Thesis for the Degree of Doctor of Technology

Contributions to Numerical Solution of Stochastic Differential Equations

Anders Muszta

CHALMERS | GÖTEBORG UNIVERSITY

Department of Mathematical Statistics
Chalmers University of Technology
and Göteborg University
SE-412 96 Göteborg, Sweden

Göteborg, Sweden 2005

Contributions to Numerical Solution of Stochastic Differential Equations
Anders Muszta

© Anders Muszta, 2005.

ISSN 0347-2809
Preprint number 2005:2
Department of Mathematical Statistics
Chalmers University of Technology and Göteborg University
SE-412 96 Göteborg
Sweden
Telephone + 46 (0)31-772 1000

Chalmers University of Technology
Göteborg, Sweden 2005

Abstract

This thesis consists of four papers:

The first paper is an overview article of recent techniques in strong numerical solutions of stochastic differential equations, driven by Wiener processes, that have appeared the last then 10 years, or so.

The second paper studies theoretical and numerical aspects of stochastic differential equations with so called volatility induced stationarity. While being of great importance in contemporary applications, these equations are particularly difficult from a numerical point of view, to the extent that most or even all standard numerical procedures fail.

The third paper develops numerical procedures for stochastic differential equations driven by Lévy processes. A general scheme for stochastic Taylor expansions is developed, together with an analysis of convergence properties.

The fourth paper shows how to reduce the common global Lipschitz condition for numerical procedures for stochastic differential equations, to a local Lipschitz condition.

Keywords: Adaptive method, change of time, CIR model, CKLS model, convergence order, Euler method, heavy-tailed SDE, hyperbolic SDE, geometric integration, global Lipschitz condition, Lévy process, Lie group method, local Lipschitz condition, local martingales, mean reversion, Milstein method, numerical method, semimartingale, stochastic differential equation, stochastic Taylor expansion, strong approximation, symplectic integration, volatility induced stationarity, waveform relaxation.

Acknowledgements

My many thanks to Patrik Albin, who supervised this thesis, for guiding, encouraging, and teaching me and for working tirelessly on helping me perfect the manuscript. Without your advice and patience over the years this thesis could not have been written. Thank you for making this the most eventful five years of my life so far.

My thanks also go to my second supervisor, Professor Roger Petersson. Thank you for your kind and unselfish help with various matters regarding this thesis.

Thank you all my colleagues and friends at Matematiska Vetenskaper for creating a friendly work environment.

Lastly I would like to thank my family for their unfailing love and support.

Anders Muszta, 6 March 2005

Introduction and Summary

Here we make an introductory and summarizing review the four papers that the thesis consists of:

Paper I: Recent advances in the numerical solution of stochastic differential equations. Joint work with J.M.P. Albin. Submitted to *SIAM Journal on Numerical Analysis*.

Paper II: On volatitilty induced stationarity for stochastic differential equations. Joint work with Martin Richter, J.M.P. Albin and Bjarne Astrup Jensen. Submitted to *Advances in Applied Probability*.

Paper III: On numerical solution of stochastic differential equations driven by Lévy processes. Submitted to *Stochastic Processes and Their Applications*.

Paper IV: A note on the Lipschitz condition for Euler approximation of stochastic differential equations. Joint work with Martin Richter, J.M.P. Albin and Bjarne Astrup Jensen. Submitted to *SIAM Journal on Numerical Analysis*.

Contents

1. Review of Paper I: Recent advances in the numerical solution of stochastic differential equations 5

1.1 Review of the paper 5

1.2 On publication and the relation to other work . . . 6

1.3 On the contribution of Anders Muszta 7

2. Review of Paper II: On volatitilty induced stationarity for stochastic differential equations 9

2.1 Review of the paper 9

2.2 On publication and the relation to other work . . . 13

2.3 On the contribution of Anders Muszta 13

3. Review of Paper III: On numerical solution of stochastic differential equations driven by Lévy processes . . . 15

3.1 Review of the paper 15

3.2 On publication and the relation to other work . . . 19

3.3 On the contribution of Anders Muszta 19

4. Review of Paper IV: A note on the Lipschitz condition for Euler approximation of stochastic differential equations 21

4.1 Review of the paper 21

4.2 On publication and the relation to other work . . . 22

4.3 On the contribution of Anders Muszta 22

1. Review of Paper I: Recent advances in the numerical solution of stochastic differential equations

This paper gives an overview of recent techniques in the strong numerical solutions of SDE, that have appeared after the publication of the standard work on numerical methods for SDE by Kloeden and Platen (1992). Many of the techniques that are overviewed are very recent.

