Conservation of Hamiltonians and Invariants in Stochastic Differential Equations

Sjur Svorkmo Bergmann

January 21, 2022

Preface

This work is dedicated to my brother, Trym Bergmann. Hopefully, you won't make the same mistakes as I did.

Acknowledgements

I would like to thank my supervisor Anne Kværnø for her considerable patience, outstanding response time and always valuable feedback and comments. Another thank you goes to Brynjulf Owren, whose lectures in geometric integration were challenging, but rewarding. Finally, I am eternally grateful that NTNU Campus Gløshaugen is available to students around the clock.

Abstract

Geometric integration in Stochastic Differential Equations has become a dynamic and developing field of study during the last two decades. In this thesis, two main cases are investigeted: the conservation of (quadratic) invariants in Stratonovich Stochastic Differential Equations, based on [1], and the preservation of drift in expectation for deterministic Hamiltonians with additive noise, based on [2]. For the first case, conditions for which two classes of Stochastic Runge-Kutta methods preserve quadratic invariants are proven to hold, before methods of these classes are applied to four test problems, comparing performance with the classic Euler-Maruyama and Milstein methods. For the second case, theoretical results on drift of Hamiltonian functions in Itô Stochastic Differential Equations are shown to be true in the Stratonovich sense as well. A composition method is then presented for which [2] demonstrates relatively good conservation properties. Test problems with the structure considered are presented in Section 5.2, But not successfully implemented. Lastly, some comments are made.

Summary

Structure-preservation has become an important field of inquiry in *Ordinary Differential Equations*(ODEs). From around the turn of the millenium until today, a significant body of research has been done in geometric integration of *Stochastic Differential Equations*(SDEs), adapting and expanding upon the results for ODEs; a short review is found in Chapter 1.

This thesis focuses mainly on two issues: conservation of quadratic invariants in Stratonovich SDEs and preservation of drift in deterministic Hamiltonians with additive noise. Before new theory is introced in Chapter 3, a preliminary into the foundations of SDEs and geometric integration is given in Chapter 2.

Based a paper by Hong, Xu and Wang [1], after offering definitions of stochastic invariants in Section 3.1, two different classes of *Stochastic Runge-Kutta methods*(SRK methods) are presented in Section 4.1.2. Then, generalizing a result already known for ODEs (see e.g. [3, p.101]), conditions under which these classes of methods conserve quadratic invariants are given and proven. To support the theory, four numerical schemes are presented in Section 4.1: Euler-Maruyama, Milstein, Stochastic Midpoint and Scheme 5.2 from [1]. Some of the methods are conservative, whereas some are not; all methods have a theoretical mean-square order of convergence 0.5 or 1.0. In Section 5.1, four test problems and the corresponding numerical results are presented. Approximations to single realizations are plotted, both in terms of solutions to the SDEs and the nominally conserved invariants. Mean-square convergence plots are also given for all the problems in the same section. In Section 5 some particular results are commented, especially where the numerical experiments deviates from the theory presented in Section 4.1.2; therefore, an alternative explanation is offered. In particular, issues with the implementation of Scheme 5.2 are discussed in Section 6.3.

The second issue studied finds its background in a paper by Burrage and Burrage [2]. Considering a Hamiltonian differential equation in a deterministic setting, addive noise is introduced in one of the variables. As in [2], it is demonstrated that the expectation of the Hamiltonian funciton drifts linearly with time for SDEs in the Itô sense. Then, a similar result is shown for Stratonovich SDEs. Both results are found in Section 3.2. To preserve said phenomenon, a composition method framework combining Euler-Maruyama steps and *implicit Runge-Kutta*(IRK) methods are presented in Section 4.2; four structure-preserving, albeit not necessarily symplectic IRK methods are suggested applied. Burrage and Burrage [2] shows that the four given IRK methods actually preserves the Hamiltonian drift well. In Section 5.2, two test problems are suggested: the Double Well and the Hénon-Heiles problem. No numerical results are presented, however; a short comment is made in Section (6.3).

The thesis is reviewed, in Chapter 7, with comments on theory and results. In addition, improvements to the thesis as well as directions in which a Master's thesis or other further work might move are suggested.

Contents

Tables

Figures

Chapter 1

Introduction

Upon first encountering *stochastic invariants*, it might seem an oxymoron. *Stochastic*, from the greek word *stochos* meaning aim or guess, pertains to something which is adequately modelled by a random probability distribution; its opposite would be *deterministic*, literally meaning that its value would be predetermined or given from before. On the other hand, *invariant* quite literally means something that does not change. However contradictory these terms might seem, *Stochastic Differential Equations* (SDEs) for which there exists invariants - some even have several - can be used to model a huge variety of practical problems. Methods preserving invariants and other structural qualities of differential equations might be just as numerous.

1.1 Structure-Preservation in Ordinary Differential Equations

Although *Ordinary Differential Equations* (ODEs) with important *first integrals* have been known and studied methodically at least since 17th century [4], especially within classical mechanics, interest in first integrals compounded significantly in the 19th century [3, p. 256]; the term *invariant*, which in the setting of differential equations is synonymous, was coined as late as in 1874 [5]. Hamiltonian functions was introduced in 1833, and with the formulations of Hamiltonian mechanics, a more elegant framework for studying invariant functions was developed. Already in 1836, Jacobi established the connection between Hamiltonian functions and symplectic transformations, in that they preserve the Hamiltonian function of an ODE [3]. Poincaré established in 1899 that the flow of a Hamiltonian ODE in itself is a symplectic mapping [3].

Structure-preserving numerical methods has been an important area of investigation in the study of ODEs from the last decades 20th century [3]. Classes of methods known since the 19th century, such as *Runge-Kutta* (RK) and *linear multistep* methods (LMMs), can preserve linear invariants. However, conditions for conserving quadratic invariants for Runge-Kutta methods were given as late as 1987 and for partitioned Runge-Kutta methods in 1993 [3]; similar results holds for *partitioned Runge-Kutta* (PRK) methods. Importantly, there exists no single method which can conserve all first integrals of arbitrary order larger than 2, that is, quadratic invariants. Any method or family of methods aiming to preserve first integrals of cubic or higher order must therefore be adapted to the type of problem at hand. Moreover, no symplectic numerical method can conserve a Hamiltonian function H exactly in general, nor can any method exactly preserving general Hs be symplectic [6].

One such class of energy-preservin methods is called *Discrete Gradient* (DG) methods, where the differential equation is expressed through the gradient of the invariant function, which is then discretized. The advantage of DG methods is that, with good discretization, they can conserve first integrals and Lyapunov functions in ODEs exactly (see e.g. [7]). An important newer class of DG methods was introduced by Quispel and McLaren in [8], dubbed the *Averaged Vector Field* (AVF) method; Celledoni, Owren and Sun shows in [9] that the AVF method applied to polynomial Hamiltonians in fact is the RK method of the fewest stages preserving Hamiltonians. In another paper, Celledoni, Grimm, et al. [10] expands the AVF framework to Hamiltonian *Partial Differential Equations* (PDEs), further demonstrating its powerful energy-preserving properties.

Another class of methods, which can be considered a form of collocation methods, are the so-called Hamiltonian Boundary Value Methods (HBVMs) introduced by Brugnano, Iavernaro and Trigiante in [11–13]. As one could deduce from the name, this class of methods is especially suited to Hamiltonian problems; in fact an $s-$stage HBVM method with a parameter $k \geq s$ ($k = s$ + number of *silent stages*) can preserve any polynomial Hamiltonian of degree $p \leq 2k/s$ and guarantee a local energy error of size $\mathcal{O}(\Delta t^{2k+1})$ for general Hamiltonians $H \in C^{(2k+1)}(\mathbb{R})$ [11], where Δt is step-size. Incidentally, the AVF method can be derived from HBVMs [14]. Burrage and Burrage adapt a symplectic variant of HBVMs, dubbed SLIRK methods (Symplectic Low-rank Implicit Runge Kutta), where instead of imposing exact solutions of Hamiltonians, they impose symplecticity; however, they do not show similarly good results for preserving the Hamiltonian function [15]. In [16], Brugnano and Iavernaro builds on the theory developed for HBVMs, creating a more general class of methods they denote *Line Integration Methods*(LIM), which can be made to conserve several suitably smooth invariants without adding stages to the method.

1.2 Structure-Preservation in Stochastic Differential Equations

Whereas white noise had been used informally in differential equations model earlier, the stochstic calculus traces its origin to Kiyosi Itô's pioneering paper "Stochastic Integral" from 1944 [17]. In it, he defines the Itô integral and the Itô formula, the stochastic equivalent to the chain rule in classic calculus, which together form the foundations of Itô calculus [18]. Ruslan Leont'evich Stratonovich presented a different definition of the stochastic integral in 1966, for which calculus more closely resemble that of deterministic theory [19]; it is much utilized in physics [20]. Both Itô and Stratonovich calculus are used within several scientific domains, and can in many cases be used interchangeably, due to a simple conversion formula.

Although there are significant differences to the way such methods are constructed, there exists anologous classes of one-step methods for stochastic differential equations relative to those of ordinary differential equations. There exists adaptations of RK methods (so-called Stochastic Runge-Kutta methods, short SRK) both for Itô SDEs [21] and for Stratonovich SDEs [22]. To design such methods, one usually applies the stochastic equivalent to Taylor expansions. These are known as Wagner-Platen series, or Itô-Taylor or Stratonovich-Taylor expansions [21].

Several ways to measure stability and convergence of numerical methods applied to SDEs, which is not as trivial as for ODEs. The most commonly applied convergence concepts are that of *mean-square*, *strong* and *weak* sense [20, 21, 23]. Mean-square(ms) convergence implies strong convergence [20]; both concern pathwise approximation of a stochastic process. Weak convergence concerns the convergence in probability distribution rather than individual realizations (e.g. [21]). It should be noted that problems for which one is only concerned with weak convergence allows for significantly more methodical flexibility, and it is easier for methods to obtain higher order convergence in the weak sense. In fact, methods of ms or strong order p is guaranteed weak order p or higher; however, increasing the strong order of a method does not guarantee an increased weak order [20, p.98] Both ways to measure convergence have their uses, depending on the problem at hand. To evaluate the order of convergence of a method, as well as test other statistical properties, Monte Carlo simulations must be performed.

As for deterministic equations, it is possible to work with first integrals or invariant functions for stochastic integrals. Hong, Xu et al. [1] gives order conditions for which the SRK methods of [22], as well as partitioned SRK methods (SPRK), preserves quadratic invariants of Stratonovich SDEs. This can be seen as adapting the results of Hairer, Lubich and Wanner in [3, p.97.104] to an SDE setting. Based on these results and the B-series theory for SRK methods formulated by P. and K. Burrage in [24] for strong convergence and by Komori in [25] and Rößler in [26] for weak convergence, Anmarkrud and Kværnø gives order conditions for which SRK methods preserve quadratic invarints in [27]. Furthermore, Debrabant and Kværnø goes on to show in [28] that for single integrand SDEs, which might be conservative, deterministic RK methods of order p attain order $\lfloor p/2 \rfloor$ in both weak and ms sense. Additionally, Li, Zhang, Ma and Ding [29] adapts discrete gradient methods (as well as linear projection methods) to a stochastic setting and demonstrates that they conserve invariants.

The aforementioned papers on structure-preservation pertains mainly to Stratonovich integrals, as they share or "inherit" many properties from deterministic calculus; that is not to say structure-preservation is exlusive to Stratonovich SDEs. In [2], Burrage and Burrage show that Itô SDEs with (separable) Hamiltonian drift terms and additive diffusion terms of possibly multidimensional noise produce Hamiltonian functions whose expectation drifts linearly in time; in contrast, the invariants in earlier cited articles are conserved not only in expectation, but pathwise. In the same article, they give a framework for energy-preserving methods that conserves this property. This result builds on that of [15] by the same authors, though here they only consider scalar noise. Elaborating on this problem format, Chen, Cohen, D'Ambrosio and Lang [30] presents a scheme of ms and weak order of convergence 1. This scheme preserves the linear drift in the expectation of the Hamiltonian exactly through evaluations of the potential vector field, similarly to what is done in the AVF method. However, as noted by D'Ambrosio, Giordano, Paternoster and Ventola [31], perturbation analysis shows that for sufficiently large diffusion terms, the error in the numerical solution compared to the expected value of the Hamiltonian function cannot be reduced efficiently by decreasing stepsize. Taking an alternative route, [32] constructs symplectic SRK methods for Itô problems with additive noise, also giving conditions for ms and weak order order 1.5; in the case of particular second-order Hamiltonian systems, they give conditions for weak and ms convergence of order 2.0. As a final apropos: LMMs can be adapted to a stochastic setting (see e.g. [21, Ch.11.4]); analysis of long-term properties of two-step methods for certain second order problems can be found in [33].

1.3 In This Thesis

From the two previous sections, it should be clear that the study of structure-preserving methods for SDEs is dynamic and ongoing, much as the SDEs are. There are also many known applications within a range of domains, including Molecular Dynamics, Bayesian statistics, Physics, Chemistry and Biology [20, Ch.5].

This thesis is mainly based on the paper by Hong, Xu, and Wang[1], as well as the paper by Burrage and Burrage [2]. In it, the theory of said papers are treated and slightly adapted, whereas the numerical results are reproduced. The goal has mainly been for the author to familiarize himself with the subject matter at hand to prepare for a Masters thesis; consequently, no major new findings are reported.

The thesis is structured thus:

- Chapter 2 presents some preliminary rules of stochastic calculus, as well as geometric integration theory for ODEs.
- In Chapter 3, the concepts from the aforementioned papers are explained and elaborated on.
- Methods to solve the SDEs presented in Chapter 3 are presented in Chapter 4 and their properties are investigated.
- In Chapter 5, the methods from Chapter 4 are applied to test problems from classes of SDEs presented in Chapter 3 and the numerical properties are reported.
- In Chapter 6, the numerical results from Chapter 5, as well as their correspondance with theoretical properties explored in Chapters 3 and 4.
- Chapter 7 summarizes the findings of the rest of the thesis. Also, areas of further investigation are suggested.

Chapter 2

Preliminaries

This is a thesis about SDEs. The theory of SDEs depends on a more rigorous probability theory than generally applied in statistics, with foundations in measure theory. The numerical study of SDEs largely builds on theory and methods developed for ODEs; similarly, the theory and methods of Stochastic Partial Differential Equations(SPDEs) builds on the theory and methods developed for Partial Differential Equations(PDEs). This demands a good understanding of and some experience with numerical mathematics.

At the Norwegian University of Science and Technology (NTNU), MA8109 - Stochastic Processes and Differential Equations and its prerequisites would give the reader the necessary theoretical understanding of SDEs. MA8404 - Numerical Integration of Time Dependent Differenetial Equations and its prerequisites would offer a decent introduction to geometrical integration in ODEs relevant to this thesis. For those without much knowledge of the relevant fields, this section should provide the necessary background to understand the rest of the thesis. However, it should be possible to implement methods and replicate the results purely from reading the other chapters; a link to code repository will also be given in the appendix.

2.1 Notation

Notation	Description	Example
i, j, k, l	Variables used for indices.	b_i, σ_{jk}
$a_{d_1 \times \cdots \times d_n}$	Array of order n with dimensions $d_1, \ldots, d_n$ with all entries set to $a \in \mathbb{R}$.	$1_d \in \mathbb{R}^d$, $0_{d \times m} \in \mathbb{R}^{d \times m}$
$\text{diag}\{v\}$	Diagonal matrix with v on its diagonal.	$\text{diag}\{1_d\} =: I_d$
I_d	Identity matrix in $\mathbb{R}^{d \times d}$.	
$\partial_u f$	Partial derivative of a function f w.r.t. scalar input u.	$\partial_t f = \frac{\partial f}{\partial t}$
$\dot{x}$	$\partial_t x$, i.e. time derivative of variable x	
$\nabla_x f$	Gradient of f w.r.t. input vector x.	$(\nabla_x f)_i = \partial_{x_i} f$
$\nabla_x^2 f$	Hessian of f w.r.t. input vector x.	$\left(\nabla_x^2 f\right)_{ij} = \partial_{x_i} \partial_{x_j} f$
$W(t)$	Wiener Process, with dimensions from context; Section 2.2.1.	
$X(t), Y(t)$	Stochastic processes; see Section 2.3	
$b(t), f(t), g_0(t)$	Drift term	
$\sigma(t)$, $g_i(t)$, $i = 1, \ldots, m$	Diffusion term	
$\text{tr}\{A\}$	(Generalized) trace of an array A; Definition 2.5.	
$\lVert\cdot\rVert_F$	Frobenius norm.	
$\lVert\cdot\rVert_2$	2-norm or Euclidean norm.	
$\lVert\cdot\rVert_{\mathcal{L}(\Omega)}$	Probability norm.	

$\mathrm{E}[\cdot]$	Expectation of a (random) variable.	$\mathrm{E}[X]$, $\mathrm{E}\left[\left(\int_{t_0}^{T}\varphi(\tau)\mathrm{d}W(\tau)\right)^2\right]$
$\perp$	Independence of random variables.	$X(t)\perp Y(t)$, $W_1(t)\perp W_2(t)$
$\omega\in\Omega$	Sample path in sample space.	
$\mathcal{F}$	Event space ($\sigma-$algebra).	
$\mathcal{F}(t)$	Filtration of $\mathcal{F}$ w.r.t. time t.	
$P(\cdot)$	Probability measure.	
a.e., a.a., a.s.	Almost everywhere, almost all, almost surely (with probability 1).	
$\otimes$	Outer or Kronecker product.	
$\wedge$	Wedge product (see Definition 3.3).	$\dot{x}(t)\wedge\dot{x}(t)$

Table 2.1: Often used notation

2.2 Probabilitiy Theory

In this thesis, when concerned with stochastic differential equations, operating in complete probability spaces $\{\Omega,\mathcal{F},\mathcal{F}(t),P\}$ is considered a given. As is convention, when considering stochastic processes $X(t,\omega):[t_0,T]\times\Omega\to\mathbb{R}^d$, the notation $X(t)$ with *sample path* ω implicit is often applied. For an introduction to the foundations of probability theory, the reader is referred to [34].

2.2.1 Wiener Processes

Essential to the mathematical foundation of SDEs is the concept of the Wiener Processes.[1]

Definition 2.1. [17] A function $W(t)$ for $t\in[0,T]$ is a *Wiener Process* if it has

1. $W(0)=0$ a.s. (Zero initial value)
2. $W(t)-W(s)\sim N(0,t-s)$, (Normally distributed increments)
3. $W(t_2)-W(t_1)\perp W(s_2)-W(s_1)$ for $0\le s_1<s_2<t_1<t_2\le T$ (Independent increments)
4. $\lim_{t\to s}W(t,\omega)-W(s,\omega)=0\quad\forall\omega\in\Omega$ a.s. (Continuous sample paths)

This can be generalized to an m dimensions with

$$W(t)=\{W_1(t),W_2(t),\ldots,W_m(t)\}^T,$$

where W_i and W_j are independent Wiener Processes $\forall i,j=1,2,\ldots,m:i\neq j$

Properties 2.2. Properties of the Wiener processes $W(t)$

1. $\mathrm{E}[W(s)W(t)]=\min\{s,t\}$
2. $\mathrm{E}[W(t)|W(s)]=\mathrm{E}[W(t)|\mathcal{F}(s)]=W(s)$ for $s<t$
3. $\widetilde{W}(t):=W(t_0+t)-W(t_0)$ is a Wiener Process for any fixed $t_0\ge 0$ (Translation invariance)
4. $\widetilde{W}(t):=W(\lambda t)/\sqrt{\lambda}$ is a Wiener Process for any $\lambda\in\mathbb{R}^+$ (Scale invariance)
5. $W(t)$ is nowhere differentiable a.s.

