenez. Linear parabolic differential equations as limits of space-time jump Markov processes. *J. Math. Anal. Appl.*, 116(1):42–76, 1986.

[77] P. Kotelenez. Fluctuations near homogeneous states of chemical reactions with diffusion. *Adv. Appl. Probab.*, 19(2):352–370, 1987.

[78] P. Kotelenez. High density limit theorems for nonlinear chemical reactions with diffusion. *Prob. Th. Rel. Fields*, 78(1):11–37, 1988.

[79] T. G. Kurtz. Solutions of ordinary differential equations as limits of pure jump Markov processes. *J. Appl. Prob.*, 7:49–58, 1970.

[80] T. G. Kurtz. Limit theorems for a sequence of jump Markov processes approximating ordinary differential equations. *J. Appl. Prob.*, 8:344–356, 1971.

[81] L. Lamberti. Solutions to the Hodgkin-Huxley equations. *Appl. Math. Comput.*, 18:43–70, 1986.

[82] H. Li, Y. Cao, L. R. Petzold, and D. T. Gillespie. Algorithms and software for stochastic simulation of biochemical reacting systems. *Biotech. Progress*, 24(1):56–62, 2008.

[83] T. Lipniackia, P. Paszekb, A. Marciniak-Czochrad, A. R. Brasiere, and M. Kimmela. Transcriptional stochasticity in gene expression. *J. Theor. Biol.*, 238:348–367, 2006.

[84] K. Liu. *Stability of Infinite Dimensional Stochastic Differential Equations.* Chapman & Hall/CRC, Boca Raton, 2006.

[85] J. Lygeros, K. Koutroumpas, S. Dimopoulos, I. Legouras, P. Kouretas, C. Heichinger, P. Nurse, and Z. Lygerou. Stochastic hybrid modeling of DNA replication across a complete genome. *Proceedings of the National Academy of Science*, 105(34):12295–12300, 2008.

[86] X. Mao and C. Yuan. *Stochastic Differential Equations with Markovian Switching.* Imperial College Press, London, 2006.

[87] M. Mascagni. An initial-boundary value problem of physiological significance for equations of nerve conduction. *Comm. Pure Appl. Math.*, 42(2):213–227, 1989.

[88] A. J. Matsuoka, J. T. Rubinstein, P. J. Abbas, and C. A. Miller. The effects of interpulse interval on stochastic properties of electrical stimulation: Models and measurements. *IEEE Trans. Biomed. Eng.*, 48(4):416–424, 2001.

[89] B. Mélykúti, K. Burrage, and K. C. Zygalakis. Fast stochastic simulation of biochemical reaction systems by alternative formulations of the chemical Langevin equation. *J. Chem. Phys.*, 132:164109, 2010.

[90] M. Métivier. *Semimartingales.* deGruyter, Berlin, 1982.

[91] M. Métivier. Convergence faible et principe d'invariance pour des martingales à valeurs dans des espaces de Sobolev. *Ann. Inst. Henri Poincaré, Sec. B*, 20(4):329–348, 1984.

[92] G. N. Milstein and M. Tretyakov. *Stochastic Numerics for Mathematical Physics.* Springer, Berlin, 2004.

[93] H. Mino, J. T. Rubinstein, and J. White. Comparison of algorithms for the simulation of action potential with stochastic sodium channels. *Ann. Biomed. Eng.*, 30:578–587, 2002.

[94] H. Mino, Rubinstein J. T., C. A. Miller, and P. J. Abbas. Effects of electrode-to-fiber distance on temporal neural response with electrical stimulation. *IEEE Trans. Biomed. Eng.*, 52(1):13–20, 2004.

[95] K. Pakdaman, M. Thieullen, and G. Wainrib. Diffusion approximation of birth-death processes: comparison in terms of large deviations and exit points. *Statist. Probab. Lett.*, 80(13-14):1121–1127, 2010.

[96] K. Pakdaman, M. Thieullen, and G. Wainrib. Fluid limit theorems for stochastic hybrid systems with application to neuron models. *Adv. in Appl. Probab.*, 42(3):761–794, 2010.

[97] K. R. Parthasarathy. *Probability Measures on Metric Spaces*. Academic Press, New York, 1967.

[98] L. E. Payne and H. F. Weinberger. An optimal Poincaré inequality for convex domains. *Arch. Rat. Mech. Anal.*, 5:286–292, 1960.