As, argubly, the new techniques numerical solutions of SDE have somewhat changed the area, we belive that it is important to write an overview, to communicate these techniques to a wide audience, in an accessible manner. For completeness, and to give a background for the understanding of the new techniques, some of the most important ideas in Kloeden and Platen (1992) are also shortly reviewed.

On the personal level, we were advised to write this overview, to learn the state-of-the-art of the area. In this manner we could combine learning about the area with writing a research paper.

1.1 Review of the paper

As the paper itself, being an overview, is a review of other works, we do not review it here, as we would have ended up with virtually the same review text that makes up the paper. Instead we merely indicate what are the different components of the paper.

The paper starts with a preparatory section, that introduces the concepts of stochastic integral and SDE, together with the phenomenally important Itô formula, in a compart manner, in part to set the notation of the latter sections.

The second section covers the mentioned background knowledge from the book by Kloeden and Platen, which includes the concepts of *strong approximation* and *stochastic Taylor expansion*, together with the *Euler method* and the *Milstein method*.

In the third section we report on some important recent progress to relax the usual conditions about global Lipschitz coefficients with linear growth, for strong numerical methods. Some authors call SDE which do not satisfy these latter conditions *non-linear SDE*.

In Section 4 we briefly mention the idea behind the *stochastic theta method*, to obtain long-time stability of strong solutions of SDE.

The fifth section give the ideas behind so called *stochastic delay differential equations* (SDDE), and their numerical solution. This fairly recent research

area deals with non-diffusion non-Markov type SDE, where futute values of the solution depends on the history of the equation in a more complicated manner than just through the present value of the solution.

In Section 6 we review the family of *waveform relaxation methods*, for dealing with high-dimensional SDE with weakly interacting subsystems, that possibly have clocks that run with quite different speeds, in which case they are called stiff. As with many numerical techniques for SDE, waveform relaxation methods origin in methods for ordinary differential equations, with a similar flavour.

The seventh section desribe the ideas behind *adaptive numerical methods* for SDE. The idea with adaptive methods is to select a step-length for the numerical solution that is related to the sizes of drift and diffusions coefficients of the SDE: For small values of the drift and diffusions coefficients, a large step-length is selected, as there is a rather small contribution to the error for such coefficients, while for large values of the drift and diffusion coefficients, a small step-length is selected. The idea is that this should give the most efficient use of the computer time.

In Section 8 we discuss *symplectic numerical methods* for solution of Hamiltonian SDE. The idea behind such methods is to replace the original Hamiltonian SDE with another approximating Hamiltonian SDE, which is then solved exactly, more or less, in part in order to gain better long-time stability of the numerical approximation than with more traditional methods, by preservation of the Hamiltonian structure.

The final ninth section describe *Lie group methods*, for numerical methods of solution of SDE that evolve on Lie groups. Such methods give numerical schemes that stay on the Lie group, which is not the case with traditional numerical approximations.

1.2 On publication and the relation to other work

The paper is a substantially edited verison of the licentiate thesis of Anders Muszta (2003). During the process of editing the paper, we found out about the overview of recent numerical methods for weak solutions of SDE, that feature as a chapter in Schuëller (1997). This overview is not easy to find, as it does not feature in article indices, as for example, MathSciNet: We found out about the existence of this overview at the very last stages of the preparation of our paper.

As the sections on weak methods in Muszta (2003), as well as in our originally planned paper were quite short, compared with the chapters on strong methods, and as they had a considerable overlap with the overview featuring

in Schuëller (1997), we decided to omit weak methods from our overview, at a very late stage. However, on the other hand, compared with the licentiate thesis of Muszta, we have added the sections on non-linear SDE, on the stochastic theta method, and on the numerical solution of stochastic delay differential equations.

Platen (1999) gives an overview of numerical methods for SDE. However, this is more or less a condensed version of Kloeden and Platen (1992), and the overlap with our paper is non-sigificant.

Burrage, Burrage and Tian (2004) gives an overview of strong numerical methods for SDE that overlap slightly more with our oversight, than does the overview by Platen. We became aware of the existence of the former overview at the very last stages of the preparation of our paper, as it did not start to feature in article indices, as for example, MathSciNet, before then. However, as in the overview by Platen, the focus in the overview by Burrage, Burrage and Tian is on strong numerical methods for SDE in general, rather than on recent methods as in our paper, so that, in particular, it does again to a large extent build on the book by Kloeden and Platen.

The paper was submitted to SIAM Journal on Numerical Analysis, on March 2, 2005.

1.3 On the contribution of Anders Muszta

The paper is a joint work of Anders Muszta and J.M.P. Albin. However, the role of J.M.P. Albin has been restricted to a quite considerable typographical work with the manuscript, and to an advisory role, while it is Anders Muszta that has contributed with all of the actual content of the paper.