Remark. $\mathcal{F}(t)\subset\mathcal{F}$ is the $\sigma-$algebra generated by $W(t)$.

2.3 Stochastic Differential Equations

The following definitions and results are necessary to work with SDEs, and follows notation from [34] and [17].

[1] The terms and notation Brownian motion $B(t)$ and Wiener process $W(t)$ are used interchangeably in the literature. Following the reasoning of Kloeden and Platen[21, p.40] rather than the notation used by Øksendal[34], the author considers Wiener processes in this thesis.

Definition 2.3. Let $\{\Omega, \mathcal{F}, \mathcal{F}(t), P\}$ be a complete probability space with probability measure P. Then, the *space of stochastic processes* $\mathcal{L}_{ad}(L^2[t_0, T], \Omega) = \mathcal{L}^2_{ad}$ is the space of all functions $f, g : [t_0, T] \times \Omega \to \mathbb{R}$ such that

- f is $\mathcal{B}[t_0, T] \times \mathcal{F}-$ measurable
- f is $\mathcal{F}(t)-$ adapted
- $\int_{t_0}^{T} \mathrm{E}[|f|^2]\mathrm{d}t < \infty \quad$ for a.a. $\omega \in \Omega$

equipped with an inner product

$$< f, g >_{\mathcal{L}^2_{ad}} := \int_{t_0}^{T} \mathrm{E}[|f\, g|]\, \mathrm{d}t$$

and a norm

$$\|f\|_{\mathcal{L}^2_{ad}} := < f, f >^{\frac{1}{2}}_{\mathcal{L}^2_{ad}} .$$

2.3.1 Itô Calculus

Definition 2.4. [34, p.29] Let $f \in \mathcal{L}^2_{ad}$ and let $\{\varphi_n\}_{n\in\mathbb{N}} \in M^2_{step} \subset \mathcal{L}^2_{ad}$ be a sequence of stochastic step functions such that

$$\mathrm{E}\left[\int_{t_0}^{T} (f(t,\omega) - \varphi_n(t,\omega))^2\right] \xrightarrow{n\to\infty} 0.$$

Then, the *Itô integral* $I(f) \in L^2(\Omega)$ is given by

$$I(f) := \int_{t_0}^{T} f(t)\mathrm{d}W(t) = \lim_{n\to\infty} \int_{t_0}^{T} \varphi_n \mathrm{d}W(t). \tag{2.1}$$

Remark. For a better insight into the steps needed in this construction, see [17, Ch.4.3] and [34, p.26-29].

Lemma 2.1. *Properties of the Itô integral.*
Let $f, g \in \mathcal{L}^2_{ad}$ and $t_0 < t_1 < T$. Then the following properties holds for the Itô integral.

- *Linearity:* $\int_{t_0}^{T}(af + bg)dW(t) = a\int_{t_0}^{T} f dW(t) + b\int_{t_0}^{T} g dW(t) \qquad \forall a, b \in \mathbb{R}$
- *Partition:* $\int_{t_0}^{T} f dW(t) = \int_{t_0}^{t_1} f dW(t) + \int_{t_1}^{T} f dW(t) \quad$ *for a.a.* $\omega \in \Omega$
- *Expectation:* $E\left[\int_{t_0}^{T} f dW(t)\right] = 0 \quad$ *for a.a.* $\omega \in \Omega$
- *Itô isometry:* $E\left[\left(\int_{t_0}^{T} f dW(t)\right)^2\right] = E\left[\int_{t_0}^{T} |f|^2 dt\right]$
- $\int_{t_0}^{T} f dW(t)$ *is* $\mathcal{F}(T)-$ *measurable*

Definition 2.5. Let $A \in \mathbb{R}^{n\times n}$ be any real-valued matrix. Then the *trace of A* is an operator $\mathrm{tr}\{\cdot\} : \mathbb{R}^{n\times n} \to \mathbb{R}$ such that

$$\mathrm{tr}\{A\} = \sum_{i=1}^{n} A_{ii}, \tag{2.2}$$

i.e. the sum of the diagonal elements of A.

For multidimensional arrays of order greater than 2, the *generalized trace* contracts the first and last dimension of the array by summation, implicitly assuming the dimensions have the same length.

Remark. Higher order arrays occuring here are usually a composition of the Gradient or Hessian of a function with some matrix functions. The gradient and Hessian of a vector function (e.g. $b(t,X) \in \mathbb{R}^d$) would be multidimensional arrays of order 2 and 3, respectively, i.e. $\nabla_x b(t,X) \in \mathbb{R}^{d\times d}$ and $\nabla^2_x b(t,X) \in \mathbb{R}^{d\times d\times d}$. The gradient and Hessian of a matrix function (e.g. $\sigma(t,X) \in \mathbb{R}^{d\times m}$) would be multidimensional arrays of order 3 and 4, respectively, i.e. $\nabla_x \sigma(t,X) \in \mathbb{R}^{d\times d\times m}$ and $\nabla^2_x \sigma(t,X) \in \mathbb{R}^{d\times d\times d\times m}$.

As an example, let $\Sigma \in \mathbb{R}^{d\times m}, g \in C^{(2)}(\mathbb{R}^d, \mathbb{R}^d) \Rightarrow \Sigma^T \nabla_x^2 g(X)\Sigma \in \mathbb{R}^{m\times d\times m}$. Then, $\mathrm{tr}\{\Sigma^T \nabla_x^2 g(X)\Sigma\} : \mathbb{R}^{m\times d\times m} \to \mathbb{R}^m$ such that

$$\mathrm{tr}\{\Sigma^T \nabla_x^2 g(X)\Sigma\} = \sum_{k=1}^{m}\sum_{i,j=1}^{d} \Sigma_{ik}\partial_{x_i}\partial_{x_j} g(X)\Sigma_{jk}.$$

Definition 2.6. [34, p.44, 48] Let $W(t)$ be an $m-$dimensional Wiener process on $\{\Omega, \mathcal{F}, P\}$. A *$d-$dimensional Itô process* $X(t)$ on $\{\Omega, \mathcal{F}, P\}$ is a stochastic process or integral such that

$$X(t) = X_0 + \int_{t_0}^{t} b(\tau,\omega)\mathrm{d}\tau + \int_{t_0}^{t} \sigma(\tau,\omega)\mathrm{d}W(\tau), \tag{2.3}$$

where $\sigma \in \mathcal{L}_{ad}^2$, such that

$$P\left(\int_{t_0}^{t} \sigma(\tau,\omega)^2\mathrm{d}\tau < \infty \quad \forall t \geq t_0\right) = 1, \quad \text{(almost surely finite)}$$

b is $\mathcal{F}(t)-$adapted and

$$P\left(\int_{t_0}^{t} |b(\tau,\omega)|\mathrm{d}\tau < \infty \quad \forall t \geq t_0\right) = 1.$$

Theorem 2.2. *[18] Itô's Formula.*
Let $X(t)$ be a d_1-dimensional stochastic process satisfying

$$dX(t) = b(t,X)dt + \sigma(t,X)dW(t), \qquad X(t_0) = X_0, \quad t \in [t_0, T],$$

where W is an $m-$dimensional Wiener process.
Furthermore, let $g \in C^{(2)}([t_0,T]\times\mathbb{R}^{d_1}, \mathbb{R}^{d_2})$ and $Y(t) = g(t,X(t))$. Then,

$$\begin{aligned} dY(t) &= \partial_t g dt + (\nabla_x g)^T dX + \frac{1}{2} dX^T \left(\nabla_x^2 g\right) dX \\ &= \left(\partial_t g + (\nabla_x g)^T b + \frac{1}{2} tr\left\{\sigma^T \left(\nabla_x^2 g\right)\sigma\right\}\right) dt + (\nabla_x g)^T \sigma dW(t), \end{aligned} \tag{2.4}$$

where, using usual convention, $dW_j dW_k = 0$ if $j \neq k$, $(dW_j)^2 = dt$ and $dW_j dt = dt dW_j = (dt)^2 = 0$, $\forall j$.
Element-wise, for $i = 1, \ldots, d_2$, this becomes

$$\begin{aligned} dY_i(t) &= \partial_t g_i dt + \sum_{j=1}^{d_1} \partial_{x_j} g_i dX_j + \frac{1}{2}\sum_{j,k=1}^{d_1} \partial_{x_j}\partial_{x_k} g_i dX_j^2 \\ &= \left(\partial_t g_i + \sum_{j=1}^{d_1}\left(\partial_{x_j} g_i b_j + \frac{1}{2}\sum_{k=1}^{d_1}\partial_{x_j}\partial_{x_k} g_i \sum_{l=1}^{m}\sigma_{jl}\sigma_{kl}\right)\right) dt + \sum_{j=1}^{d_1}\partial_{x_j} g_i \sum_{l=1}^{m}\sigma_{jl} dW_l(t) \end{aligned} \tag{2.5}$$

Theorem 2.3. *Existence and Uniqueness.[17, 34]*
Let $b : [t_0,T]\times\mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : [t_0,T]\times\mathbb{R}^d \to \mathbb{R}^{d\times m}$ be measurable functions satisfying

1. *Lipschitz condition:*

$$\|b(t,x) - b(t,y)\|_2 + \|\sigma(t,x) - \sigma(t,y)\|_F \leq C\|x-y\|_2 \quad \forall x, y \in \mathbb{R}^d \tag{2.6}$$

   for some constant C.
2. *Linear growth condition:*

$$\|b(t,x)\|_2 + \|\sigma(t,x)\|_F \leq K(1 + \|x\|_2) \quad \forall x \in \mathbb{R}^d \tag{2.7}$$

   for some constant K.

Let $X_0 \in \mathcal{F}_0 \subset \mathcal{F}$ be a random variable independent of the $\sigma-$algebra $\mathcal{F}(\infty)$ generated by the $m-$dimensional Wiener Process $W(t,\omega)$. Assume that X_0 is finite in the mean square, i.e. $E[|X_0|^2] < \infty$.

Then, if $X(t)$ satisfies the equation

$$dX(t) = b(t,X)dt + \sigma(t,X)dW(t), \qquad X(t_0) = X_0, \quad t \in [t_0, T], \tag{2.8}$$

it is a unique and $t-$continuous solution adapted to the filtration $\mathcal{F}^{X_0}(t) = \mathcal{F}_0 \otimes \mathcal{F}(t)$ with a finite mean-square bound, i.e.

$$E\left[\int_{t_0}^{T} |X(t)^2| dt\right] < \infty. \tag{2.9}$$

Remark. [17, p. 190] For $X(t)$ to solve (2.8), it holds that

- $\sigma_{ij} \in \mathcal{L}^2_{ad}$ for $i = 1,\dots,d$ and $j = 1,\dots,m$
- $f_i \in \mathcal{L}^1_{ad}$ for $i = 1,\dots,d$
- (2.8) is satisfied for every $t \in [t_0, T]$ a.s.

2.3.2 Stratonovich Calculus

Definition 2.7. [17, p. 120] Let $X(t)$ and $Y(t)$ be [possibly multidimensional] Itô processes. The *Stratonovich integral of $X(t)$ w.r.t. $Y(t)$* is given by

$$\int_{t_0}^{t} X(\tau) \circ \mathrm{d}Y(\tau) = \int_{t_0}^{t} X(t)\mathrm{d}Y(t) + \frac{1}{2}\int_{t_0}^{t} \mathrm{d}X(\tau)\cdot \mathrm{d}Y(\tau). \tag{2.10}$$

Equivalently, in symbolic form, we get that

$$X(t) \circ \mathrm{d}Y(t) = X(t)\mathrm{d}Y(t) + \frac{1}{2}\mathrm{d}X(t)\cdot \mathrm{d}Y(t). \tag{2.11}$$

Theorem 2.4. *[19] Let $X(t)$ be an Itô process such that*

$$dX(t) = \overline{b}(t,X)dt + \sigma(t,X)dW(t), \qquad b(t,X) \in \mathbb{R}^d,\ \sigma(t,X) \in \mathbb{R}^{d\times m}, \tag{2.12}$$

where $W(t)$ is an $m-$ dimensional Wiener Process. This corresponds to the Stratonovich process

$$\begin{aligned} dX(t) &= \underline{b}dt + \sigma \circ dW(t) \\ &= \left(\overline{b} - \frac{1}{2} tr\left\{\sigma^T \nabla_x \sigma\right\}\right) dt + \sigma \circ dW(t). \end{aligned} \tag{2.13}$$

Equivalently, $\overline{b} = \underline{b} + \frac{1}{2} tr\left\{\sigma^T \nabla_x \sigma\right\}$.

Remark. Consider $\sigma(t,X(t)) \in \mathbb{R}^{d\times 1}$ and $\nabla_x \sigma(t,X(t)) \in \mathbb{R}^{d\times d\times 1}$ for $m = 1$, i.e.

$$\mathrm{tr}\{\sigma^T \nabla_x \sigma\}_i = \sum_{k=1}^{1}\sum_{j=1}^{d} \sigma_{jk}\partial_{x_j}\sigma_{ik} = \sum_{j=1}^{d} \sigma_{j1}\partial_{x_j}\sigma_{i1}.$$

Proof. From the definition of the Stratonovich integral, we get that

$$\sigma \circ \mathrm{d}W(t) = \sigma \mathrm{d}W(t) + \mathrm{d}\sigma \cdot \mathrm{d}W(t). \tag{2.14}$$

Using Itô's formula (2.4), it follows that

$$\begin{aligned} \mathrm{d}\sigma(t,X)\cdot \mathrm{d}W &= \left(\left(\partial_t \sigma + (\nabla_x \sigma)^T b + \frac{1}{2}\mathrm{tr}\left\{\sigma^T\left(\nabla_x^2 \sigma\right)\sigma\right\}\right)\mathrm{d}t + (\nabla_x \sigma)^T \sigma \mathrm{d}W\right)\cdot \mathrm{d}W \\ &= \left((\nabla_x \sigma)^T \sigma \mathrm{d}W\right)^T \mathrm{d}W \\ &= \mathrm{d}W^T \sigma^T (\nabla_x \sigma)\mathrm{d}W \\ &= \mathrm{tr}\left\{\sigma^T \nabla_x \sigma\right\}\mathrm{d}t. \end{aligned} \tag{2.15}$$

as $\mathrm{d}t\mathrm{d}W_i(t) = 0 = \mathrm{d}W_i\mathrm{d}W_j$ for $i \neq j$ and $\mathrm{d}W_i^2 = \mathrm{d}t\ \forall i, j = 1, \ldots, m$.

Then, inserting into the symbolic expression for the SDE,

$$\begin{aligned}\mathrm{d}X(t) &= \overline{b}\mathrm{d}t + \sigma\mathrm{d}W(t) = \overline{b}\Delta t + \sigma \circ \mathrm{d}W(t) - \mathrm{d}\sigma \cdot \mathrm{d}W(t)\\ &= \left(\overline{b} - \mathrm{tr}\left\{\sigma^T \nabla_x \sigma\right\}\right)\mathrm{d}t + \sigma \circ \mathrm{d}W(t)\\ &= \underline{b}\mathrm{d}t + \sigma \circ \mathrm{d}W(t).\end{aligned}$$

□

Corollary 2.4.1. *For SDEs with additive noise, i.e. $\sigma(t, X(t)) = \sigma(t) \in \mathbb{R}^{d\times m}$, the Itô and Stratonovich formulations coincide; that is, $\overline{b} = \underline{b}$.*

Proof.

$$\nabla_x \sigma = \nabla_x \sigma(t) = 0_{d\times d\times m} \qquad \Rightarrow \qquad \underline{b} = \overline{b} + \frac{1}{2}\mathrm{tr}\left\{\sigma^T \nabla_x \sigma\right\} = \overline{b} + 0_d = \overline{b}.$$

□

Remark. The chain rule and expectation of Stratonovich and Itô integrals with additive noise do not coincide.

Theorem 2.5. *Stratonovich Chain Rule.[34]*
Let $X(t)$ be a stochastic Stratonovich process such that

$$dX(t) = \underline{b}(t, X)dt + \sigma(t, X) \circ dW(t) \tag{2.16}$$

and let $g \in C^2([t_0, T] \times \mathbb{R}^{d_1}, \mathbb{R}^{d_2})$. Then,

$$\begin{aligned}dY(t) = dg(t, X(t)) &= \partial_t g dt + \nabla_x g \cdot dX\\ &= \left(\partial_t g + (\nabla_x g)^T \underline{b}\right)dt + (\nabla_x g)^T \sigma \circ dW(t).\end{aligned} \tag{2.17}$$

Proof. The Stratonovich process in Theorem 2.5 corresponds with the Itô process

$$\begin{aligned}\mathrm{d}X(t) &= \left(\underline{b} + \frac{1}{2}\mathrm{tr}\left\{\sigma^T \nabla_x \sigma\right\}\right)(t, X)\mathrm{d}t + \sigma(t, X)\mathrm{d}W(t)\\ &= \overline{b}(t, X)\mathrm{d}t + \sigma(t, X)\mathrm{d}W(t).\end{aligned} \tag{2.18}$$

Using the Chain Rule for Itô processes (Theorem 2.2), it follows that

$$\mathrm{d}g(t, X(t)) = \left(\partial_t g + (\nabla_x g)^T\ \overline{b} + \frac{1}{2}\mathrm{tr}\left\{\sigma^T \left(\nabla_x^2 g\right)\sigma\right\}\right)\mathrm{d}t + (\nabla_x g)^T \sigma \mathrm{d}W(t) \tag{2.19}$$

Looking at the last term and expanding using Definition 2.7:

$$\begin{aligned}(\nabla_x g)^T \sigma \mathrm{d}W(t) &= (\nabla_x g)^T \sigma \circ \mathrm{d}W(t) - \frac{1}{2}\mathrm{d}\left((\nabla_x g)^T \sigma\right) \cdot \mathrm{d}W(t)\\ &= (\nabla_x g)^T \sigma \circ \mathrm{d}W(t) - \frac{1}{2}\left(\mathrm{tr}\left\{\sigma^T \left(\nabla_x^2 g\right)\sigma\right\} + (\nabla_x g)^T \mathrm{tr}\left\{\sigma^T \nabla_x \sigma\right\}\right)\mathrm{d}t,\end{aligned} \tag{2.20}$$

using the same argumentation as for (2.15). Hence, inserting this into (2.19),

$$\begin{aligned}\mathrm{d}g(t, X(t)) &= \left(\partial_t g + (\nabla_x g)^T \overline{b} + \frac{1}{2}\mathrm{tr}\left\{\sigma^T \left(\nabla_x^2 g\right)\sigma\right\}\right)\mathrm{d}t + (\nabla_x g)^T \sigma \circ \mathrm{d}W(t)\\ &\quad - \frac{1}{2}\left(\mathrm{tr}\left\{\sigma^T \left(\nabla_x^2 g\right)\sigma\right\} + (\nabla_x g)^T \mathrm{tr}\left\{\sigma^T \nabla_x \sigma\right\}\right)\mathrm{d}t\\ &= \left(\partial_t g + (\nabla_x g)^T \underline{b}\right)\mathrm{d}t + (\nabla_x g)^T \sigma \circ \mathrm{d}W(t)\\ &= \partial_t g \mathrm{d}t + \nabla_x g \cdot \mathrm{d}X.\end{aligned} \tag{2.21}$$

□

2.3.3 Numerical Simulation

Definition 2.8. A *one-step approximation* $Y_{t,x}(t+\Delta t)$ to a stochastic process $X(t)$ as given by (2.8) is given as

$$Y_{t,x}(t+\Delta t) = x + \Phi(t,x,\Delta t; W(\theta)-W(t), \theta \in [t, t+\Delta t]), \quad t \in [t_0, T], \tag{2.22}$$

with $Y_0 = X_0 = X(t_0)$ and $Y_{k+1} = Y_{t_k,Y_k}(t_{k+1}) = Y_{t_0,X_0}(t_{k+1})$, where $t_k = t_0 + k\cdot\Delta t$ for $k = 0,\dots,N$; $t_N = T$.