[99] E. Platen and N. Bruti-Liberati. Strong approximations of stochastic differential equations with jumps. *J. Comput. Appl. Math.*, 205(2):982–1001, 2007.

[100] C. Prévôt and M. Roeckner. *A Concise Course on Stochastic Partial Differential Equations*. Springer, Berlin, 2007.

[101] J. W. Prüss and M. Wilke. *Gewöhnliche Differentialgleichungen und dynamische Systeme*. Birkhäuser, Basel, 2010.

[102] J. Rauch. Global existence for the Fitzhugh-Nagumo equations. *Comm. Partial Differential Equations*, 1(6):609–621, 1976.

[103] M. Renardy and R. C. Rogers. *Introduction to Partial Differential Equations, 2nd Ed.* Springer, New York, 2004.

[104] J. C. Robinson. *Inifinite-dimensional Dynamical Systems*. Cambridge University Press, Cambridge, 2001.

[105] J. T. Rubinstein. Threshold fluctuations in an N sodium channel model of the Node of Ranvier. *Biophys. J.*, 68:779–785, 1995.

[106] S. Rüdiger, J. W. Shuai, W. Huisinga, C. Nagaiah, G. Warnecke, I. Parker, and M. Falcke. Hybrid stochastic and deterministic simulations of calcium blips. *Biophys. J.*, 93:1847–1857, 2007.

[107] H. Salis and Y. Kaznessis. Accurate hybrid stochastic simulation of coupled chemical or biochemical reactions. *J. Chem. Phys.*, 122, 2005.

[108] L. F. Shampine and S. Thompson. Event location for ordinary differential equations. *Comput. Math. Appl.*, 39(5-6):43–54, 2000.

[109] M. Sharpe. *General Theory of Markov Processes*. Academic Press, San Diego, 1988.

[110] E. Skaugen and L. Walløe. Firing behaviour in a stochastic nerve membrane model based upon the Hodgkin-Huxley equations. *Acta Physiol. Scand.*, 107(4):343–363, 1979.

[111] J. Smoller. *Shock Waves and Reaction-Diffusion Equations*. Springer, New York, 1983.

[112] S. Stolze. Stochastic equations in Hilbert space with Lévy noise and their applications in finance. Master's thesis, Universität Bielefeld, 2005.

[113] D. W. Stroock and S. R. S. Varadhan. *Multidimensional Diffusion Processes*. Springer, Berlin, 1979.

[114] A. M. Stuart and A. R. Humphries. *Dynamical Systems and Numerical Analysis*. Cambridge University Press, Cambridge, 1996.

[115] S. Swillens, P. Campeil, L. Combettes, and G. Dupont. Stochastic simulation of a single inositol 1,4,5-triphosphate-sensitive Ca^{2+} channel reveals repetitive openings during 'blip-like' Ca^{2+} transients. *Cell Calcium*, 23(5):291–302, 1998.

[116] R. Temam. *Infinite-dimensional Dynamical Systems in Mechanics and Physics*. Springer, New York, 1988.

[117] H. C. Tuckwell. Stochastic equations for nerve membrane potential. *J. Theoret. Neurobiol.*, 5:87–99, 1986.

[118] N. G. van Kampen. *Stochastic Processes in Physics and Chemistry*. North-Holland, Amsterdam, 1981.

[119] G. Wainrib. *Randomness in neurons : a multiscale probabilistic analysis*. PhD thesis, Ecole polytechnique, Paris, 2010.

[120] J. J. Westman and F. B. Hanson. State dependent jump models in optimal control. *Proc. 38th IEEE Conf. on Decision and Control*, pages 2378–2384, 1999.

[121] G. Yin, X. Mao, C. Yuan, and D. Cao. Approximation methods for hybrid diffusion systems with state-dependent switching processes: Numerical algorithms and existence and uniqueness of solutions. *SIAM J. Math. Anal.*, 41:2335–2352, 2010.

[122] E. Zeidler. *Nonlinear Functional Analysis and its Applications, Vol. 2*. Springer, New York, 1990.

[123] S. Zeisler, U. Franz, O. Wittich, and V. Liebscher. Simulation of genetic networks modelled by piecewise deterministic Markov processes. *IET Syst. Bio.*, 2(3):113–135, 2008.