2. Review of Paper II: On volatitilty induced stationarity for stochastic differential equations

This paper studies theoretical and numerical aspects of SDE, with so called volatility induced stationarity (VIS). While being of great importance in contemporary applications, these equations are particularly difficult from a numerical point of view, to the extent that most or even all standard numerical procedures fail.

On the personal level, we were advised to write this paper, to get in close down-to-earth contact with some specific SDE, of great interest in contemporary applications, that are very difficult to handle from a numerical point of view. This in-depth study of a few examples complements the other three papers, that are set in more general contexts. In addition, we wanted to cooperate with people from economy: On of the coauthors, Martin Richter, recently took up work at a Danish bank, but was previously affiliated with the Department of Finance, at Copenhagen Businees School. Another of the coauthors, Bjarne Astrup Jensen, is affiliated with the Department of Finance, at Copenhagen Businees School, as well as with the Center of Finance, at the Gothenburg School of Economics.

2.1 Review of the paper

This paper deals with SDE with VIS. Although the VIS effect is more or less well-known, misinterpretations in the literature are common. We clerify matters by giving a general definition of VIS, and explain why SDE with VIS often have a local martingale term that is not a martingale. We investigate the relation between VIS and stationary moments and mean reversion.

We study a class of stationary and ergodic SDE, for which stationarity is ensured by a high volatility - VIS. SDE with VIS escape from the stationary level often, but in short periods if the VIS is strong. Between these excursions, the process is very steady. As SDE with stationarity from mean reversion by the drift do not behave in this manner, SDE with VIS are important additions to more conventional SDE models.

The VIS appears at the boundaries of the SDE. Either we have a local martingale part of the solutions that is not a martingale, with a mean reverting drift, or we have a martingale or non-martingale local martingale part that is mean reverting, simply because the clock runs quicker near VIS boundaries. We conclude that interpretations and statistical methods for SDE with VIS in the literature have to be revised, as SDE with VIS do not behave as white noise perturbed ODE. In particular, we point out mistakes in the literature, on inference for SDE with VIS.

SDE with VIS feature to model, for example, interest rates and eletricity prices. But such SDE can be difficult from a statistical point of view, so that basic estimation procedures fail. Hence computer simulations to evaluate statistics are important. As virtually all traditional simulation schemes may fail for SDE with VIS, we consider alternative methods.

We provide three examples of SDE with VIS, that all are different in terms of their VIS effect, and that require different simulation procedures. The main example is the CKLS model, which has been studied by many authors, but where we find some new features. The others are a class of hyperbolic SDE, and a class of heavy-tailed SDE.

Followed by an introduction is Section 2, were the notation is set: Let $\{W_t\}_{t\geq 0}$ be a *standard Brownian motion*, and $I = (l, r)$, $-\infty \leq l < r \leq \infty$ an open interval. For a continuous *drift* $b : I \to \mathbb{R}$ and a locally Hölder continuous of order $\frac{1}{2}$ *volatility* $\sigma :\to (0, \infty)$, consider the SDE

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dW_t, \quad X_0 = \zeta,$$

where ζ is a random variable with values in I, that is independent of W.

Given an $x_0 \in I$, the *scale* function S and *speed* measure m are given by

$$S(x) = \int_{x_0}^{x} \exp\left\{-2\int_{x_0}^{y} \frac{b(z)}{\sigma(z)^2}\,dz\right\} dy \quad \text{and} \quad \frac{dm(x)}{dx} = \frac{2}{\sigma(x)^2 S'(x)}.$$

We will require that $S(l^+) = -\infty$, $S(r^-) = \infty$, and $m(I) < \infty$.

With the above assumptions, the SDE will have a unique strong solution that does not explode, and that has the *stationary distribution* $\pi = m/m(I)$. Further, the SDE is *ergodic*, *β-mixing*, and *ρ-mixing*.

In Section 3 we define the concept of VIS:

Definition. *The stationary solution to the SDE has* VIS *at l [r], and we call l [r] a* VIS boundary, *if $S'(l^+) < \infty$ [$S'(r^-) < \infty$]. If $S'(l^+) > 0$ [$S'(r^-) > 0$] we call l [r] a* positive VIS boundary *– otherwise it is a* null VIS boundary.

Because of our assumptions in Section 2, we can only have VIS boundaries at $\pm\infty$.

In Section 4 we refer some results from the literature, on simulation of SDE, for easy reference for later use.

In addition, we prove that the Euler scheme often is transient with positive probability, for SDE with VIS. In particular, we show that this applies when

$b = 0$, so that X is a local martingale, provided that

$$\int_{-\infty}^{\infty} |x|^p \, d\pi(x) < \infty \quad \text{for some } p > 1.$$

By a result of Gyöngy (1998), the Euler scheme converges almost surely as the step-length $\Delta \downarrow 0$. But for many SDE with VIS, for any specific Δ, the Euler scheme may diverge. In fact, for many specific SDE with VIS, the Euler scheme breaks down almost immediately, even for impossible small step-lengths.