Definition 2.9. Let $X_{t,x}(t+\Delta t)$ denote a stochastic process which takes the value x at time t, and let $Y_{t,x}(t+\Delta t)$ denote a one-step approximation of $X_{t,x}(t+\Delta t)$. The *local error* $\delta_{t,x}(\Delta t)$ is given by

$$\delta_{t,x}(\Delta t) = X_{t,x}(t+\Delta t) - Y_{t,x}(t+\Delta t), \tag{2.23}$$

while the *global error* $\varepsilon_n(\Delta t)$ is given by

$$\varepsilon_n(\Delta t) = X(t_n) - Y_n. \tag{2.24}$$

Definition 2.10. A one-step method has *strong mean-square order of convergence* γ if

$$\sqrt{\mathrm{E}[|\varepsilon_n(\Delta t)|^2]} \le K(\Delta t)^\gamma \tag{2.25}$$

for some constant K.

Theorem 2.6. *Milstein's Fundamental Theorem of Mean-Square Order of Convergence [20].*
Suppose that

1. $\left|E\left[\delta_{t,x}(\Delta t)\right]\right| \le K_1\sqrt{1+|x|^2}(\Delta t)^{p_1}$
2. $\sqrt{E\left[|\delta_{t,x}(\Delta t)|^2\right]} \le K_2\sqrt{1+|x|^2}(\Delta t)^{p_2}$
3. $0 \le p := p_2 - \frac{1}{2} \le p_1 - 1$

Then, for any $N \in \mathbb{N}$,

$$\sqrt{E[|\varepsilon_n(\Delta t)|^2]} \le K_3\sqrt{1+E[|X_0|^2]}(\Delta t)^p \quad \text{for } n = 1,\dots,N. \tag{2.26}$$

In other words, the one-step method has mean-square order of convergence p.

Remark. The constants K_1, K_2, K_3 are independent of X_0 and N; they only depend on the formulation of the problem and the one-step method.

2.4 Geometric Integration of Ordinary Differential Equations

Within this section, mainly *autonomous* ODEs, i.e. ODEs not explicitly time-dependent, of the format

$$\dot{x} = f(x(t)), \qquad x(t_0) = x_0, \qquad f : U \to \mathbb{R}^d, \quad U \subset \mathbb{R}^d \tag{2.27}$$

although it is quite possible to generalize to non-autonomous ODEs as well. This theory is mainly adapted from Hairer, Lubich and Wanner [3, Ch.IV & VI].

2.4.1 Invariants

Definition 2.11. [3] An *invariant* or a *first integral* is a function $I \in C^{(1)}(\mathbb{R}^d, \mathbb{R})$ such that for any x satisfying an Ordinary Differential Equation (ODE) of the form (2.27), it holds that

$$\frac{\mathrm{d}}{\mathrm{d}t} I(x) = I'(x)\,\dot{x} = \nabla I^T f(t,x) = 0 \qquad \forall t \in [0,T]. \tag{2.28}$$

From the definition above, it follows that

$$\begin{aligned}\int_{t_0}^{t} \frac{\mathrm{d}}{\mathrm{d}\tau} I(x(\tau))\,\mathrm{d}\tau = I(x(t)) - I(x_0) = 0 \quad \forall t \in [t_0, T] \\ \iff I(x(t)) = I(x_0) \quad \forall t \in [t_0, T].\end{aligned} \tag{2.29}$$

Linear invariants are of the form

$$I(x) = d^T x, \qquad d \in \mathbb{R}^d, \tag{2.30}$$

whereas quadratic invariants are of the form

$$I(x) = x^T C x, \qquad C \in \mathbb{R}^{d\times d} \text{ s.t. } C^T = C. \tag{2.31}$$

Note that invariants can also be of cubic or higher polynomial order, or not polynomial at all. It is also quite possible that certain problems conserve several invariants at once - see e.g. [3, Ch.VI.6.4].

McLahclan, Quispel and Robidoux [7] offers a reformulation of the system (2.27) to the following:

$$\dot{x} = S(t,x)\nabla I, \qquad S : [t_0, T] \times \mathbb{R}^d \to \mathbb{R}^{d\times d}, \quad S^T = -S; \tag{2.32}$$

in other words, the ODE can be expressed by a combination of some skew-/anti-symmetric matrix function S and its invariant I. It follows readily that

$$\frac{\mathrm{d}}{\mathrm{d}t} I(x(t)) = \nabla I^T S\, \nabla I = 0 \qquad \forall t \in [t_0, T] \tag{2.33}$$

owing to the skew-symmetry of S.

Given that the first integral doesn't vanish, they supply the following more general form of S:

$$S = \frac{f \cdot \nabla I^T - \nabla I \cdot f^T}{\|\nabla I\|^2}, \qquad \|\nabla I\|^2 = \nabla I^T \nabla I, \tag{2.34}$$

where f is as in (2.27) Next, an important class of invariants are presented.

2.4.2 Hamiltonian Systems

The derivation of the Hamiltonian system is adapted from [3].

The Hamiltonian comes from a reformulation of the *Lagrangian* of a system, given in the form

$$L(q,\dot{q}) = T(q,\dot{q}) - U(q), \qquad q(t_0), \dot{q}(t_0) = q_0, \dot{q}_0 \in \mathbb{R}^d, \tag{2.35}$$

where $T(q,\dot{q})$ is an expression for the kinetic energy and $U(q)$ is an expression for the potential energy in said system; q refers to position and $\dot{q}$ its time derivative. To arrive at a Hamiltonian expression for the system, the Lagrangian needs to be regular, i.e. have that the *Legendre transform*

$$p = \phi(q,\dot{q}) = \nabla_q L(q,\dot{q}), \tag{2.36}$$

is bijective (injective and surjective). The definition of the *Hamiltonian* of a system is given by

$$H(q,p) := p^T \dot{q} - L(q,\dot{q}). \tag{2.37}$$

Taking the derivative of (2.37), using (2.36) and the *Euler-Lagrange equations*

$$\nabla_q L(q,\dot{q}) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\nabla_{\dot{q}} L(q,\dot{q})\right), \tag{2.38}$$

this can be restated as a system of equations on the form

$$\dot{q} = \nabla_p H(q,p), \tag{2.39a}$$

$$\dot{p} = -\nabla_q H(q,p). \tag{2.39b}$$

By setting $x = (q^T, p^T)^T$, it follows that

$$\dot{x} = f(x) = J\nabla H(x), \qquad x(0) = x_0 \in \mathbb{R}^{2d}, \quad J = \begin{bmatrix} 0 & I_d \\ -I_d & 0 \end{bmatrix}, \tag{2.40}$$

where I_d refers to the identity matrix with d rows and d columns. Systems of ODEs of the form of (2.39) and (2.40) are known as *canonical Hamiltonian systems*.

Remark. It is assumed here and after that $H \in C^{(2)}(U, \mathbb{R})$, $U \subset \mathbb{R}^{2d}$, i.e. that H is a twice continuously differentiable function on a subset U of $\mathbb{R}^{2d}$.

Remark. The Hamiltonian function associated with an ODE is often denoted as the *energy* of the system.

Importantly, taking the derivative of $H(y(t))$ with respect to t leads to the expression

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} H(x(t)) = \nabla H(x(t))^T \dot{x} = \nabla H(x(t))^T J \nabla H(x(t)) = 0 \\ \Rightarrow H(x(t)) - H(x_0) = \int_0^t \frac{\mathrm{d}}{\mathrm{d}t} H(x(\tau))\mathrm{d}\tau = 0 \quad \forall t. \end{aligned} \tag{2.41}$$

From comparing (2.32) and (2.29) with (2.40) and (2.41), it becomes clear that Hamiltonians are invariants of their associated differential equation.

Definition 2.12. The *flow of an ODE* from an initial value x_0 at a (fixed) time t, denoted $\varphi_t(x_0)$, is

$$\varphi_t(x_0) = x(t),$$

where x satisfies (2.27) for the initial value x_0.

Remark. Variational equations, arising from $\partial_{x_0}\varphi_t$, are closely tied to symplecticity, yet outside the scope of this thesis.

Definition 2.13. [3, p. 185] An ODE of the form (2.40) is *locally Hamiltonian* if there exists a neighbourhood for every $x_0 \in U$ where $f(x) = J\nabla H(x)$ for some function H.

Symplecticity and the Connection to Hamiltonians

Definition 2.14. [3, p.183] Let ω be a bilinear mapping $\omega : \mathbb{R}^{2d} \times \mathbb{R}^{2d} \to \mathbb{R}$ such that

$$\omega(\xi, \eta) = \xi^T J_d \eta, \quad J = \begin{bmatrix} 0 & I_d \\ -I_d & 0 \end{bmatrix}.$$

A differentiable mapping $g : U \to \mathbb{R}^{2d}$ for an open set $U \subset \mathbb{R}^{2d}$ is called *symplectic* if

$$\omega(g'(x)\xi, g'(x)\eta) = \xi^T g'(x)^T J g'(x)\eta = \xi^T J \eta = \omega(\xi, \eta) \quad \forall \xi, \eta \in \mathbb{R}^{2d}. \tag{2.42}$$

Remark. For symplectic linear mappings $g(x) \mapsto Ax$, it follows that

$$\omega(Ax, Ay) = (Ax)^T JAy = x^T A^T JAy = x^T J y^T = \omega(x, y) \qquad \forall x, y \in \mathbb{R}^{2d}.$$

More succinctly, $A^T JA = J$.

Theorem 2.7. *[7, p.185] An ODE $\dot{x} = f(x)$, $x(t_0) = x_0 \in U$ for $f \in C^{(1)}(U, \mathbb{R}^{2d})$ is said to be locally Hamiltonian if and only if its flow φ_t (see Definition 2.12) is symplectic.*

Theorem 2.8. *[7, p.184] Let φ_t be the flow of a Hamiltonian system* (2.40)*. Then φ_t is a symplectic mapping.*

Remark. In fact, the symplecticity condition is a quadratic invariant of the variational equation of Hamiltonian systems.

2.4.3 Runge-Kutta Methods for ODEs

In this section, we are generally concerned with autonomous ODEs, i.e.

$$\dot{x} = f(x(t)), \quad x(t_0) = x_0, \quad t \in [t_0, T]. \tag{2.43}$$

Definition 2.15. An $s-$stage Runge-Kutta method applied the ODE

$$\dot{x} = f(t, x(t)), \quad x(t_0) = x_0, \quad t \in [t_0, T],$$

take the form

$$k_i = x_n + \Delta t \sum_{j=1}^{s} a_{ij} f(t_n + c_j \Delta t, k_j), \qquad i = 1, \dots, s,$$

$$x_{n+1} = x_n + \Delta t \sum_{i=1}^{s} b_i f(t_n + c_i \Delta t, k_i), \qquad n = 0, \dots, N$$

where Δt is an equidistant temporal step-length, with x_0 a given initial value.

These methods can be written in the form of a Butcher table (Table 2.2), where $b^T = (b_1, \cdots, b_s)$, $c = A \cdot 1_s$ are s-dimensional vectors and A an $s \times s$-matrix, with each element i, j corresponding to coefficients in the RK-methods. Note that

c	A
	b^T

Table 2.2: The basic Butcher table.

- Methods with only zero entries on and above the diagonal of A, i.e. $a_{ij} = 0$ for $j \geq i$ are called *explicit*.
- Methods with zero entries above and some non-zero entries on the diagonal of A, i.e. $a_i j = 0$ for $j > i$, are called *diagonally implicit*.
- Methods with non-zero entries on and above the diagonal of A are called *implicit*.
- k_i is called *stage* i.
- Methods are called *irreducible* if they do not have equivalent stages.

2.4.4 Structure-Preservation and Symplecticity

Theorem 2.9. *[3, p.101] Let a Runge-Kutta method satisfy*

$$b_i a_{ij} + b_j a_{ji} = b_i b_j, \quad \forall i, j = 1, \dots, s, \tag{2.44}$$

or equivalently

$$M := BA + A^T B - b b^T = 0, \quad B = diag\{b\}.$$

Then it conserves quadratic invariants of the ODE.

Remark. This condition is closely related to the notions of algebraic and $B-$*stability*; a RK method is algebraically stable if M is positive semi-definite.

Definition 2.16. [3, p.42] The *adjoint method* of a one-step method Φ_h, denoted Φ_h^*, is the inverse map of the original methods with reversed time step, i.e.

$$\Phi_h^* := \Phi_{-h}^{-1}. \tag{2.45}$$

Φ_h is *symmetric* or *time-reversible* if

$$\Phi_h \circ \Phi_{-h} = I \qquad \text{or equivalently} \qquad \Phi_h = \Phi_{-h}^{-1} = \Phi_h^*; \tag{2.46}$$

in other words, the method is its own adjoint.

Theorem 2.10. *[3, p.147] The adjoint method of an s−stage RK method is itself an s−stage RK method with coefficients*

$$a^*_{ij} = b_{s+1-j} - a_{s+1-i,s+1-j}, \quad b^*_i = b_{s+1-i} \quad \forall i,j.$$

The RK method is symmetric if

$$a_{s+1-i,s+1-j} + a_{ij} = b_j \quad \forall i,j.$$

Definition 2.17. [3, p.187] A numerical one-step method Φ_h is a *symplectic method* if the mapping of the method with

$$x_{n+1} = \Phi_h(x_n)$$

is symplectic when applied to a smooth Hamiltonian system, i.e.

$$\left(\partial_{x_n} x_{n+1}\right)^T J \left(\partial_{x_n} x_{n+1}\right) = \Phi_h'(x_n)^T J \Phi_h(x_n) = J. \tag{2.47}$$

Theorem 2.11. *For any RK method, the following statements are equivalent:*

- *The method is symmetric for linear problems $\dot{x} = Lx$.*
- *The methods is symplectic for quadratic Hamiltonian problems $H(x) = \frac{1}{2}x^T Cx$, where C is symmetric.*
- *The stability function satisfies $R(-z)R(z) = 1 \; \forall z \in \mathbb{C}$.*

Theorem 2.12. *[3, p.106] No RK method can conserve all cubic or higher order polynomial invariants.*

Another interesting and related result from [6]:

Theorem 2.13. *[6] Numerical methods exactly preserving energy or other conserved quantities[2] beyond quadratic degree of a Hamiltonian ODE cannot be symplectic.*

Remark. Although symplectic methods cannot preserve energy exactly in general, they still have strong energy conservation properties[2, 3, 6].

[2]Here, only invariants are considered.

Chapter 3

Theory

3.1 Structures in Stratonovich SDEs

Consider now autonomous SDEs in the Stratonovich sense (see Section 2.3.2), written in the form

$$\mathrm{d}X(t) = f(X(t))\mathrm{d}t + \sum_{k=1}^{m} g_k(X(t)) \circ \mathrm{d}W(t), \qquad X(t_0) = X_0 \in \mathbb{R}^d, \quad t \in [t_0, T]; \tag{3.1}$$

where $W(t)$ is an $m-$dimensional Wiener process, with $f, g_k : \mathbb{R}^d \to \mathbb{R}^d \; \forall k = 1, \dots m$. Rewriting (3.1) with

$$\begin{aligned} g(X(t)) &= \{f(X(t)), g_1(X(t)), \dots, g_m(X(t))\} \text{ and} \\ W(t) &= \{t, W_1(t), \dots, W_m(t)\}, \end{aligned}$$

i.e. $g_0 := f$, $g : \mathbb{R}^d \to \mathbb{R}^{d\times(m+1)}$ and $W_0 := t$, it follows that

$$\mathrm{d}X(t) = \sum_{k=0}^{m} g_k(X(t)) \circ \mathrm{d}W_k(t) = g(X(t)) \circ \mathrm{d}W(t). \tag{3.2}$$

3.1.1 Invariants

The concept of invariants or first integrals from Section 2.4.1 can easily be extended to the Stratonovich framework [1]:

Definition 3.1. A function $I \in C^{(1)}(\mathbb{R}^d, \mathbb{R})$ of (3.2) is a *stochastic invariant* if it holds that

$$\nabla I^T g_k(X(t)) = 0 \quad \forall t \in [t_0, T] \quad \text{for } k = 0, \dots, m. \tag{3.3}$$

Remark. Linear and quadratic stochastic invariants have the same form as those for deterministic systems - see Section 2.4.1.

As for ODEs, SDEs might have several invariants - see e.g. [28] or Section 5. In Section 4.1.3, conditions for which numerical schemes preserve invariants are presented.

3.1.2 Hamiltonians

Similarly to the invariants of the previous section, Hamiltonian functions in the deterministic setting can be extended to Stratonovich SDEs[1].

Definition 3.2. A *stochastic Hamiltonian system* takes the form

$$\mathrm{d}X(t) = J\sum_{k=0}^{m}\nabla H_k(X(t))\circ\mathrm{d}W_k(t) = J\nabla H\circ\mathrm{d}W(t),\quad X(t_0)=X_0\in\mathbb{R}^{2d},\quad J=\begin{bmatrix}0 & I_d\\ -I_d & 0\end{bmatrix},\tag{3.4}$$

or, alternatively, for $X(t)=(q(t)^T,p(t)^T)^T$:

$$\begin{aligned}\mathrm{d}q(t) &= \sum_{k=0}^{m}\nabla_p H_k(q,p)\circ\mathrm{d}W_k(t),\quad q(t_0)=q_0\in\mathbb{R}^d,\\ \mathrm{d}p(t) &= -\sum_{k=0}^{m}\nabla_q H_k(q,p)\circ\mathrm{d}W_k(t)\quad p(t_0)=p_0\in\mathbb{R}^d;\end{aligned}\tag{3.5}$$

Here, $H:\mathbb{R}^{2d}\to\mathbb{R}^{m+1}$, and consequently $\nabla H:\mathbb{R}^{2d}\to\mathbb{R}^{2d\times(m+1)}$.

Remark. Here, the Stratonovich sense is preferred to the Itô as the calculus is slightly nicer and closer to the deterministic case; it is also more conventional in physics, which offers many applications of the theory.