In Sections 5 we consider the *CKLS model* of Chan, Koralyi, Longstaff and Sanders (1992)

$$dX_t = (\alpha + \beta X_t)\, dt + \sigma X_t^\gamma \, dW_t, \quad X_0 = \zeta > 0,$$

with parameters that gives the SDE a VIS boundary at ∞, namely $\{\frac{1}{2} < \gamma < 1, \alpha > 0, \beta = 0\} \cup \{\gamma = 1, \alpha > 0, 0 \le \beta < \frac{1}{2}\sigma^2\} \cup \{\gamma > 1, \alpha > 0\} \cup \{\gamma > 1, \alpha = 0, \beta > 0\}$.

The VIS boundary ∞ is null for $\gamma = 1$ and positive otherwise. The local martingale part of the stationary solution is not a martingale. The SDE has a medium degeree of VIS.

The stationary mean is infinite for $\gamma \le 1$: Moments of order $p < 2\gamma - 1$ exist for $\{\frac{1}{2} < \gamma < 1, \alpha > 0, \beta = 0\}$, and of order $p < 1 - 2\beta/\sigma^2$ for $\{\gamma = 1, \alpha > 0, 0 \le \beta < \frac{1}{2}\sigma^2\}$. When $\gamma > 1$, stationary moments of order $p < 2\gamma - 1$ exist, and we find a formula for these moments, as well as for the stationary distribution, which can be used to check the fit of the CKLS model to data.

We show that the stationary local martingale has a drift

$$\mathbf{E}^\pi \left\{ \sigma \int_0^t X_s^\gamma dW_s \right\} = -\left(\alpha + \beta \int_0^\infty x \, d\pi(x) \right) t,$$

that can contribute to the mean reversion, or conteract it, depending on the values of α and β. In any case, the clock is very quick at high levels, so that the process necessarily spends little time there.

As the stationary local martingale has a drift, it follows that the so called econometric approach to statistical inference for the CKLS model, by the generalized method of moments, suggested by Chan, Koralyi, Longstaff and Sanders (1992), is incorrect.

For $\gamma \le 1$, Paper IV gives uniform convergence in mean-square of the Euler scheme. For $\gamma > 1$, the Euler scheme breaks down with positive probability.

For $\gamma > 1$, we simulate the SDE by considering the transformed process $Z_t = f(X_t)$, where $f(x) = x^{1-\gamma}$, satisfying the SDE

$$dZ_t = \varphi(Z_t)\, dt + (1 - \gamma)\sigma \, dW_t,$$

where

$$\varphi(z) = \alpha(1-\gamma)z^{\gamma/(\gamma-1)} + (1-\gamma)\beta z + \frac{1}{2}\gamma(\gamma-1)\sigma^2 z^{-1}.$$

This SDE in turn is simulated by various implicit Euler methods.

In Section 6 we consider a *hyperbolic diffusion*, given by the SDE

$$dX_t = \sigma \exp\Big\{\frac{1}{2}\Big(\alpha\sqrt{\delta^2 + (X_t-\mu)^2} - \beta(X_t-\mu)\Big)\Big\}\, dW_t, \quad X_0 = \zeta,$$

with parameters $\alpha > |\beta| \geq 0$, $\delta, \sigma > 0$ and $\mu \in \mathbb{R}$. This SDE has positive VIS boundaries at $\pm\infty$. The stationary solution has all moments.

The stationary local martingale is uniformly integrable, but is not a martingale. The mean reversion is not due to bias of the local martingale, but to the fact that the clock is very quick at large values. The SDE has a high degree of VIS.

The Euler scheme is transient with positive probability. This should then apply also to higher order explicit schemes.

We simulate this SDE in the following way: Let $[X]_t = \int_0^t \sigma^2(X_s)\, ds$ be the quadratic variation process. If T_t solves $[X]_{T_t} = t$, then $B_t = X_{T_t} - X_0$ is a Brownian motion. The idea then is to simulate B, and get X from

$$T_t = \int_0^t \frac{du}{\sigma^2(X_0 + B_u)} \quad \text{and} \quad X_{T_t} = X_0 + B_t.$$

We consider a grid $0 = t_0 < t_1 < \ldots < t_N = T$, with spacing Δ. Given T_{t_n}, we have

$$X_{T_{t_{n+1}}} = X_0 + B_{t_{n+1}} \quad \text{where} \quad T_{t_{n+1}} - T_{t_n} = \int_{t_n}^{t_{n+1}} \frac{du}{\sigma^2(X_0 + B_u)}.$$

In Section 7 we consider the family of SDE

$$dX_t = 3X_t^a\, dt + 3X_t^{2/3}\, dW_t, \quad X_0 = \zeta > 0,$$

where $a < \frac{1}{3}$ is a constant. This SDE has a positive VIS boundary at $r = \infty$. Stationary moments only exist up to order $p < \frac{1}{3}$ exist, making the SDE *heavy tailed*.

The SDE has a quite low degeree of VIS, allowing it to make quite long excursions from the stationary level. The local martingale part of the stationary solution is not a martingale.

By Paper IV, the Euler scheme is uniformly convergent in mean-square for $a \geq 0$.

For $a < 0$ explicit schemes are unstable, as they jump, from zero to high values, from which there is little reversion. This in turn, is not a property displayed by the true solution.

To simulate the SDE for $a < 0$, we first simulate the transformed process $Z_t = f(X_t)$, where $f(x) = x^{1/3}$, satisfying the SDE

$$dZ_t = \varphi(Z_t)\,dt + dW_t, \quad Z_0 = f(\zeta) > 0,$$

where $\varphi(z) = z^{3a-2} - z^{-1}$. This SDE in turn, is simulated by an implicit Euler method.

In Section 8 we summarize our conclusions in the paper.

2.2 On publication and the relation to other work

The point of take off were the unpublished work by Richter (2004). This manuscript was obtained by Anders Muszta and J.M.P. Albin, from Martin Richter and Bjarne Astrup Jensen, during late summer 2004, in order for Anders Muszta and J.M.P. Albin to make mathematical statistics, publishable in an international journal out of it.

The paper was submitted to Advances in Applied Probability, on March 1, 2005. A correction was made in the manuscript, on March 4, 2005 (the same error that is mentioned in the review of Paper IV below), which we have selected not to communicate to the journal, at this stage, as it does not play a major role in the paper. However, the correction has been implemented in the thesis, so that the version of the paper that appear in the thesis is not completely the same as the paper submitted.

2.3 On the contribution of Anders Muszta

The paper is joint work of Anders Muszta, Martin Richter, J.M.P. Albin and Bjarne Astrup Jensen. The roles of J.M.P. Albin and Bjarne Astrup Jensen have been restricted to typographically "cleaning up" the manuscript, and to an advisory role, while it is Anders Muszta and Martin Richter that have contributed with the actual content of the paper, on equal terms. On a more detailed level, it is Martin that have contributed with the "raw material" to the paper, from an economical background, while it is Anders that has made a paper of it, that can be submitted to a scientific journal, including remaking of virtually all proofs, by agreement with Martin Richter, J.M.P. Albin and Bjarne Astrup Jensen, to make it a joint publication.

3. Review of Paper III: On numerical solution of stochastic differential equations driven by Lévy processes

This paper develops numerical procedures for SDE driven by Lévy processes, on the basis on a Stochastic Taylor Formula, that has proven succesful for SDE driven by Wiener processes. A general scheme for stochastic Taylor expansions is developed, together with an analysis of convergence properties.

The literature on strong numerical approximation of SDE driven by Lévy processes is very undeveloped. We belive that a Stochastic Taylor Formula is a basic step towards a more developed literature.

On the personal level, we were advised to write this paper, to become familiar with the very active and important, but also very difficult research area, of SDE driven by Lévy process: In many contemporary applications, for example, in mathematical finance, SDE driven by other processes than Wiener processes are becoming increasingly important, and we wanted to learn about this area.

3.1 Review of the paper

In 1992, the by now standard work: Numerical Solution of Stochastic Differential Equations, by Peter E. Kloeden and Eckhard Platen was published.

The basis for the work of Kloeden and Platen is the Stochastic Taylor Formula, to represent a solution of an SDE driven by Wiener processes, as a sum of multiple stochastic integrals. Approximate solutions are constructed by retaining a subset of the full representation. Kloeden and Platen then proceed to find convergence properties of the approximations.

We derive a *Stochastic Taylor Formula* for SDE driven by Lévy processes, and get approximation schemes by retaining a subset of the full representation. Further, we find mean-square convergence properties of the approximations.

Followed by the introduction is Section 2, which sets the notation and prerequisites of the paper. We explain the Einstein summation convention, which is used througout the paper: If in an expression a letter appears as a subscript and as a superscript, then summation is carried out over that letter. For example, $\sum_{i=1}^{n}\sum_{j=2}^{m} a_{ij}b^i c^j$ is written $a_{ij}b^i c^j$, $i \in \{1,\dots,n\}$, $j \in \{2,\dots,m\}$: As the range of the summation will be obvious from the context, in our paper, we only have to write $a_{ij}b^i c^j$ for such a double sum.