Lemma 3.1. *For a stochastic process $X(t)$ as given by* (3.4)*, we have that*

$$H(X_0)=H(X(t))\quad\forall t\in[0,T].\tag{3.6}$$

Proof. Let $Y(t)=H(X(t))$. Then, from the Stratonovich chain rule (Theorem (2.5)), we have that

$$\mathrm{d}Y=\partial_t H(X)\mathrm{d}t+\nabla H(X)\cdot\mathrm{d}X=(\nabla H)^T J\nabla H\circ\mathrm{d}W(t)=0,$$

due to the skew-symmetry of J. □

Corollary 3.1.1.

$$E[H_0]=E[H(X_0)]=E[H(X(t))]\quad\forall t\in[0,T]\tag{3.7}$$

Proof. By the additive property of expectiation and the previous result, we have that

$$\mathrm{E}[H(X(t))]-\mathrm{E}[H_0]=\mathrm{E}[H(X(t))-H_0]=\mathrm{E}[0]=0.$$

□

Symplecticity

Definition 3.3. Let $x,y\in\mathbb{R}^d$. Then, the *wedge product* of x and y is given as

$$u\wedge v=u\otimes v-v\otimes u,\quad(x\wedge y)\in\mathbb{R}^{d\times d};$$

where $\otimes$ denotes the outer or Kronecker product.

Remark. The wedge product results in a skew-symmetric matrix.

Extending the result of Theorem 2.7 for deterministic systems, Milstein [20, p.313] notes that for SDEs of the form (3.5),

$$\mathrm{d}p(t)\wedge\mathrm{d}q(t)=\mathrm{d}p_0\wedge\mathrm{d}q_0,\tag{3.8}$$

i.e. that the phase flow of the differential equation preserve its symplectic structure.

3.2 Deterministic Hamiltonians With Additive Noise

In the papers [2, 15, 30, 31], a slightly different setting than seen in the previous sections is studied. Here, the drift term of the equation is the Hamiltonian function of an autonomous deterministic system, cf. (2.39). This is combined with an additive (m—dimensional) noise term only appearing in the p vector, which can be expressed through a constant matrix. The resulting SDE takes the form

$$\mathrm{d}q = \nabla_p H(q,p)\mathrm{d}t, \tag{3.9a}$$
$$\mathrm{d}p = -\nabla_q H(q,p)\mathrm{d}t + \Sigma\,\mathrm{d}W(t), \qquad \Sigma \in \mathbb{R}^{d\times m}. \tag{3.9b}$$

Letting $X(t) = (q(t), p(t))^T$, $q, p \in \mathbb{R}^d$ and $\Sigma^* = [0,1]^T \otimes \Sigma$ results in the equation

$$\mathrm{d}X(t) = J\nabla H(X(t))\mathrm{d}t + \Sigma^*\mathrm{d}W(t) \qquad X(t_0) = X_0 \in \mathbb{R}^{2d}, \quad t \in [t_0, T]. \tag{3.10}$$

Remark. From Corollary 2.4.1, it should be clear that the Stratonovich and Itô formulations of the system (3.10) coincide.

Remark. The Hamiltonian functions in this section are *not* invariants. This follows readily, as $\nabla H^T \mathrm{d}X = \nabla H^T (J\nabla H\mathrm{d}t + \Sigma^*\mathrm{d}W) = \nabla H \Sigma^* \mathrm{d}W \neq 0$.

Treating (3.10) as an Itô SDE, Burrage and Burrage [2] considers the special case when $m = d$ and $\Sigma = \mathrm{diag}\{\sigma\}, \sigma \in \mathbb{R}^d$. For this case, they derive the following result for separable Hamiltonians.

Theorem 3.2. *For SDEs of the form* (3.9) *with separable H such that*

$$H(q,p) = \frac{1}{2}p^T p + V(q), \tag{3.11}$$

where $V : \mathbb{R}^d \to \mathbb{R}$ is a smooth potential function, the expectation of the Hamiltonian grows linearly in time:

$$E[H(t)] = E[H_0] + \frac{1}{2}\sigma^T\sigma\,(t - t_0). \tag{3.12}$$

Proof. Let $Y(t) = H(X(t))$. Then, from Itô's Formula (Theorem 2.2), we get that

$$\begin{aligned}\mathrm{d}Y &= \partial_t H\mathrm{d}t + \nabla H^T\mathrm{d}X + \frac{1}{2}\mathrm{d}X^T\nabla^2 H\mathrm{d}X \\ &= 0 + \nabla H^T(J\nabla H\mathrm{d}t + \Sigma^*\mathrm{d}W) + \frac{1}{2}\mathrm{tr}\{\Sigma^*\nabla^2 H(\Sigma^*)^T\}.\end{aligned}$$

The first term disappears because the equation is autonomous, whereas $\nabla H^T J \nabla H = 0$ due to the skew-symmetry of J. Noting that $\Sigma^* = \begin{bmatrix}0\\1\end{bmatrix} \otimes \Sigma$ and that

$$\nabla^2 H = \begin{bmatrix}\nabla_{qq} & \nabla_{qp}\\ \nabla_{pq} & \nabla_{pp}\end{bmatrix} H = \begin{bmatrix}\nabla_{qq}V(q) & 0\\ 0 & 1\end{bmatrix},$$

the last term reduces to $\frac{1}{2}\mathrm{tr}\{\Sigma\Sigma^T\} = \frac{1}{2}\sigma^T\sigma$. Finally, taking the expectation of $Y(t)$, we get that

$$\begin{aligned}\mathrm{E}[Y(t)] = \mathrm{E}[H(X(t))] &= \mathrm{E}[H(X(t_0))] + \mathrm{E}\left[\int_{t_0}^t \frac{1}{2}\sigma^T\sigma\mathrm{d}\tau\right] + \mathrm{E}\left[\int_{t_0}^t \nabla H(X(\tau))\Sigma^*\mathrm{d}W(\tau)\right] \\ &= \mathrm{E}[H_0] + \frac{1}{2}\sigma^T\sigma(t - t_0),\end{aligned}$$

with the last integral disappearing being a property of Itô integrals (see e.g. Lemma 2.1). □

Remark. It is clear from the proof above that the slightly more general result

$$\mathrm{E}[H(t)] = \mathrm{E}[H_0]\frac{1}{2}\mathrm{tr}\{\Sigma^T\Sigma\}(t-t_0), \quad t \in [t_0, T] \tag{3.13}$$

holds for general matrices $\Sigma \in \mathbb{R}^{d\times m}$ where m and d might share no relation as well; this is also noted in [30].

Somewhat interestingly, the same result holds for Stratonovich system of the form

$$\mathrm{d}q(t) = \nabla_p H \mathrm{d}t, \quad q(t) = q_0, \tag{3.14a}$$
$$\mathrm{d}p(t) = -\nabla_q H \mathrm{d}t + \Sigma \circ \mathrm{d}W(t), \quad p(t) = p_0, \tag{3.14b}$$

which can be stated as a theorem:

Theorem 3.3. *Let $q(t), p(t)$ satisfy* (3.14) *for a separable Hamiltonian function*

$$H(q,p) = \frac{1}{2}p^T p + V(q), \tag{3.15}$$

where $V : \mathbb{R}^d \to \mathbb{R}$ is a smooth potential function. Then, the expectation of the Hamiltonian function grows linearly in time, i.e.

$$E[H(q(t),p(t))] = E[H_0] + \frac{1}{2}\{\Sigma^T\Sigma\}(t-t_0), \quad t \in [t_0, T]. \tag{3.16}$$

Proof. From the assumptions, H is a function of the Stratonovich process $X(t) = (q(t)^T, p(t))^T$. Then, applying the Stratonovich Chain Rule (See Theorem 2.5):

$$\begin{aligned}\mathrm{d}H(X(t)) &= \nabla_x H \cdot \mathrm{d}X(t)\\ &= (\nabla_q H)^T \nabla_p H \mathrm{d}t + (\nabla_p H)^T(-\nabla_q H)\mathrm{d}t + (\nabla H)^T \Sigma \circ \mathrm{d}W(t)\\ &= (\nabla_p H)^T \Sigma \circ \mathrm{d}W(t).\end{aligned}$$

Then, using the form of H from (3.15) and the Stratonovich Conversion Formula (see e.g. (2.11))

$$(\nabla_p H)^T \Sigma \circ \mathrm{d}W(t) = p^T \Sigma \circ \mathrm{d}W(t) = p^T \Sigma \mathrm{d}W(t) + \frac{1}{2}\mathrm{d}p^T \Sigma \mathrm{d}W(t). \tag{3.17}$$

Noting that

$$\mathrm{d}p(t) = -\nabla_q H \mathrm{d}t + \Sigma \circ \mathrm{d}W(t) = -\nabla_q H \mathrm{d}t + \Sigma \mathrm{d}W(t),$$

following from Corollary 2.4.1,

$$\begin{aligned}\mathrm{d}p^T \Sigma \mathrm{d}W(t) &= \left(-\nabla_q H \mathrm{d}t + \Sigma \mathrm{d}W(t)\right)^T \Sigma \mathrm{d}W(t)\\ &= \mathrm{tr}\{\Sigma^T\Sigma\}\mathrm{d}t,\end{aligned}$$

treating combinations of $\mathrm{d}t$ and $\mathrm{d}W(t)$ as before. Inserting this into (3.17) and taking the expectation of the integral equation for $H(t)$ results in

$$\begin{aligned}\mathrm{E}[H(t)] &= \mathrm{E}[H_0] + \mathrm{E}\left[\int_{t_0}^{t} p(\tau)^T \Sigma \mathrm{d}W(\tau)\right] + \mathrm{E}\left[\int_{t_0}^{t} \frac{1}{2}\mathrm{tr}\{\Sigma^T\Sigma\}\mathrm{d}\tau\right]\\ &= \mathrm{E}[H_0] + \frac{1}{2}\mathrm{tr}\{\Sigma^T\Sigma\}(t-t_0),\end{aligned}$$

concluding the proof. □

Chapter 4

Method

This thesis is concerned mainly with Runge-Kutta methods adapted to SDEs. The equations discussed in Sections 3.1 and 3.2 does this in slightly different ways. For the chapter in general, let Δt refer to temporal step-size such that $\Delta t = (T - t_0)/N$, where t_0 is start time, T stop time and N number of time-steps.

4.1 Numerical Methods for Stratonovich SDEs

The methods in this section are designed to solve the general Stratonovich SDE (3.2), whose most compact form is repeated here:

$$\mathrm{d}X(t) = g(X(t)) \circ \mathrm{d}W(t), \tag{4.1}$$

where $g \in \mathbb{R}^{d\times(m+1)}$ and $W(t) = \{t,\ W_1(t), W_2(t), \ldots, W_m(t)\}$ with each component $W_k(t)$ independent Wiener processes. Here, $\Delta W_n \in \mathbb{R}^{m+1}$ represents the n—th time-step simulated Wiener increment s.t. $\Delta W_n^k \sim \sqrt{\Delta t} \cdot \mathcal{N}(0,1)$ for $k = 1, \ldots, m$ and $\Delta W_n^0 = \Delta t\ \forall n$. Additionally, $X(t_0 + nh) = X(t_n) \approx X_n$, X_n refers to approximate solution in nth timestep and $X(t_0) = X_0$ are given initial values as in Section 2.4.3.

4.1.1 Well-Known Numerical Methods

The possibly most widely known numerical method for solving SDEs is the Euler-Maruyama method, which for Itô SDEs takes the form

$$X_{n+1} = X_n + g(X_n)\Delta W_n. \tag{4.2}$$

For Stratonovich SDEs, using the standard conversion formula between Itô and Stratonovich SDEs (see Theorem 2.4), the adapted method becomes

$$X_{n+1} = X_n + g(X_n)\Delta W_n + \frac{1}{2}\left(\nabla_x g^*(X_n)\right)^T g^*(X_n)\Delta t, \tag{4.3}$$

with g^* referring to g without the diffusion term g_0. It can be compared to the Milstein method, which for Stratonovich SDEs has the formula

$$X_{n+1} = X_n + g(X_n)\Delta W_n + \frac{1}{2}(\Delta W_n)^T(\nabla_x g^*(X_n))^T g^*(X_n)\Delta W_n. \tag{4.4}$$

Some remarks:

- Both the Euler-Maruyma and the Milstein method are explicit methods.
- When applied to deterministic ODEs, both the Euler-Maruyama and the Milstein scheme are reduced to the Euler scheme
$$x_{n+1} = x_n + f(x_n)\Delta t.$$
- The Euler-Maruyama and the Milstein have mean square order 0.5 and 1.0 for scalar noise SDEs, respectively; for multiple noise terms, both reduce to order 0.5.

- Neither the Euler-Maruyama nor the Milstein scheme can conserve quadratic invariants - see Section 4.1.3 for more on the subject matter.

4.1.2 Stochastic Runge-Kutta methods

The Stochastic Runge-Kutta methods for solving (4.1) treated in [1] were first presented in [22]. Trees are not treated in this thesis, and the following simplification is made [1]: The random variables of [22, Section 3], denoted θ_k in said paper, are set to $\theta_k = \Delta W_n^k$ for $k = 1,\dots,m$,; temporal step-size is as before denoted $\Delta t = \Delta W_n^0$. This leads to the following definition.

Definition 4.1. *Stochastic Runge-Kutta methods* for solving (4.1) take the form

$$\begin{aligned} Y_i &= X_n + \sum_{j=1}^{s}\sum_{k=0}^{m} a_{ij}^{(k)} g_k(Y_j)\Delta W_n^k, \quad i = 1,\dots,s, \\ X_{n+1} &= X_n + \sum_{i=1}^{s}\sum_{k=0}^{m} b_i^{(k)} g_k(Y_i)\Delta W_n^k. \end{aligned} \tag{4.5}$$

Remark. The simplified version above means that for equations of the form (4.1) in general, only ms convergence of order 1.0 and 0.5 can be achieved for the scalar and non-scalar noise case, respectively[1, p.45]. [1] gives conditions for this, as well as for weak convergence of order 1.0 and 2.0.

c	$A^{(0)}$	$A^{(1)}$	$\cdots$	$A^{(m)}$
	$b^{(0)T}$	$b^{(1)T}$	$\cdots$	$b^{(m)T}$

(a) General SRK method

	$\frac{1}{2}$	$\cdots$	$\frac{1}{2}$
	1	$\cdots$	1

(b) Stochastic Implicit Midpoint method.

Table 4.1: Butcher Tables of Stochastic Runge-Kutta methods corresponding to (4.5).

As for the RK methods presented in Section 2.4.3, it is possible to express an SRK method through Butcher tables, as seen in 4.1a. A simple example of such a method is the *Stochastic Implicit Midpoint* method (or just the midpoint method for short):

$$X_{n+1} = X_n + \sum_{k=0}^{m} g_k\left(\frac{1}{2}(X_n + X_{n+1})\right)\Delta W_n^k = X_n + g\left(\frac{1}{2}(X_n + X_{n+1})\right)\Delta W_n. \tag{4.6}$$

The (4.6) corresponds to the Butcher table seen in Table 4.1b. Notably, the midpoint method, like the Milstein method, is of strong order 1 for scalar noise and 0.5 for multiplicative noise[1, p. 46], and while the two other Euler-Maruyama and Milstein are explicit, the midpoint method is implicit. To solve the non-linear system associated with the methods, a procedure of simplified Newton iterations is applied[35, p. 119-121].

Hong and collaborators also study a variant class of SRK methods, which has a formula independent of the dimension of the Wiener process[1, p.40]. This class of methods will be called *alternative SRK methods* here; a similar simplification as for Definition 4.1 is made here.

$$\begin{aligned} Y_i^{(0)} &= X_n + \sum_{j=1}^{s}\left(a_{ij}^{(0)} g_0(Y_j^{(0)})\Delta t + \sum_{k=1}^{m} b_{ij}^{(0)} g_k(Y_j^{(k)})\Delta W_n^k\right) \quad \text{for } i = 1,\dots,s; \\ Y_i^{(1)} = \cdots = Y_i^{(m)} &= X_n + \sum_{j=1}^{s}\left(a_{ij}^{(1)} g_0(Y_j^{(0)})\Delta t + \sum_{k=1}^{m} b_{ij}^{(1)} g_k(Y_j^{(k)})\Delta W_n^k\right) \quad \text{for } i = 1,\dots,s; \\ X_{n+1} &= X_n + \sum_{i=1}^{s}\left(\beta_i^{(0)} g_0(Y_i^{(0)})\Delta t + \sum_{k=1}^{m} \beta_i^{(1)} g_k(Y_i^{(k)})\Delta W_n^k\right). \end{aligned} \tag{4.7}$$

[1] The same simplification is made in [1, p.45]

	$A^{(0)}$	$B^{(0)}$
	$A^{(1)}$	$B^{(1)}$
	$\beta^{(0)}$	$\beta^{(1)}$

(a) Alternative SRK Butcher table.

	$\frac{1}{2}$	$1-\theta$
	θ	$\frac{1}{2}$
	1	1

(b) θ method (Scheme 5.2 from [1]) Butcher table.

Table 4.2: Alternative Stochastic Runge-Kutta method Butcher table formulation corresponding to (4.7).

This gives rise to a slightly different Butcher table variant, as seen in Table 4.2a. A basic example of such a method is proposed by Hong et al. (Scheme 5.2, see Table 4.2b) [1], where θ is an arbitrary parameter:

$$\begin{aligned} Y^{(0)} &= X_n + \frac{1}{2} g_0(Y^{(0)})\Delta t + (1-\theta)\sum_{k=1}^{m} g_k(Y^{(k)})\Delta W_n^k; \\ Y^{(k)} &= X_n + \theta\, g_0(Y^{(0)})\Delta t + \frac{1}{2}\sum_{k=1}^{m} g_k(Y_m^{(k)})\Delta W_n^k \quad \text{for } k = 1,\dots,m; \\ X_{n+1} &= X_n + \sum_{k=0}^{m} g_k(Y^{(k)})\Delta W_n^k. \end{aligned} \tag{4.8}$$

Incidentally, setting $\theta = \frac{1}{2}$ returns the midpoint method.

4.1.3 Structure Preservation

Consider again (3.4), whose compact form is restated here:

$$\mathrm{d}X(t) = J\nabla H(X)\circ \mathrm{d}W(t), \quad X(t_0) = X_0. \tag{4.9}$$

Symplecticity

As noted in Section 3.1.2, the flow of (4.9) preserves the symplectic structure $\mathrm{d}X(t)\wedge J\mathrm{d}X(t)$. The definition of a symplectic method presented for ODEs can be extended to Stratonovich SDEs as such [1]:

Definition 4.2. A numerical method solving (4.9) is symplectic if preserves the flow of the underlying equation, i.e.

$$\left(\partial_{X_n} X_{n+1}\right) J \partial_{X_n} X_{n+1} = J \quad \Leftrightarrow \quad \mathrm{d}X_{n+1}\wedge J\mathrm{d}X_{n+1} = \mathrm{d}X_n \wedge J\mathrm{d}X_n. \tag{4.10}$$

On the relationship between conserving quadratic invariants and and symplecticity, [1] adds the following result.

Lemma 4.1. *Every SRK method of* (4.5) *which preserves quadratic invariants of* (4.9) *is symplectic.*

Conservation of Quadratic Invariants

The following theorems are among the main results of the paper by Hong et al.[1]:

Theorem 4.2. *If the coefficients of an SRK method* (4.5) *satisfy*

$$b_i^{(k)} a_{ij}^{(l)} + b_j^{(l)} a_{ji}^{(k)} - b_i^{(k)} b_j^{(l)} = 0 \quad \forall i,j = 1,\dots,sandk,l = 0,\dots,m; \tag{4.11}$$

then it preserves quadratic invariants of (3.2).

Proof. Let $I(X) = \frac{1}{2}X^T CX$, $C \in \mathbb{R}^{d\times d}$ be a quadratic invariant of (4.5); whence, $x_n^T C x_n = 0$, and $Y_i C g_k(Y_i) = g_k(Y_i)^T C Y_i = 0$ for $i = 1,\dots,s$ and $k = 0,\dots,k$. To show that the methods conserve quadratic invariants, it is sufficient to show that $X_{n+1}^T C X_{n+1} = X_n^T C X_n$. Introducing the short-hand $g_k^i := g_k(Y_i)$, rewrite the first line of (4.5) to solve for X_i, i.e.

$$X_n = Y_i - \sum_{i=1}^{s}\sum_{k=0}^{m} a_{ij}^{(k)} g_k^i \Delta W_n^k. \tag{4.12}$$

Then, for $n = 0,\dots,N$

$$\begin{aligned}
X_{n+1}^T C X_{n+1} &= \left(X_n + \sum_{i=1}^{s}\sum_{k=0}^{m} b_i^{(k)} g_k^i \Delta W_n^k\right)^T C \left(X_n + \sum_{i=1}^{s}\sum_{k=0}^{m} b_i^{(k)} g_k^i \Delta W_n^k\right) \\
&= X_n^T C X_n \\
&\quad + \sum_{i=1}^{s}\sum_{k=0}^{m}\left(b_i^{(k)} g_k^i \Delta W_n^k\right)^T C \underbrace{\left(Y_i - \sum_{j=1}^{s}\sum_{l=0}^{m} a_{ij}^{(l)} g_l^j \Delta W_n^l\right)}_{=X_n} \\
&\quad + \sum_{j=1}^{s}\sum_{l=0}^{m}\underbrace{\left(Y_j - \sum_{i=1}^{s}\sum_{k=0}^{m} a_{ji}^{(k)} g_k^j \Delta W_n^l\right)^T}_{=X_n} C \left(b_j^{(l)} g_l^i \Delta W_n^l\right) \\
&\quad + \sum_{i,j=1}^{s}\sum_{k,l=0}^{m}\left(b_i^{(k)} g_k^i \Delta W_n^k\right)^T C \left(b_j^{(l)} g_k^i \Delta W_n^k\right) \\
&= X_n^T C X_n - \sum_{i,j=1}^{s}\sum_{k,l=0}^{m}\underbrace{\left(b_i^{(k)} a_{ij}^{(l)} + a_{ji}^{(k)} b_j^{(l)} - b_i^{(k)} b_j^{(l)}\right)}_{=0 \text{ by assumption}} \Delta W_n^k \Delta W_n^l \left(g_k^i\right)^T C g_l^j \\
&= X_n^T C X_n.
\end{aligned}$$

□

Remark. The original result of [1] is slightly more general, referring to the class of methods presented in [22] rather than the simplified one presented here; however, the results presented here are still valid.

Theorem 4.3. *If the coefficients of the alternative SRK method* (4.7) *satisfy*

$$\begin{aligned}
&\beta_i^{(0)} a_{ij}^{(0)} + \beta_j^{(0)} a_{ji}^{(0)} - \beta_i^{(0)} \beta_j^{(0)} = 0, \\
&\beta_i^{(0)} b_{ij}^{(0)} + \beta_j^{(1)} a_{ji}^{(1)} - \beta_i^{(0)} \beta_j^{(1)} = 0, \qquad \forall i,j = 1,\dots,s, \\
&\beta_i^{(1)} b_{ij}^{(1)} + \beta_j^{(1)} b_{ji}^{(1)} - \beta_i^{(1)} \beta_j^{(1)}, = 0
\end{aligned} \tag{4.13}$$

then it preserves quadratic invariants of (3.2).

Proof. This proof follows similar lines as the proof of Theorem 4.2: Let $I(X) = \frac{1}{2}X^T CX$ with symmetric $C \in \mathbb{R}^{d\times d}$ and let $g_k(Y^{(k)_i}) = g_k^i$, whence $(Y_i^{(k)})^T C g_k^i = 0$ for all $k = 0,\dots,m$ and $i = 1,\dots,s$. Solving the two types of stages of (4.7) for X_n, it follows thet

$$X_n = Y_i^{(l)} - \begin{cases} \displaystyle\sum_{i=1}^{s}\left(a_{ij}^{(0)} g_0^j \Delta t + \sum_{k=1}^{m} b_{ij}^{(0)} g_k^j \Delta W_n^k\right) & \text{for } l = 0 \text{ and} \\ \displaystyle\sum_{i=1}^{s}\left(a_{ij}^{(1)} g_0^j \Delta t + \sum_{k=1}^{m} b_{ij}^{(1)} g_k^j \Delta W_n^k\right) & \text{for } l = 1,\dots,m. \end{cases} \tag{4.14}$$

Then, for $n = 0, \dots, N$

$$
\begin{aligned}
X_{n+1}^T C X_{n+1} &= \left(X_n + \sum_{i=1}^{s} \left(\beta_i^{(0)} g_0^i \Delta t + \sum_{k=1}^{m} \beta_i^{(1)} g_k^i \Delta W_n^k \right) \right)^T C \left(X_n + \sum_{j=1}^{s} \left(\beta_j^{(0)} g_0^j \Delta t + \sum_{l=1}^{m} \beta_j^{(1)} g_l^j \Delta W_n^l \right) \right) \\
&= X_n^T C X_n \\
&+ \sum_{i=1}^{s} \left(\beta_i^{(0)} g_0^i \Delta t \right)^T C \left(\underbrace{Y_i^{(0)} - \sum_{j=1}^{s} \left(a_{ij}^{(0)} g_0^i \Delta t + \sum_{l=1}^{m} b_{ij}^{(0)} g_l^j \Delta W_n^l \right)}_{=X_0} + \sum_{j=1}^{s} \left(\beta_j^{(0)} g_0^j \Delta t + \sum_{l=1}^{m} \beta_j^{(1)} g_l^j \Delta W_n^l \right) \right) \\
&+ \sum_{i=1}^{s} \sum_{k=1}^{m} \left(\beta_i^{(1)} g_k^i \Delta W_n^k \right)^T C \left(\underbrace{Y_i^{(k)} - \sum_{j=1}^{s} \left(a_{ij}^{(1)} g_0^j \Delta t + \sum_{l=1}^{m} b_{ij}^{(1)} g_l^j \Delta W_n^l \right)}_{=X_0} + \sum_{j=1}^{s} \left(\beta_j^{(0)} g_0^j \Delta t + \sum_{l=1}^{m} \beta_j^{(1)} g_l^j \Delta W_n^l \right) \right) \\
&+ \sum_{j=1}^{s} \underbrace{\left(Y_j^{(0)} - \sum_{i=1}^{s} \left(a_{ji}^{(0)} g_0^i \Delta t + \sum_{k=1}^{m} b_{ji}^{(0)} g_k^j \Delta W_n^k \right) \right)^T}_{=X_0} C \left(\beta_j^{(0)} g_0^j \Delta t \right) \\
&+ \sum_{j=1}^{s} \sum_{l=1}^{m} \underbrace{\left(Y_j^{(k)} - \sum_{i=1}^{s} \left(a_{ji}^{(1)} g_0^i \Delta t + \sum_{k=1}^{m} b_{ji}^{(1)} g_k^j \Delta W_n^k \right) \right)^T}_{=X_0} C \left(\beta_j^{(1)} g_l^j \Delta W_n^l \right).
\end{aligned}
$$

Grouping terms with the same coefficients Δt^2, $\Delta W_n^k \Delta t$, $dW_n^l \Delta t$ or $\Delta W_n^k \Delta W_n^l$ results in the following four equations.

$$\sum_{ij=1}^{s} \left(-\beta_i^{(0)} a_{ij}^{(0)} + \beta_i^{(0)} \beta_j^{(0)} - a_{ji}^{(0)} \beta_j^{(0)} \right) (\Delta t)^2 (g_0^i)^T C g_0^j, \tag{4.15a}$$

$$\sum_{i,j=1}^{s} \sum_{l=1}^{m} \left(\beta_i^{(0)} \beta_j^{(1)} - \beta_i^{(0)} b_{ij}^{(0)} - a_{ji}^{(1)} \beta_j^{(1)} \right) \Delta t \Delta W_n^l (g_0^i)^T C g_l^j, \tag{4.15b}$$

$$\sum_{i,j=1}^{s} \sum_{k=1}^{m} \left(\beta_i^{(1)} \beta_j^{(0)} - \beta_i^{(1)} a_{ij}^{(1)} - b_{ji}^{(0)} \beta_j^{(0)} \right) \Delta W_n^k \Delta t (g_k^i)^T C g_0^j, \tag{4.15c}$$

$$\sum_{i,j=1}^{s} \sum_{k,l=1}^{m} \left(\beta_i^{(1)} (-b_{ij}^{(1)}) + \beta_i^{(1)} \beta_j^{(1)} - b_{ji}^{(1)} \beta_j^{(1)} \right) \Delta W_n^k \Delta W_n^l (g_k^i)^T C g_l^j. \tag{4.15d}$$

By the initial assumption of the theorem (and a slight reshuffle of the terms in (4.15b) and (4.15c)), it should be clear that all the groups are equal to zero. Thus,

$$X_{n+1}^T C X_{n+1} = X_n^T C X_n.$$

□

For the midpoint method (4.6) / Table 4.1b, we have that

$$1 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = 1 \cdot 1,$$

meaning that it it should preserve quadratic invariants.

4.2 Numerical Methods for Additive Drift Problems

This section concerns numerical methods adapted to the problem of Section 3.2, based on [2]:

$$\mathrm{d}q = \nabla_p H(q,p)\mathrm{d}t, \tag{4.16a}$$
$$\mathrm{d}p = -\nabla_q H(q,p)\mathrm{d}t + \Sigma\,\mathrm{d}W(t), \qquad \Sigma \in \mathbb{R}^{d\times m} \tag{4.16b}$$

or equivalently

$$\mathrm{d}X(t) = J\nabla H(X)\mathrm{d}t + \Sigma^* \circ \mathrm{d}W(t), \quad \Sigma^* = [0,1]^T \otimes \Sigma. \tag{4.17}$$

where H is a separable Hamiltonian function of the form $H(q,p) = \frac{1}{2}p^T p + V(q)$ for some potential function $V \in C^{(1)}(\mathbb{R}^d, \mathbb{R})$.

Burrage and Burrage consider in [2, 15] the numerical framework where an RK method is applied to (4.17). First, they disprove that treating the whole of (4.17) as an ODE work to preserve the drift in expectation of the Hamiltonian. To counter this problem, they split the vector field of (4.17) into a field of pure drift(deterministic) and a field of pure diffusion(stochastic), i.e.

$$\begin{aligned} \mathrm{d}X(t) &= \mathrm{d}X^{[0]}(t) + \mathrm{d}X^{[1]}(t), \quad \text{where} \\ \mathrm{d}X^{[0]}(t) &= J\nabla H(X)\mathrm{d}t =: f(X)\mathrm{d}t \quad \text{and} \\ \mathrm{d}X^{[1]}(t) &= \Sigma^*\mathrm{d}W(t). \end{aligned} \tag{4.18}$$

Now, they propose the following three-step procedure:

$$\tilde{X}_n = X_n + \frac{1}{\sqrt{2}}\Sigma^*\Delta W_{n,1} \quad =: \Phi_{h/2}^{[EM]}(X_n), \tag{4.19a}$$
$$\tilde{X}_{n+1} = \Phi_h^{[IRK]}(\tilde{X}_n), \tag{4.19b}$$
$$X_{n+1} = \tilde{X}_{n+1}\frac{1}{\sqrt{2}}\Sigma^*\Delta W_{n,2} \quad = \Phi_{h/2}^{[EM]}(\tilde{X}_{n+1}). \tag{4.19c}$$

Here, $\Delta W_{n,i}^k \sim \sqrt{\Delta t/2}\cdot N(0,1)$ for $i = 1,2$, while $\Phi_h^{[IRK]}(\tilde{X}_n)$ corresponds to numerically integrate one step in the deterministic vector field $\mathrm{d}X^{[0]}$ with a symmetric IRK method. In [2], four symmetric IRK methods are applied, namely the implicit midpoint method, Kahan's method, the MQ method [2] and the Lobatto IIIA method of order 4. For the four methods considered here, the Butcher tables can be

$$\begin{array}{c|cc} 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \hline & 0 & 1 \end{array}$$

(a) Implicit midpoint method

$$\begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ \frac{1}{2} & -\frac{1}{4} & 1 & -\frac{1}{4} \\ 1 & -\frac{1}{2} & 2 & -\frac{1}{2} \\ \hline & -\frac{1}{2} & 2 & -\frac{1}{2} \end{array}$$

(b) Quadratic Kahan's method

$$\begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{12} & \frac{1}{3} & \frac{1}{12} \\ 1 & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \\ \hline & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \end{array}$$

(c) MQ method

$$\begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{5}{24} & \frac{1}{3} & -\frac{1}{24} \\ 1 & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \\ \hline & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \end{array}$$

(d) Lobatto IIIA

Table 4.3: Butcher tables for the four RK methods considered in the report.

found in Table 4.3. Using the values from the Butcher tables in Table 4.3, one step of each IRK-method,

[2] It is named after McLaren and Quispel, as it belongs to a class of B-series methods that they derived in [8].

i.e. $\Phi_h^{[IRK]}$, can be rewritten as

$$x_{n+1} = x_n + \Delta t f\left(\frac{x_n + x_{n+1}}{2}\right), \quad \text{midpoint} \tag{4.20a}$$

$$x_{n+1} = x_n + \frac{\Delta t}{2}\left(-f(x_n) + 4f\left(\frac{x_n + x_{n+1}}{2}\right) - f(x_{n+1}\right), \quad \text{Kahan} \tag{4.20b}$$

$$x_{n+1} = x_n + \frac{\Delta t}{6}\left(f(x_n) + 4f\left(\frac{x_n + x_{n+1}}{2} + \frac{\theta \Delta t}{8}(f(x_n) - f(x_{n+1}))\right) + f(x_{n+1})\right); \tag{4.20c}$$

where $\theta = 0$ returns the MQ method and $\theta = 1$ returns the Lobatto IIIA method. As all the chosen methods are implicit, the need arises to solve systems of nonlinear equations. To do this, the method described in [35, p.118-120] is applied. For the derivation of the respective systems, see Appendix A.

Using the same kind of notation, note that (4.19a) and (4.19c) corresponds to taking a step of length $h/2$ with the Euler-Maruyama method in the purely stochastic vector field $\mathrm{d}X^{[1]}$. Writing the steps in the short-hand form frequently used in composition methods (see e.g. [3, p.43]) results in the following scheme:

$$X_{n+1} = \Psi_h(X_n) := \Phi_{h/2}^{[EM]} \circ \Phi_h^{[IRK]} \circ \Phi_{h/2}^{[EM]}(X_n). \tag{4.21}$$

Remark. By definitions in [3, p.47, 49] and [36, p.389-390], (4.21) is a composition method with a Strang splitting of the vector field $\mathrm{d}X(t)$. Also, twice as many Wiener increments must be simulated for every simulation compared to the methods of Section 4.1.2.

Theorem 4.4. *For a 2-dimensional linear problem with additive noise*

$$dX(t) = JX dt + \sigma^* dW(t), \quad \sigma^* = [0, \sigma]^T, \, \sigma \in \mathbb{R}, \tag{4.22}$$

The composition method (4.21) *utilizing the methods in* (4.20) *preserves the temporal drift of* $E[H]$ *satisfying* (4.17).

Proof. To show that the method actually preserves the drift of the problem (4.22), it must be demonstrated that

$$\mathrm{E}\left[X_{n+1}^T X_{n+1}\right] = \mathrm{E}[X_{n+1}] + \frac{\sigma^2}{2}\Delta t. \tag{4.23}$$

The composition method (4.21) is written in [2, p.529] as

$$\begin{aligned} \tilde{X}_n &= X_n + \frac{1}{\sqrt{2}} z_1 \sigma^*, \\ \tilde{X}_{n+1} &= R(\Delta t J)\tilde{X}_n, \\ \tilde{X}_{n+1} &= X_n + \frac{1}{\sqrt{2}} z_2 \sigma^*; \end{aligned} \tag{4.24}$$

where $z_1, z_2 \sim \sqrt{\frac{\Delta t}{2}} \cdot N(0,1)$ independently and $R(\Delta t J)$ is a symmetric Padé approximation of order 2 (equations (4.20a) or (4.20c) with $\theta = 0$)

$$R(\Delta t J) = \left(I_2 - \frac{\Delta t}{2} J\right)^{-1}\left(I_2 + \frac{\Delta t}{2} J\right) \tag{4.25}$$

or 4 ((4.20c) with $\theta = 1$),

$$R(\Delta t J) = \left(I_2 - \frac{\Delta t}{2} J + \frac{(\Delta t)^2}{12} J^2\right)^{-1}\left(I_2 + \frac{\Delta t}{2} J + \frac{(\Delta t)^2}{12} J^2\right). \tag{4.26}$$

Composing the steps in (4.24), it follows that $X_{n+1} = R(\Delta tJ)\left(X_n + \frac{1}{\sqrt{2}}\sigma^* z_1\right) + \frac{\sigma^*}{\sqrt{2}} z_2$. Noting that for symmetric Padé-approximations $R(\Delta tJ)^T R(\Delta tJ) = R(\Delta tJ^T)R(\Delta tJ) = R(-\Delta tJ)R(\Delta tJ) = I_2$,

$$
\begin{aligned}
\mathrm{E}\left[X_{n+1}^T X_{n+1}\right] &= \mathrm{E}\left[\left(R(\Delta tJ)\left(X_n + \frac{1}{\sqrt{2}}\sigma^* z_1\right) + \frac{1}{\sqrt{2}}\sigma^* z_2\right)^T \left(R(\Delta tJ)\left(X_n + \frac{1}{\sqrt{2}}\sigma^* z_1\right) + \frac{1}{\sqrt{2}}\sigma^* z_2\right)\right] \\
&= \mathrm{E}\left[\left(X_n + \frac{1}{\sqrt{2}}\sigma^* z_1\right)^T \underbrace{R(\Delta tJ)^T R(\Delta tJ)}_{=I_2} \left(X_n + \frac{1}{\sqrt{2}}\sigma^* z_1\right)\right] \\
&\quad + \mathrm{E}\left[\left(X_n + \frac{1}{\sqrt{2}}\sigma^* z_1\right)^T R(\Delta tJ)^T \frac{\sigma^*}{\sqrt{2}} z_2 + \frac{\sigma^*}{\sqrt{2}} z_2 R(\Delta tJ)\left(X_n + \frac{1}{\sqrt{2}}\sigma^* z_1\right)\right] \\
&\quad + \mathrm{E}\left[\frac{(\sigma^*)^T}{\sqrt{2}} z_2 \frac{\sigma^*}{\sqrt{2}} z_2\right] \\
&= \mathrm{E}\left[X_n^T X_n\right] + \frac{\sigma^2}{2}\underbrace{\mathrm{E}\left[z_1^2\right]}_{=\Delta t/2} + \underbrace{\mathrm{E}\left[X_n^T R(-\Delta tJ)\frac{\sigma^*}{\sqrt{2}} z_2 + \frac{(\sigma^*)^T}{\sqrt{2}} z_2 R(\Delta tJ) X_n\right]}_{=0 \text{ as } \mathrm{E}[z_1]=\mathrm{E}[z_2]=0} \\
&\quad + (R(-\Delta tJ) + R(\Delta tJ))\frac{\sigma^2}{2}\underbrace{\mathrm{E}[z_1 z_2]}_{=0 \text{ as } z_1 \perp z_2} + \frac{\sigma^2}{2}\underbrace{\mathrm{E}[z_2^2]}_{=\Delta t/2} \\
&= \mathrm{E}[X_n^T X_n] + \frac{\sigma^2}{2}\left(\frac{\Delta t}{2} + \frac{\Delta t}{2}\right) \\
&= \mathrm{E}[X_n^T X_n] + \frac{\sigma^2}{2}\Delta t.
\end{aligned}
$$

□

Chapter 5

Numerical Results

In this section, the numerical results used to validate the theoretical properties studied in Chapter 4 are reported. In Section 5.1, the conservative properties of methods presented in Section 4.1 are investigated numerically. In Section 5.2, problems for assessing whether the methods of Section 4.2 preserve the drift in Hamiltonian are presented.