We consider an SDE

$$dX_t^j = a^j(t, X_{t-})\,dt + b_p^j(t, X_{t-})dL_t^p, \quad j = 1, \ldots, m, \tag{1}$$

where $L = (L^1, \ldots, L^m)$ is an m-dimensional Lévy process, built of independent one dimensional square integrable Lévy component processes $\{L_t^p\}_{t\in[0,\infty)}$, the law of which are determined by triples $(\gamma^p, \sigma(p), \nu^p)$, where $\gamma^p \in \mathbb{R}$ and $\sigma(p) \geq 0$ are constants and ν^p a Lévy measure.

We show that the Itô Formula for the solution X of (1) takes the form

$$f(t, X_t) = f(s, X_s) + \int_s^t \mathcal{L}_k f(r, X_{r-})\,d\mathcal{L}_r^k,$$

where the differential $d\mathcal{L}_t^k$ is defined by

$$d\mathcal{L}_t^k = \begin{cases} dt & \text{for } k = 0 \\ L_t^k & \text{for } k \in \{1, \ldots, m\} \\ dJ_t^{k-m} & \text{for } k \in \{m+1, \ldots, 2m\} \end{cases},$$

and where the operators $\mathcal{L}_{m+p}$ are given by

$$\begin{aligned} \mathcal{L}_0 f(r, X_{r-}) &= \partial_r f(r, X_{r-}) + a^j(r, X_{r-})\partial_j f(r, X_{r-}) \\ &\qquad + \frac{\sigma^2(p)}{2} b_p^i(u, X_{u-}) b_p^j(u, X_{u-})\partial_i\partial_j f(r, X_{r-}), \\ \mathcal{L}_p f(r, X_{r-}) &= b_p^j(r, X_{r-})\partial_j f(r, X_{r-}), \\ \mathcal{L}_{m+p} f(r, X_{r-}) &= b_p^j(r, X_{r-})\big(\partial_j f(r, \theta_r(X)) - \partial_j f(r, X_{r-})\big). \end{aligned}$$

Here $\theta(X) = \{\theta_t(X)\}_{t\geq 0}$ is a stochastic process such that, if $\Delta X_t^j \equiv X_t^j - X_{t-}^j = 0$ then $\theta_t(X)^j = X_t^j$, while if $\Delta X_t^j \neq 0$ then $\theta(X)_t^j \in (X_{t-}^j \wedge X_t^j, X_{t-}^j \vee X_t^j)$ and

$$f(t, X_t) - f(t, X_{t-}) = \partial_j f(t, \theta_t(X))\Delta X_t^j.$$

Further, J is the Lévy process of the jumps of L

$$J_t = \sum_{0<s\leq t} \Delta L_s \quad \text{for } t \geq 0.$$

Consider the set $\mathcal{M}$ of all *multi-indices* $(j_1, \ldots, j_k)$, where $j_1, \ldots, j_k \in \{0, 1, \ldots, m, m+1, \ldots, 2m\}$ and $k \in \mathbb{N}$, and where v denotes the empty multi-index.

A finite and non-empty set of multi-indices $\mathcal{A} \subseteq \mathcal{M}$ is *hierarchical* if it has the property that $(j_2, \ldots, j_k) \in \mathcal{A}$ if $(j_1, \ldots, j_k) \in \mathcal{A} \setminus \{v\}$.

The *remainder* of a hierarchical set $\mathcal{A}$ is defined

$$\mathcal{B}(\mathcal{A}) = \{(j_1, \ldots, j_k) \in \mathcal{M} \setminus \mathcal{A} : (j_2, \ldots, j_k) \in \mathcal{A}\}.$$

Define the *coefficient function* f_α for a sufficiently smooth function f and $(j_1, \ldots, j_k) \in \mathcal{M}$ recursively, by $f_{(j_1,\ldots,j_k)} = \mathcal{L}_{j_1} f_{(j_2,\ldots,j_k)}$ with $f_v = f$.

Define *multiple Itô integrals* for adapted cáglád processes f and $(j_1, \ldots, j_k) \in \mathcal{M}$ recursively, by

$$I_{(j_1,\ldots,j_k)}[f]_{s,t} = \int_s^t I_{(j_1,\ldots,j_{k-1})}[f]_{s,r}\, d\mathcal{L}_r^{j_k} \quad \text{with} \quad I_v[f]_{s,t} = f(t).$$

In Section 3, we establish the Stochastic Taylor Formula:

Stochastic Taylor Formula. *If X satisfies the SDE* (1)*, then we have for hierarchical $\mathcal{A} \subset \mathcal{M}$,*

$$f(t, X_t) = \sum_{\alpha \in \mathcal{A}} I_\alpha[f_\alpha(s, X_s)]_{s,t} + \sum_{\alpha \in \mathcal{B}(\mathcal{A})} I_\alpha[f_\alpha(\cdot, X_{(\cdot,-)})]_{s,t}.$$

The Stochastic Taylor Formula incorporates, as special cases, the Itô Formula, as well as the Taylor Formula from real analysis.