5.1 Quadratic Invariants

Problem \ Method	EM(4.3)	Milstein(**??**)	SymMi(5.4)	Midpoint(4.6)	Scheme 5.2(4.8)
Linear test equation	X	X		X	X
Scalar Kubo Oscillator	X	X	X	X	X
Kubo Oscillator (2 noise terms)	X	X	X	X	X
3-dimensional Rigid Body System	X	X		X	X

Table 5.1: Table over methods used to evaluate the properties described in Section 3.1, whose formulas can be found in Section 4.1.2, with the exception Symplectic Milstein [20, p.342], for which the formula is in Section 5.1

Here, the methods described in Section 4.1 are tested on four different problems: the Linear Test Equation (Section 5.1.1), the Kubo Oscillator with scalar (Section 5.1.2) and double (Section 5.1.3) noise, and a 3-dimensional Rigid Body Sytstem with Scalar noise (Section 5.1.4). For all the problems aside from the Linear Test problem, the plot of the Hamiltonian function resulting from a single realization is made, as well as convergence plots. For the Kubo Oscillator problems, plots of q and p for a single realization approximated by each relevant method are given as well. For all the convergence plots, 10^5 simulations of each step-size are compared to a reference solution in the mean-square sense, calculated exactly for Linear Test Problem and by a Midpoint approximation with step-size $\Delta t = 2^{-11}$ for the rest. In all problems, θ parameter of (4.8) is set to 0.5.

5.1.1 Linear Test problem

The first problem considered, commonly used as a test equation in stochastic numerics (see e.g. [21]), is the linear test problem with scalar noise of the form

$$\mathrm{d}X(t) = \mu X(t)\mathrm{d}t + \sigma X(t) \circ \mathrm{d}W(t), \qquad X(0) = X_0 \in \mathbb{R}^1, \quad W(t) \in \mathbb{R}. \tag{5.1}$$

Figure 5.1 shows the average of 10 simulations with each of the methods as described in the figure text, visually demonstrating convergence towards the same solution. From Figure 5.2, where the mean-square expectation of the global error $\sqrt{\mathrm{E}[|\varepsilon_n(\Delta t)|^2]}$ (see Definition 2.10) is compared with stepsize Δt, the order of the methods can be inferred. Euler-Maruyama achieves as expected only order 0.5,

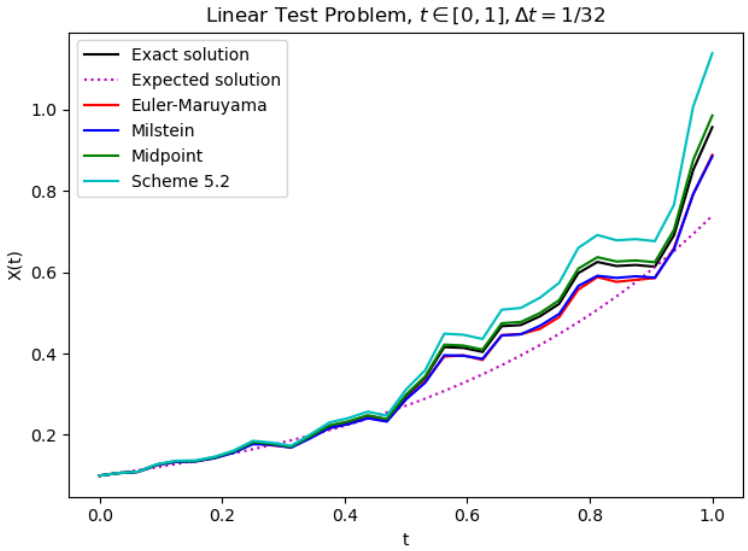


Figure 5.1: Average of 10 simulations of linear test problem with $\mu = 1.5$ and $\sigma = 1$. The dotted line shows the expected solution.

whereas the Midpoint and Milstein method both have order 1.0. As for Scheme 5.2, the error is very large for large stepsizes, but from the smaller stepsizes, it seems to approach order 1.0 as well.

5.1.2 Kubo Oscillator

The one-dimensional Stochastic Harmonic Oscillator or Kubo oscillator [20] with scalar noise can be expressed by the equations

$$\begin{aligned} \mathrm{d}q(t) &= p(t)\mathrm{d}t + \sigma p(t) \circ \mathrm{d}W(t), \quad q(0) = q_0, \\ \mathrm{d}p(t) &= -q(t)\mathrm{d}t - \sigma q(t) \circ \mathrm{d}W(t), \quad p(0) = p_0, \end{aligned} \qquad \sigma, W(t) \in \mathbb{R}, \tag{5.2}$$

or equivalently

$$\mathrm{d}X(t) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} X(t)(1\mathrm{d}t + \sigma \circ \mathrm{d}W(t)), \quad X(0) = X_0 \in \mathbb{R}^2, \quad \sigma, W(t) \in \mathbb{R}. \tag{5.3}$$

It is possibly the easiest example of a quadratic stochastic Hamiltonian system, thus making it especially suited as a basic comparison. As noted in [28], it is a single-integrand problem, which should be apparent from the formulation in (5.3). For both this and the case of two noise terms (Section 5.1.3), the symplectic partitioned RK method described in [20, p.342] and as Symplectic Milstein(SymMI) in [1, p.47] is applied. For (5.2) and (5.6), the formula of [1] becomes

$$\begin{aligned} \mathrm{d}q_{n+1} &= q_n + p_n \Delta t + (\sigma^T \otimes p_n)\Delta W_n - \frac{\sigma^T \sigma}{2} q_{n+1} \Delta t, \\ \mathrm{d}p_{n+1} &= p_n - q_n \Delta t - (\sigma^T \otimes q_n)\Delta W_n + \frac{\sigma^T \sigma}{2} p_{n+1} \Delta t. \end{aligned} \tag{5.4}$$

Here, $\sigma, \Delta W_n \in \mathbb{R}^m$ for $m = 1$ or 2, depending on the number of noises included, and $\Delta W_n^k \sim \sqrt{\Delta t} \cdot N(0,1)$ for $k = 1, 2$.

This is a Hamiltonian SDE, where the Hamiltonian function is

$$H(q,p) = \frac{1}{2}\left(q^2 + p^2\right) = \frac{1}{2}\|X\|_2^2. \tag{5.5}$$

To demonstrate the long-time behaviour, a single realization for each of the methods is made. From (5.5) and (2.41) and with initial values $X_0 = (q,p) = (0,1)$, a conservative numerical method would

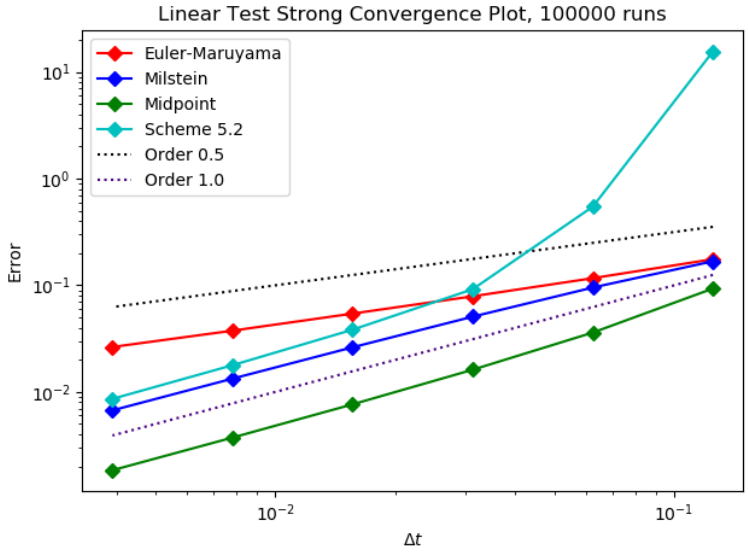
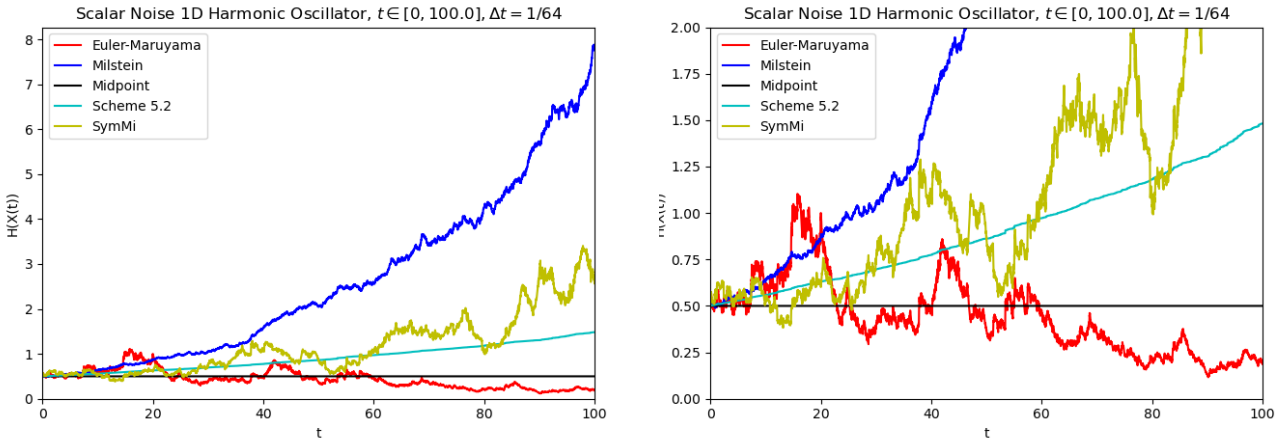


Figure 5.2: Convergence plot with 10^5 simulations of (5.1) with $\mu = 1.5$ and $\sigma = 1$.



(a) Full plot **(b)** Cropped view

Figure 5.3: Hamiltonian function $H(X(t))$ (5.5) for harmonic oscillator with scalar noise (5.2) with $\sigma = 1, t \in [0, 100], \Delta t = 1/64, X_0 = (q, p) = (1, 0)$.

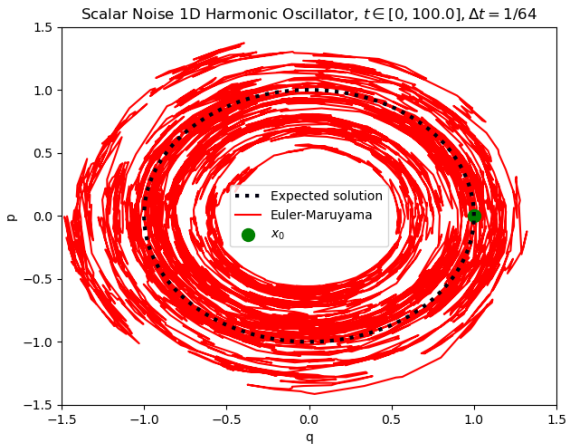
have the solution stay on the unit circle. Plots of one realization can be seen in Figure 5.4, while plots of the resulting Hamiltonian function can be seen in Figure 5.3.

Figure 5.5 shows the methods convergence for the same problem, with stepsizes $\Delta t \in \{2^{-i}\}_{i=3}^{9}$ as before.

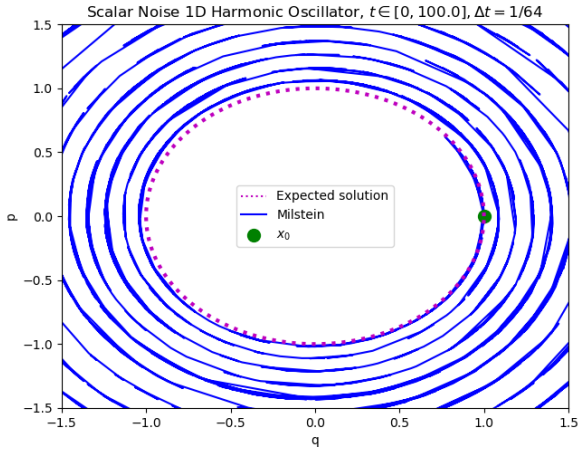
5.1.3 Kubo Oscillator With Two Noise Terms

Adding a second noise term, the problem changes to

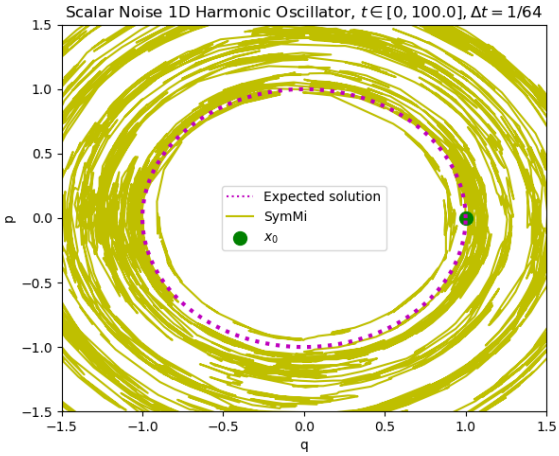
$$\begin{aligned} \mathrm{d}q &= p\mathrm{d}t + \sum_{k=1}^{2} \sigma_k p \circ \mathrm{d}W_k(t), \quad q(0) &&= q_0, \\ \mathrm{d}p &= -q\mathrm{d}t - \sum_{k=1}^{2} \sigma_k q \circ \mathrm{d}W_k(t), \quad p(0) &&= p_0, \end{aligned} \qquad \begin{aligned} W(t) &= (W_1(t), W_2(t))^T \in \mathbb{R}^2, \\ \sigma &= (\sigma_1, \sigma_2)^T \in \mathbb{R}^2 \end{aligned} \tag{5.6}$$



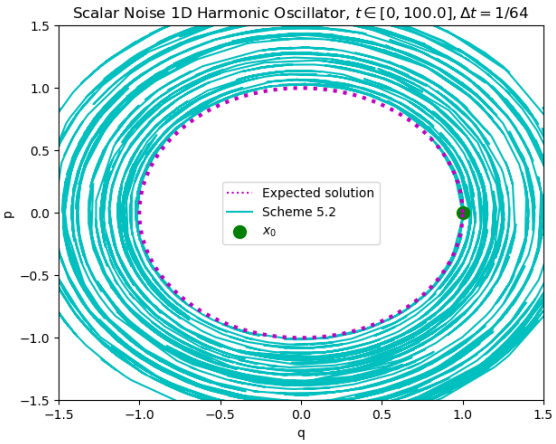
(a) Euler-Maruyama



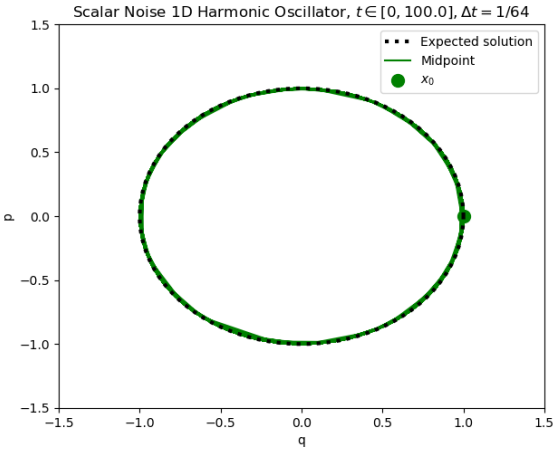
(b) Milstein



(c) Symplectic Milstein



(d) Scheme 5.2



(e) Midpoint

Figure 5.4: Plots solutions of 1 realization of the stochastic harmonic oscillator (5.2) with $\sigma = 1$, $t \in [0, 100]$, $\Delta t = 1/64$, $X_0 = (q, p) = (1, 0)$.

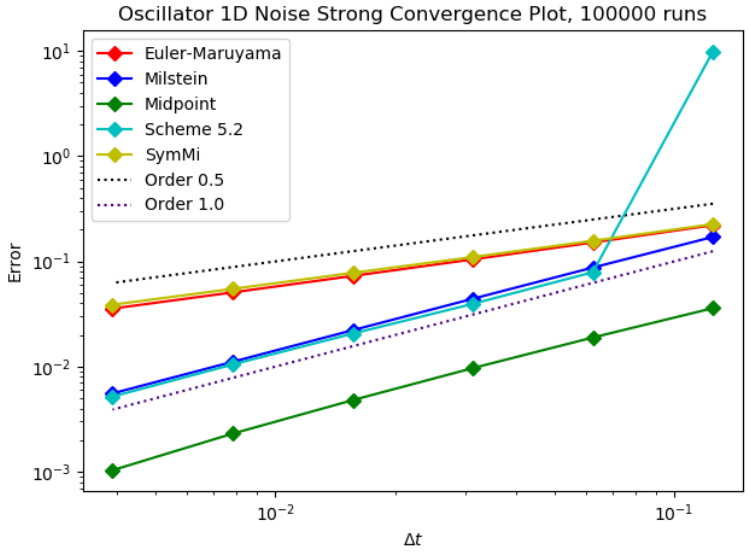


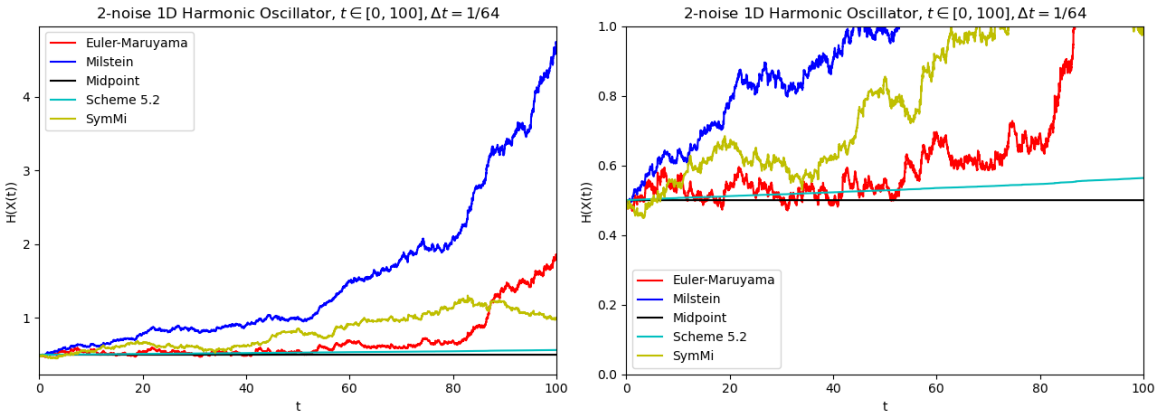
Figure 5.5: Convergence plot of 1-dimensional harmonic oscillator with scalar noise term (5.2) with 10^5 simulations and stepsizes $\Delta t \in \{2^{-i}\}_{i=3}^{9}$.

or

$$\mathrm{d}X = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} X(t)\big(1\mathrm{d}t + (\sigma^T \otimes 1_2) \circ \mathrm{d}W(t)\big), \quad X(0) = X_0, \sigma, W(t) \in \mathbb{R}^2. \tag{5.7}$$

It still retains the same Hamiltonian function (5.5), and is still a single-integrand problem.

In Figures 5.7 and 5.6, a single realization is plotted as coordinates and Hamiltonian function as a function of time, respectively. Figure 5.8 demonstrates the mean square convergence behaviour of the methods applied to this problem.



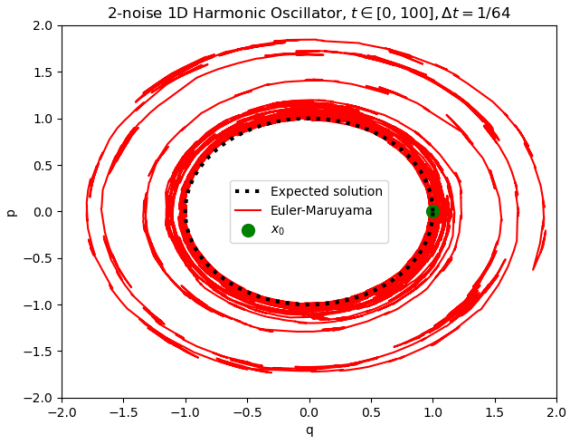
(a) Full view **(b)** Cropped view

Figure 5.6: Hamiltonian function $H(X(t))$ (5.5) for harmonic oscillator (5.6) with $\sigma^T = (0.5, 0.25), t \in [0, 100], \Delta t = 1/64, X_0 = (q, p) = (1, 0)$.

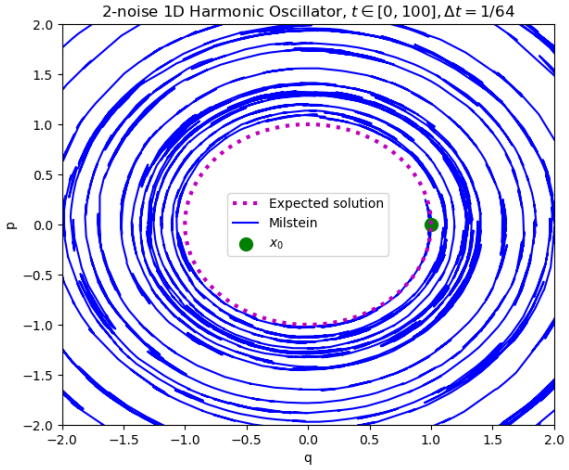
5.1.4 Rigid Body Problem

Next, consider the scalar noise Rigid Body Problem

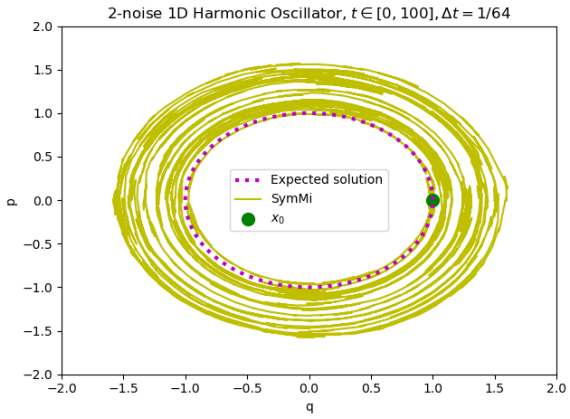
$$\mathrm{d}X(t) = A(X)X(t)\mathrm{d}t + \Sigma X(t) \circ dW(t), \tag{5.8}$$



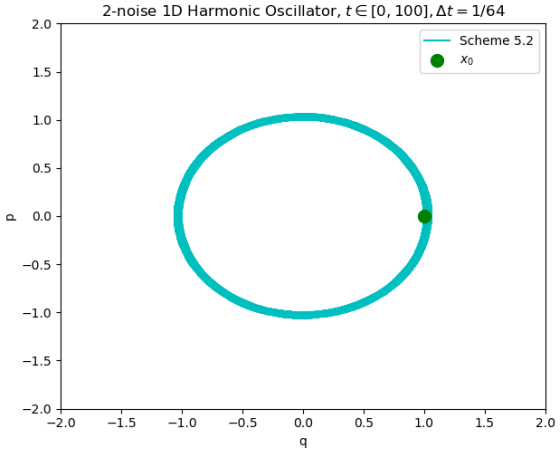
(a) Euler-Maruyama



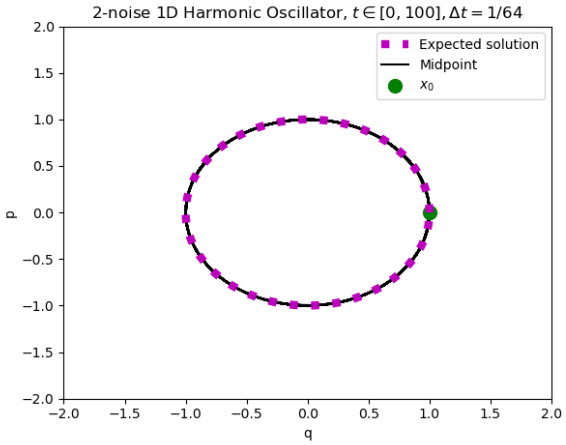
(b) Milstein



(c) Symplectic Milstein



(d) Scheme 5.2



(e) Midpoint

Figure 5.7: Plots solutions of 1 realization of the stochastic harmonic oscillator (5.6) with $(\sigma_1, \sigma_2) = (0,5, 0.25)$, $t \in [0, 100]$, $\Delta t = 1/64$, $X_0 = (q, p) = (1, 0)$.

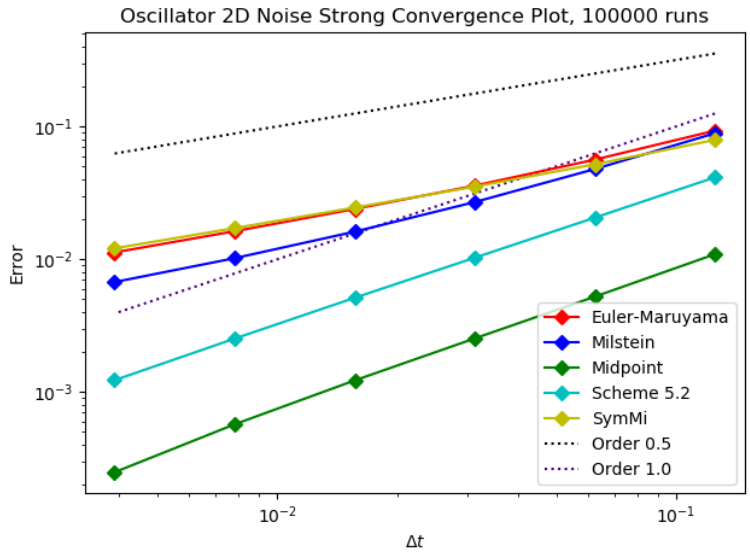


Figure 5.8: Convergence plot for Harmonic Oscillator with 2 noise terms (5.6) over 10^5 simulations, with $[\sigma_1, \sigma_2] = [0.5, 0.25]$, $t \in [0,1]$ and $\Delta t \in \{2^{-i}\}_{i=3}^{9}$.

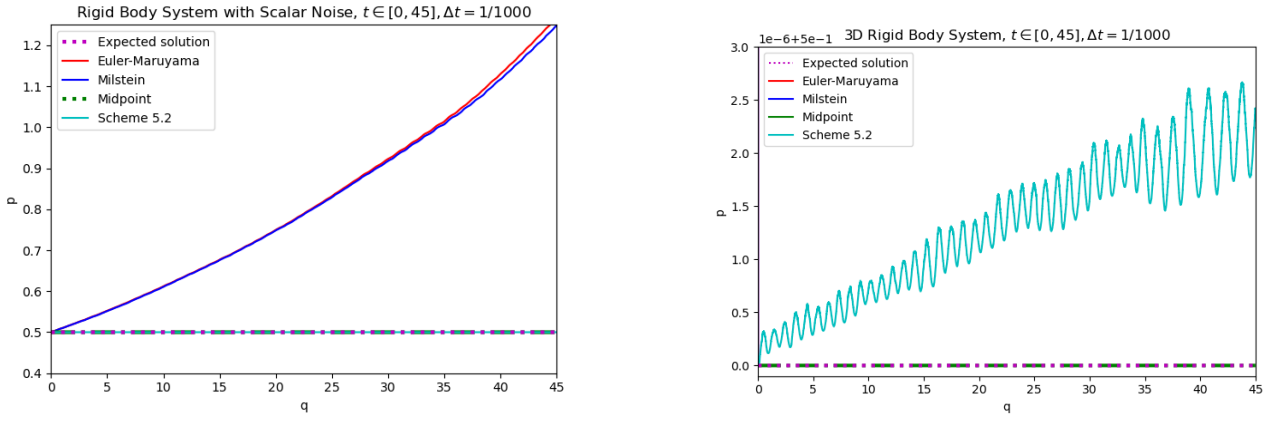
where

$$A = \begin{pmatrix} 0 & X_3(t)/I_3 & -X_2(t)/I_2 \\ -X_3(t)/I_3 & 0 & X_1(t)/I_1 \\ X_2(t)/I_2 & -X_1(t)/I_1 & 0 \end{pmatrix}, \quad \Sigma = \sigma \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}, \quad \begin{aligned} &I_1, I_2, I_3 \in \mathbb{R}^+, \sigma \in \mathbb{R}_0^+, \\ &X(0) = X_0 \in \mathbb{R}^3, W(t) \in \mathbb{R}. \end{aligned}$$

Note that both A and Σ are anti-symmetric matrices. This problem conserves the Casimir

$$C(X(t)) = \frac{1}{2}\|X(t)\|_2^2 = \frac{1}{2}\left(X_1^2 + X_2^2 + X_3^2\right). \tag{5.9}$$

In Figure 5.9, the resulting Hamiltonian plot is shown. Figure 5.10 is a convergence plot for the



(a) Full view **(b)** Cropped view

Figure 5.9: Rigid body problem (5.8) with $(I_1, I_2, I_3) = (0.8, 0.6, 0.2)$, $\sigma = 0.1$, $X_0 = (\cos(1.1), 0, \sin(1.1))^T$ and stepsize $\Delta t = 10^{-3}$.

same problem.

5.2 Drifting Hamiltonians

In this section, problems used in [2, 15] to test efficacy of the composition method described in Section 4.2 are presented. In contrast with the previous section, these methods are expected to work well

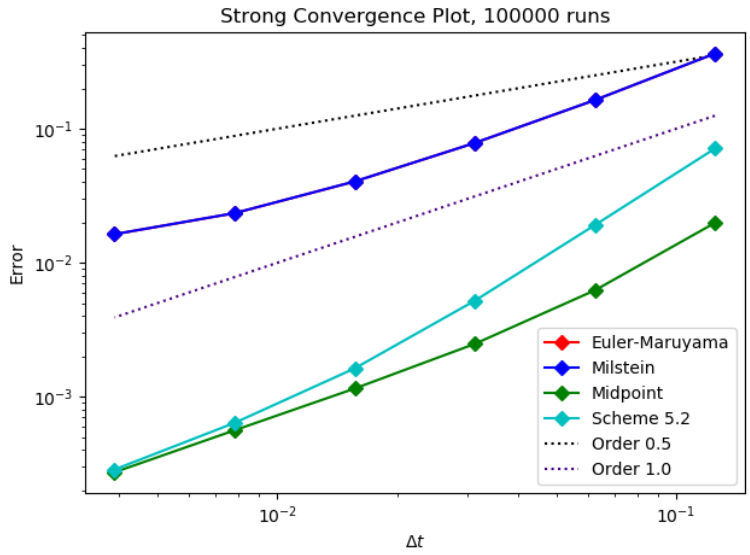


Figure 5.10: Convergence of methods on rigid body problem (5.8) with $(I_1, I_2, I_3) = (0.8, 0.6, 0.2)$, $\sigma = 0.1$, $X_0 = (\cos(1.1), 0, \sin(1.1))^T$, $t \in [0, 1]$.

on Hamiltonian functions of greater polynomial degree than 2. The Double Well problem presented in Section 5.2.1 has a quartic Hamiltonian but $m = d = 1$, i.e. only scalar noise. The Hènon-Heiles problem of Section 5.2.2 has a cubic Hamiltonian, however $m = d = 2$, adding two noises. For both problems, different IRK methods were to be compared with varying parameters σ and Δt, as was done in [2].

Neither of them were successfully implemented; the author offers his apologies.

5.2.1 Double Well

The double well potential problem with $q, p \in \mathbb{R}$ has a Hamiltonian of the form

$$H(q,p) = \frac{1}{2}p^2 + \frac{1}{2}\left(\frac{1}{2}q^4 - q^2\right), \tag{5.10}$$

leading to the SDE

$$\mathrm{d}p = -\left(q^3 - q\right)\mathrm{d}t + \sigma \mathrm{d}W, \qquad \sigma, W(t) \in \mathbb{R}; \tag{5.11a}$$

$$\mathrm{d}q = p, \qquad (q(t_0), p(t_0)) = (q_0, p_0). \tag{5.11b}$$

No results are reported.

5.2.2 Hénon-Heiles

The Hénon-Heiles problem has a Hamiltonian

$$H(q,p) = \frac{1}{2}\left(p_1^2 + p_2^2\right) + \frac{1}{2}\left(q_1^2 + q_2^2\right) + \alpha\left(q_1 q_2^2 - \frac{1}{3}q_1^3\right), \tag{5.12}$$

with the resulting SDE

$$\mathrm{d}p = -\begin{pmatrix} q_1 + \alpha\left(q_2^2 - q_1^2\right) \\ q_2 + 2\alpha q_1 q_2 \end{pmatrix} + \begin{pmatrix} \sigma_1 \\ \sigma_2 \end{pmatrix}\mathrm{d}W, \tag{5.13a}$$

$$\mathrm{d}q = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}. \tag{5.13b}$$

No results are reported.

Chapter 6

Discussion

In this section, the results from Chapter 5 are discussed.

6.1 Kubo Oscillator

6.1.1 Scalar Noise

As expected, neither Euler-Maruyama (Figure 5.4a) nor Milstein (Figure 5.4b) preserves the Hamiltonian, and as demonstrated in [1], neither does the symmetric Milstein scheme (Figure 5.4c). Somewhat surprisingly, nor does Scheme 5.2 from [1], which should be conservative. However, Figure 5.3 indicates that this might be a flaw of the implementation: The solution drifts smoothly away from the expected solution, qualitatively different from the non-conservative methods. Finally, the Stochastic Midpoint Method (Figure 5.4e) demonstrates the expected conservative behaviour.

From Figure 5.5, it appears that all the methods achieve mean-square order 1.0 bar the symmetric Milstein (SymMi) and Euler-Maruyama method, which has mean-square order 0.5. Both results are as expected for scalar noise problems. It should also be noted that Scheme 5.2 has a very large error for the largest step-size, before falling down to error almost identical to that of the Milstein method.

6.1.2 Two Noise Terms

Figures 5.6 and 5.7 demonstrate the same qualitative behaviour as Figures 5.3 and 5.4, albeit with slightly smaller noise impact, which can be explained by the smaller coefficients $\sigma_1 = 0.5$, $\sigma_2 = 0.25$ compared to $\sigma = 1$ for (5.2).

The strong convergence plot of Figure 5.8 demonstrates slightly different behaviour than earlier. Now, the order of the Milstein method is reduced to 0.5, which is expected for multiple noise terms. However, both Scheme 5.2 and the Midpoint method retains order 1.0. To explain this discrepancy from the order expected in [1], consider the following result from [28]:

Theorem 6.1. *For the* single-integrand *SDE*

$$X(t) = x_0 + \int_{t_0}^{t} f(X(\tau)) \circ d\mu(\tau), \quad \mu(\tau) := \lambda\tau + \sigma \cdot W(s), \tag{6.1}$$

consider an SRK method of the form

$$Y_i = X_n + \Delta\mu_n \sum_{j=1}^{s} a_{ij} f(Y_j), \quad \textit{for } i = 1, \ldots, s; \tag{6.2a}$$

$$X_{n+1} = X_n + \Delta\mu_n \sum_{i=1}^{s} b_i f(Y_i), \quad \textit{where } \Delta\mu_n = \lambda\Delta t + \sigma \cdot \Delta W_n \tag{6.2b}$$

If the problem allows a (unique) solution and the single-integrand SRK method of (6.2) *has deterministic order p_d, the method has mean square and weak convergence of order $p_\mu = \lfloor p_d/2 \rfloor$.*

As noted in Section 5.1.2 and 5.1.3, the Kubo oscillator is a single-integrand problem.[1] For the dereministic case, both Euler-Maruyama and Milstein reduces to the Euler method, as noted in Section 4.1.1; as the Euler method has order 1, neither achieves more than order 0.5. The same is true for the Symplectic Milstein method. On the other hand, both Stochastic Midpoint and Scheme 5.2 corresponds to deterministic methods of order 2; consequently, for single-integrand problems, they have the observed order of 1.0.

6.2 Rigid Body System

Notably in Figure 5.9a, the two explicit methods diverge rapidly from the correct value of 0.5, whereas both Scheme 5.2 and the Midpoint method seemingly conserves the invariant. However, upon closer inspection (Figure 5.9b), it is apparent that Scheme 5.2 still moves away from the expected solution. Although the effect happens at a much slower rate than in the other test problems, it most likely reflects the smaller stepsize used (10^{-3} vs 2^{-6} used earlier).

In the mean-square convergence plot of Figure 5.10, the neither of the methods display a consequent polynomial convergence. For stepsizes $\Delta t \geq 2^{-8}$, Euler-Maruyama and Milstein demonstrates a convergence of order 1, before falling down to order 0.5 for $2^{-9} \leq \Delta t \leq 2^{-8}$. Similarly, Scheme 5.2 seems to achieve order 2 for $\Delta t \in [2^{-8}, 2^{-3}]$, before falling down to order 1.0 for $\Delta t \leq 2^{-8}$; the picture is less clear for the Midpoint method.

Most likely, this behaviour reflects two different phenomena: For the larger step-sizes, the error of the deterministic method dominates that of the stochastic method; For the smaller step-sizes, the methods again achieve mean-square order of convergence as expected for single-integrand problems.

6.3 Implementation Issues

6.3.1 Scheme 5.2

There are several indications that Scheme 5.2 of (4.8) suffers from a flawed implementation:

- The very high errors for the large step-sizes seen in Figure 5.2 and 5.5
- The drifting Hamiltonian observed in Figure 5.3 and 5.6, as well as the oscillating drift observed in Figure 5.9b
- The error being consistently larger than that of the Midpoint method, although with $\theta = 0.5$, these should be identical.

Unfortunately, at the time of submission, no correct implementation has been achieved.