The Stochastic Taylor Formula carries over to general semimartingales.

In Section 4 we prove two inequalities that are crucial for the latter sections of the paper.

In Section 5 we prove strong convergence of the Stochastic Taylor Formula:

Strong Convergence. *Consider the hierarchical sets*

$$\mathcal{A}_k = \{\alpha \in \mathcal{M} : l(\alpha) + n(\alpha) \leq k\} \cup \{\alpha \in \mathcal{M} : l(\alpha) = n(\alpha) = (k+1)/2\}$$

for $k \in \mathbb{N}$. Consider the SDE (1) *with $d = 1$, take $f(t, x) = x$, and consider the* stochastic Taylor scheme

$$X_r^k = \sum_{\alpha \in \mathcal{A}_k} I_\alpha[f_\alpha(s, X_s)]_{s,r}.$$

Under technical conditions on the functions f_α, we have

$$\mathbb{E}\left\{ \sup_{r \in [s,t]} |X_r - X_r^k|^2 \right\} \leq C_1 (t-s)^k C_2^k \quad \textit{for } t - s \leq 1,$$

for some constants $C_1, C_2 > 0$. In particular,

$$\mathbb{P}\left\{ \lim_{k \to \infty} \sup_{r \in [s,t]} \left| X_r^k - X_r \right| = 0 \right\} = 1.$$

The approximations of the solution of an SDE take the form

$$X_t^k = \sum_{\alpha \in \mathcal{A}_k} I_\alpha[f_\alpha(s, X_s)]_{s,t}.$$

Since $I_\alpha[g]$ is linear in g, we get

$$I_\alpha[f_\alpha(s, X_s)]_{s,t} = f_\alpha(s, X_s) \int_s^t I_{\alpha-}[1]_{s,s_{l(\alpha)}} \, d\mathcal{L}_{s_{l(\alpha)}}^{j_{l(\alpha)}} = \ldots = f_\alpha(s, X_s) I_\alpha^L(s, t),$$

where

$$I_\alpha^L(s, t) = I_\alpha[1]_{s,t} = \int_{s_1=s}^t \ldots \int_{s_{l(\alpha)}=s}^{s_{l(\alpha)-1}} d\mathcal{L}_{s_1}^{j_1} \ldots d\mathcal{L}_{s_{l(\alpha)}}^{j_{l(\alpha)}}.$$

And so we may write the approximation

$$X_t^k = X_s + \sum_{\alpha \in \mathcal{A}_k \setminus \{v\}} f_\alpha(s, X_s) I_\alpha^L(s, t). \tag{2}$$

In Section 6 we continue with the SDE (1) with $d = 1$, we take $f(t, x) = x$, and consider a uniform partition $s = \tau_0 < \cdots < \tau_{n_t} \leq t$ of $[s, t]$, with $\tau_{j+1} - \tau_j = h \leq 1$ and $\max(t - \tau_{n_t}, \sup_j\{\tau_j - \tau_{j-1}\}) \leq h$, where $n_t = \sup\{k \in \mathbb{N} : \tau_k \leqslant t\}$.

The *stochastic Taylor approximation* $Y^k = \{Y_t^k\}_{t \geq 0}$ interpolates the values $\{X_{\tau_j}^k\}_{j=0}^{n_t}$ of the discrete approximations (2), and are given by

$$X_{\tau_{j+1}}^k = X_{\tau_j}^k + \sum_{\alpha \in \mathcal{A}_k \setminus \{v\}} f_\alpha(\tau_j, X_{\tau_j}^k) I_\alpha[1]_{\tau_j, \tau_{j+1}},$$

$$Y_r^k = X_{\tau_{n_r}}^k + \sum_{\alpha \in \mathcal{A}_k \setminus \{v\}} f_\alpha(\tau_{n_r}, X_{\tau_{n_r}}^k) I_\alpha[1]_{\tau_{n_r}, r}. \tag{3}$$

We have the following central result:

Order of Convergence. *With the notation of this section, there exists a constant $C > 0$ such that*

$$\mathbb{E}\Big\{ \sup_{r \in [s,t]} |X_r - Y_r^k|^2 \Big\} \leq 3\mathbb{E}\big\{ |X_s - Y_s^k|^2 \big\} + Ch^k.$$

In the final Section 7, we consider the analogues of the Euler scheme and the Milstein scheme, for SDE driven by Lévy processes, obtained from selecting $k = 1$ and $k = 2$, respectively, in (3).