6.3.2 Working progress

A significant amount of time has been spent on the implementation of the methods discussed in this thesis. In hindsight, far too much time has been spent on redundant or superfluous code architecture rather than actually producing the needed numerical results. Using a class-based approach, both the problems and their respective [S/I]RK methods were attempted built as input variables or class methods. The reason for this is the ease for which the code would be reusable for other test problems, as well as the simplicity by which new RK methods could be added if desirable. However, this superstructure has more often than not been a hindrance than help, and thus been rewritten far too many times. The consequence is that far fewer numerical results than initially planned can be reported. Hopefully, the author will learn from his mistakes and make more meaningful progress in the Master's thesis.

An advantage for further work is the relatively comprehensive literary review in the Introduction (Chapter 1). Albeit undertaking the review slightly late in the working process, several new insights into the problems here discussed were gathered. Moreover, the comprehensive reference list makes it

[1] It is also one of the test examples of [28].

easier for both reader and author to validate claims made herein and expand on the existing body of work.

Chapter 7

Conclusion

Two slightly different approaches to structure preservation has been considered in this thesis, with slightly different outcomes.

The first approach of the thesis, in studying invariants and in particular quadratic invariants in Stratonovich SDEs, replicates many of the results of [1], as well as demonstrating that certain aspects of the article are lacking: the conservation conditions of the presented SRK methods are proven to be correct; however, the proposed order conditions cannot sufficiently explain the behaviour of the methods for the same test problems used in the same article. Here, however, only the simplest of the methods presented are implemented, and only mean square order of convergence is reported. Immediate improvements would be efficient and functional implementations of more of the methods presented in [1], as well as reporting both weak and mean-square order of convergence for a larger set of step-sizes would be of interest. Additionally, stability and order analysis could be performed.

There are several ways in which the work on invariants in Stratonovich integrals can be continued from this thesis. Based on [27] and [28], using B-series, formulas for higher order conservative methods could be derived and tested. New and more involved test problems could also be explored. Another possible continuation of the work is to consider methods attempting to conserve invariants exactly, such as dicrete gradient and projection methods presented in [29], and compare to the performance of the SRK methods.

The second approach investigated here is the attempt to preserve drift of a deterministic Hamiltonian with additive noise. Some of the theoretical results from the initial paper [2] has been replicated; additionally, the theoretical results for Itô SDEs has been shown to hold true for Stratonovich SDEs as well. However, there are no reported numerical experiments to back up the analysis. To continue investigating this approach, producing valid numerical result would be imperative.

There are immediate ways to continue on the work done here and in [2]. Firstly, Burrage and Burrage does not report convergence order for the composition method - this would be a natural addition. A second addition, which is actually suggested in the same paper, is expanding the analysis on the adaptation of discrete line integral methods for the stochastic setting. Further work on the methods inspiring the initial article [15] has since been done, examples seen in [16, 37]. On the other hand, several studies have already followed [2], some of which have demonstrated significant limitations of the approach - see e.g. [31].

A goal of future investigation could be an attempt of unifying the two approaches already taken here. Would it, for instance, be possible to give conditions under which SRK methods performed well on higher order invariants or possibly on Hamiltonians with additive drift? Conversely, could the discrete line integral approach be adapted to invariants in Stratonovich SDEs, with which methods of arbitrary order could conserve invariants of arbitrary polynomial degree? It appears that there are many aspects yet to examine.

Bibliography

[1] J. Hong, D. Xu and P. Wang, 'Preservation of quadratic invariants of stochastic differential equations via runge-kutta methods,' English, *Applied Numerical Mathematics*, vol. 87, pp. 38–52, 2015. [Online]. Available: https://doi.org/10.1016/j.apnum.2014.08.003.

[2] P. M. Burrage and K. Burrage, 'Structure-preserving Runge-Kutta methods for stochastic Hamiltonian equations with additive noise,' *Numer. Algorithms*, vol. 65, no. 3, pp. 519–532, 2014, ISSN: 1017-1398. DOI: 10.1007/s11075-013-9796-6. [Online]. Available: https://doi.org/10.1007/s11075-013-9796-6.

[3] E. Hairer, C. Lubich and G. Wanner, *Geometric numerical integration*, Second, ser. Springer Series in Computational Mathematics. Springer-Verlag, Berlin, 2006, vol. 31, pp. xviii+644, Structure-preserving algorithms for ordinary differential equations, ISBN: 978-3-540-30663-4.

[4] E. Hairer, S. P. Nørsett and G. Wanner, *Solving ordinary differential equations. I*, Second, ser. Springer Series in Computational Mathematics. Springer-Verlag, Berlin, 1993, vol. 8, pp. xvi+528, Nonstiff problems, ISBN: 3-540-56670-8.

[5] *Invariant*, Web Page, Dictionary Entry, Accesed 11.01.2022, Merriam-Webster. [Online]. Available: https://www.merriam-webster.com/dictionary/invariant.

[6] G. Zhong and J. E. Marsden, 'Lie-Poisson Hamilton-Jacobi theory and Lie-Poisson integrators,' *Phys. Lett. A*, vol. 133, no. 3, pp. 134–139, 1988, ISSN: 0375-9601. DOI: 10.1016/0375-9601(88)90773-6. [Online]. Available: https://doi.org/10.1016/0375-9601(88)90773-6.

[7] R. I. McLachlan, G. R. W. Quispel and N. Robidoux, 'Geometric integration using discrete gradients,' *R. Soc. Lond. Philos. Trans. Ser. A Math. Phys. Eng. Sci.*, vol. 357, no. 1754, pp. 1021–1045, 1999, ISSN: 1364-503X. DOI: 10.1098/rsta.1999.0363. [Online]. Available: https://doi.org/10.1098/rsta.1999.0363.

[8] G. R. W. Quispel and D. I. McLaren, 'A new class of energy-preserving numerical integration methods,' *J. Phys. A*, vol. 41, no. 4, pp. 045206, 7, 2008, ISSN: 1751-8113. DOI: 10.1088/1751-8113/41/4/045206. [Online]. Available: https://doi.org/10.1088/1751-8113/41/4/045206.

[9] E. Celledoni, B. Owren and Y. Sun, 'The minimal stage, energy preserving Runge-Kutta method for polynomial Hamiltonian systems is the averaged vector field method,' *Math. Comp.*, vol. 83, no. 288, pp. 1689–1700, 2014, ISSN: 0025-5718. DOI: 10.1090/S0025-5718-2014-02805-6. [Online]. Available: https://doi.org/10.1090/S0025-5718-2014-02805-6.

[10] E. Celledoni, V. Grimm, R. I. McLachlan, D. I. McLaren, D. O'Neale, B. Owren and G. R. W. Quispel, 'Preserving energy resp. dissipation in numerical PDEs using the "average vector field" method,' *J. Comput. Phys.*, vol. 231, no. 20, pp. 6770–6789, 2012, ISSN: 0021-9991. DOI: 10.1016/j.jcp.2012.06.022. [Online]. Available: https://doi.org/10.1016/j.jcp.2012.06.022.

[11] L. Brugnano, F. Iavernaro and D. Trigiante, 'Analysis of hamiltonian boundary value methods (hbvms): A class of energy-preserving runge–kutta methods for the numerical solution of polynomial hamiltonian systems,' *Communications in Nonlinear Science and Numerical Simulation*, vol. 20, no. 3, pp. 650–667, Mar. 2015, ISSN: 1007-5704. DOI: 10.1016/j.cnsns.2014.05.030. [Online]. Available: http://dx.doi.org/10.1016/j.cnsns.2014.05.030.

[12] L. Brugnano, F. Iavernaro and D. Trigiante, 'A note on the efficient implementation of hamiltonian bvms,' *Journal of Computational and Applied Mathematics*, vol. 236, no. 3, pp. 375–383, Sep. 2011, ISSN: 0377-0427. DOI: 10.1016/j.cam.2011.07.022. [Online]. Available: http://dx.doi.org/10.1016/j.cam.2011.07.022.

[13] L. Brugnano, G. Frasca Caccia and F. Iavernaro, 'Efficient implementation of gauss collocation and hamiltonian boundary value methods,' *Numerical Algorithms*, vol. 65, no. 3, pp. 633–650, Jan. 2014, ISSN: 1572-9265. DOI: 10.1007/s11075-014-9825-0. [Online]. Available: http://dx.doi.org/10.1007/s11075-014-9825-0.

[14] L. Brugnano, F. Iavernaro and D. Trigiante, 'Hamiltonian boundary value methods (energy preserving discrete line integral methods),' *JNAIAM. J. Numer. Anal. Ind. Appl. Math.*, vol. 5, no. 1-2, pp. 17–37, 2010, ISSN: 1790-8140.

[15] K. Burrage and P. M. Burrage, 'Low rank Runge-Kutta methods, symplecticity and stochastic Hamiltonian problems with additive noise,' *J. Comput. Appl. Math.*, vol. 236, no. 16, pp. 3920–3930, 2012, ISSN: 0377-0427. DOI: 10.1016/j.cam.2012.03.007. [Online]. Available: https://doi.org/10.1016/j.cam.2012.03.007.

[16] L. Brugnano and F. Iavernaro, 'Line integral methods which preserve all invariants of conservative problems,' *J. Comput. Appl. Math.*, vol. 236, no. 16, pp. 3905–3919, 2012, ISSN: 0377-0427. DOI: 10.1016/j.cam.2012.03.026. [Online]. Available: https://doi.org/10.1016/j.cam.2012.03.026.

[17] H.-H. Kuo, *Introduction to stochastic integration*, ser. Universitext. Springer, New York, 2006, pp. xiv+278, ISBN: 978-0387-28720-1; 0-387-28720-5.

[18] K. Itô, 'Stochastic integral,' *Proc. Imp. Acad. Tokyo*, vol. 20, pp. 519–524, 1944, ISSN: 0369-9846. [Online]. Available: http://projecteuclid.org/euclid.pja/1195572786.

[19] R. L. Stratonovich, 'A new representation for stochastic integrals and equations,' *SIAM J. Control*, vol. 4, pp. 362–371, 1966, ISSN: 0363-0129.

[20] G. N. Milstein and M. V. Tretyakov, *Stochastic numerics for mathematical physics*, Second, ser. Scientific Computation. Springer-Verlag, Berlin, 2021, pp. xxv+736, ISBN: 3-540-21110-1. DOI: 10.1007/978-3-030-82040-4. [Online]. Available: https://doi.org/10.1007/978-3-030-82040-4.

[21] P. E. Kloeden and E. Platen, *Numerical solution of stochastic differential equations*, ser. Applications of Mathematics (New York). Springer-Verlag, Berlin, 1992, vol. 23, pp. xxxvi+632, ISBN: 3-540-54062-8. DOI: 10.1007/978-3-662-12616-5. [Online]. Available: https://doi.org/10.1007/978-3-662-12616-5.

[22] K. Burrage and P. M. Burrage, 'High strong order explicit Runge-Kutta methods for stochastic ordinary differential equations,' in 1-3, vol. 22, Special issue celebrating the centenary of Runge-Kutta methods, 1996, pp. 81–101. DOI: 10.1016/S0168-9274(96)00027-X. [Online]. Available: https://doi.org/10.1016/S0168-9274(96)00027-X.

[23] D. J. Higham, 'An algorithmic introduction to numerical simulation of stochastic differential equations,' *SIAM Rev.*, vol. 43, no. 3, pp. 525–546, 2001, ISSN: 0036-1445. DOI: 10.1137/S0036144500378302. [Online]. Available: https://doi.org/10.1137/S0036144500378302.

[24] K. Burrage and P. M. Burrage, 'Order conditions of stochastic Runge-Kutta methods by B-series,' *SIAM J. Numer. Anal.*, vol. 38, no. 5, pp. 1626–1646, 2000, ISSN: 0036-1429. DOI: 10.1137/S0036142999363206. [Online]. Available: https://doi.org/10.1137/S0036142999363206.

[25] Y. Komori, 'Multi-colored rooted tree analysis of the weak order conditions of a stochastic Runge-Kutta family,' *Appl. Numer. Math.*, vol. 57, no. 2, pp. 147–165, 2007, ISSN: 0168-9274. DOI: 10.1016/j.apnum.2006.02.002. [Online]. Available: https://doi.org/10.1016/j.apnum.2006.02.002.

[26] A. Rößler, 'Rooted tree analysis for order conditions of stochastic Runge-Kutta methods for the weak approximation of stochastic differential equations,' *Stoch. Anal. Appl.*, vol. 24, no. 1, pp. 97–134, 2006, ISSN: 0736-2994. DOI: 10.1080/07362990500397699. [Online]. Available: https://doi.org/10.1080/07362990500397699.

[27] S. Anmarkrud and A. Kværnø, 'Order conditions for stochastic Runge-Kutta methods preserving quadratic invariants of Stratonovich SDEs,' *J. Comput. Appl. Math.*, vol. 316, pp. 40–46, 2017, ISSN: 0377-0427. DOI: 10.1016/j.cam.2016.08.042. [Online]. Available: https://doi.org/10.1016/j.cam.2016.08.042.

[28] K. Debrabant and A. Kværnø, 'Cheap arbitrary high order methods for single integrand SDEs,' *BIT*, vol. 57, no. 1, pp. 153–168, 2017, ISSN: 0006-3835. DOI: 10.1007/s10543-016-0619-8. [Online]. Available: https://doi.org/10.1007/s10543-016-0619-8.

[29] X. Li, C. Zhang, Q. Ma and X. Ding, 'Discrete gradient methods and linear projection methods for preserving a conserved quantity of stochastic differential equations,' *Int. J. Comput. Math.*, vol. 95, no. 12, pp. 2511–2524, 2018, ISSN: 0020-7160. DOI: 10.1080/00207160.2017.1408803. [Online]. Available: https://doi.org/10.1080/00207160.2017.1408803.

[30] C. Chen, D. Cohen, R. D'Ambrosio and A. Lang, 'Drift-preserving numerical integrators for stochastic Hamiltonian systems,' *Adv. Comput. Math.*, vol. 46, no. 2, Paper No. 27, 22, 2020, ISSN: 1019-7168. DOI: 10.1007/s10444-020-09771-5. [Online]. Available: https://doi.org/10.1007/s10444-020-09771-5.

[31] R. D'Ambrosio, G. Giordano, B. Paternoster and A. Ventola, 'Perturbative analysis of stochastic Hamiltonian problems under time discretizations,' *Appl. Math. Lett.*, vol. 120, Paper No. 107223, 7, 2021, ISSN: 0893-9659. DOI: 10.1016/j.aml.2021.107223. [Online]. Available: https://doi.org/10.1016/j.aml.2021.107223.

[32] W. Zhou, J. Zhang, J. Hong and S. Song, 'Stochastic symplectic Runge-Kutta methods for the strong approximation of Hamiltonian systems with additive noise,' *J. Comput. Appl. Math.*, vol. 325, pp. 134–148, 2017, ISSN: 0377-0427. DOI: 10.1016/j.cam.2017.04.050. [Online]. Available: https://doi.org/10.1016/j.cam.2017.04.050.

[33] R. D'Ambrosio, M. Moccaldi and B. Paternoster, 'Numerical preservation of long-term dynamics by stochastic two-step methods,' *Discrete Contin. Dyn. Syst. Ser. B*, vol. 23, no. 7, pp. 2763–2773, 2018, ISSN: 1531-3492. DOI: 10.3934/dcdsb.2018105. [Online]. Available: https://doi.org/10.3934/dcdsb.2018105.

[34] B. Øksendal, *Stochastic differential equations*, Sixth, ser. Universitext. Springer-Verlag, Berlin, 2003, pp. xxiv+360, An introduction with applications, ISBN: 3-540-04758-1. DOI: 10.1007/978-3-642-14394-6. [Online]. Available: https://doi.org/10.1007/978-3-642-14394-6.

[35] E. Hairer and G. Wanner, *Solving ordinary differential equations. II*, Second, ser. Springer Series in Computational Mathematics. Springer-Verlag, Berlin, 1996, vol. 14, pp. xvi+614, Stiff and differential-algebraic problems, ISBN: 3-540-60452-9. DOI: 10.1007/978-3-642-05221-7. [Online]. Available: https://doi.org/10.1007/978-3-642-05221-7.

[36] R. I. McLachlan and G. R. W. Quispel, 'Splitting methods,' *Acta Numer.*, vol. 11, pp. 341–434, 2002, ISSN: 0962-4929. DOI: 10.1017/S0962492902000053. [Online]. Available: https://doi.org/10.1017/S0962492902000053.

[37] L. Brugnano, G. Gurioli and F. Iavernaro, 'Analysis of energy and quadratic invariant preserving (EQUIP) methods,' *J. Comput. Appl. Math.*, vol. 335, pp. 51–73, 2018, ISSN: 0377-0427. DOI: 10.1016/j.cam.2017.11.043. [Online]. Available: https://doi.org/10.1016/j.cam.2017.11.043.

Appendix A

Newton iterations

To solve the nonlinear systems of equations arising from (4.20), Newton iterations are applied to find approximate solutions to the next step x_{n+1}. The full derivation is shown for the implicit midpoint method, while only the final expressions are given for the rest.

The expression (4.20a) can be reformulated by setting $y = x_{n+1} - x_n$ to

$$G(y) = y - \Delta t f\left(y_0 + \frac{1}{2}y\right) = 0, \qquad y_0 = x_n. \tag{A.1}$$

The Newton iteration takes the form

$$y_{k+1} = y_k - (\partial_y G(y_k))^{-1} \cdot G(y_k), \tag{A.2}$$

converging towards $x_{n+1} - x_n$. Using the chain rule to calculate an expression for $\partial_y G(y)$:

$$\begin{aligned} \partial_y G(y) &= I - \Delta t \nabla f\left(y_0 + \frac{1}{2}y\right) \cdot \frac{1}{2}, \qquad \nabla f\left(y_0 + \lambda y\right) \approx \nabla f(y_0) =: J_n, \quad 0 \le \lambda \le 1, \\ &\approx I - \frac{\Delta t}{2} J_n =: \Gamma_n, \end{aligned}$$

which is constant with respect to y. Setting $\Delta y_k = y_{k+1} - y_k$ and reformulating (A.2), if follows that

$$\Gamma_n \Delta y_k = -G(y_k), \tag{A.3a}$$

$$y_{k+1} = y_k + \Delta y_k. \tag{A.3b}$$

(A.3a) is system of linear equations for δy_k, which can be solved with a single LU-decomposition of Γ_n for each time-step n, calculating $G(y_k)$ according to (A.1). All the methods presented here can be rewritten into such a system with different expressions for Γ_n and $G(y)$.
For Kahan's method, the expressions become

$$\begin{aligned} G(y) &= y - \frac{\Delta t}{2}\left(-f(y_0) + 4f\left(y_0 + \frac{1}{2}y\right) - f\left(y_0 + y\right)\right), \\ \Gamma_n &= I - \frac{\Delta t}{2} \cdot J_n. \end{aligned}$$

For the MQ and Lobatto IIIA methods, the expressions become

$$\begin{aligned} G(y) &= y - \frac{\Delta t}{6}\left(f(y_0) + 4f\left(y_0 + \frac{1}{2}y + \frac{\theta \Delta t}{8}\left(f(y_0) - f(y_0 + y)\right)\right) + f(y_0 + y)\right) \\ \Gamma_n &= I - \frac{\Delta t}{2} J_n + \frac{\theta (\Delta t)^2}{12} J_n^2. \end{aligned}$$

Link to code repository: TBA