3.2 On publication and the relation to other work

The point of take off were the Stochastic Taylor Formula, presented by Kloeden and Platen (1992), to represent solutions of SDE driven by Wiener processes, as sums of multiple stochastic integrals. Approximate solutions are then constructed by retaining a subset of the full representation. We simply carry this idea over to the context of SDE driven by semimartingales, and in particular Lévy processes.

This paper will be submitted to Stochastic Processes and Their Applications, in the form it is presented in the thesis.

3.3 On the contribution of Anders Muszta

The paper is work of Anders Muszta himself. However, J.M.P. Albin has given advice, including some strong advice on the actual presentation of the article, so that it was cut to less than a third of its original length.

4. Review of Paper IV: A note on the Lipschitz condition for Euler approximation of stochastic differential equations

This paper shows how the usual *global Lipschitz condition* on the coefficients, for convergence of the Euler scheme for numerical approximation of SDE, can be relaxed to *local Lipschitz conditions*, at finite boundaries.

On the personal level, we were advised to write this paper, to exercise ourselves in some technical skills that should feature in an ideally complete research education in numerical theory of SDE. In addition, we wanted cooperate with people from economy, see Sections 4.2 and 4.3 below.

4.1 Review of the paper

For an open interval $I = (l, r)$, and measurable functions $a : I \to \mathbb{R}$ and $b : I \to [0, \infty)$, consider a one dimensional time homogeneous SDE

$$dX_t = a(X_t)dt + b(X_t)dW_t \quad \text{for } t \in [0, T], \quad X_0 = x, \tag{4}$$

driven by a standard Brownian motion $\{W_t\}_{t\geq 0}$, where $x \in I$ is a constant.

Assume that X does not explode, and thus stays within the interval I for finite times, necessary and sufficients conditions for which are given by Feller's test for explosion (see e.g., Karatzas and Shreve, 1991).

Consider a uniform partition $0 = t_0 < \ldots < t_n = T$ of $[0, T]$ with $t_k - t_{k-1} = \Delta > 0$ for $k = 1, \ldots, n$. Writing $\Delta W_k = W(t_k) - W(t_{k-1})$, the Euler scheme for numerical approximation of the SDE (4) is given by

$$Y_k = Y_{k-1} + a(Y_{k-1})\Delta + b(Y_{k-1})\Delta W_k \quad \text{for } k = 1, \ldots, n. \tag{5}$$

The common condition in the literature for uniform mean-square convergence

$$\lim_{\Delta \downarrow 0} \mathbf{E}\Big\{ \sup_{0\leq k\leq n} |X_{t_k} - Y_k|^2 \Big\} = 0, \tag{6}$$

is that the coefficients a and b satisfy global Lipschitz conditions on I (which implies *linear growth* of the coefficients).

In the following theorem, that is the main result of the paper, we show that local Lipschitz conditions and linear growth are sufficient for the convergence (6) for the SDE (4), when I is finite. If I is half-finite, then the global Lipschitz condition is required at the infinite boundary, while local Lipschitz is enough at the finite boundary.

Theorem. *Assume that the coefficients a and b of the SDE* (4) *are globally Lipshitz on any closed sub-interval of I if I is finite, and globally Lipshitz on any closed half-finite sub-interval of I if I is half-finite. Further, assume that a and b have linear growth on I, and that the SDE does not explode. The Euler scheme* (5) *is uniformly mean-square convergent* (6).

The proof of the theorem makes use of Higham, Mao and Stuart (2002), Theorem 2.2, who assume that

$$\sup_{\Delta>0} \mathbf{E}\Big\{ \sup_{0\leq k\leq n} |Y_k|^p \Big\} < \infty \quad \text{and} \quad \mathbf{E}\Big\{ \sup_{0\leq t\leq T} |X_t|^p \Big\} < \infty \tag{7}$$

for some $p > 2$, together with local Lipschitz conditions on a and b, in order to establish (6). And so the proof consists of verifying (7), which lead to quite delicate martingale arguments, to establish the first part of that assumption.

As an example for which the theorem is useful, we consider the SDE

$$dX_t = \alpha X_t^a dt + \beta X_t^b dW_t \quad \text{for } t \in [0,T], \quad X_0 = x,$$

on the interval $I = (0,\infty)$, where $\alpha, \beta > 0$ and $a, b \in (0,1)$ are constants that satisfy $2b > a + 1$, and $x \in I$.

4.2 On publication and the relation to other work

The point of take off were the unpublished work by Richter (2004). This manuscript was obtained by Anders Muszta and J.M.P. Albin, from Martin Richter and Bjarne Astrup Jensen, during late summer 2004, in order for Anders Muszta and J.M.P. Albin to make mathematical statistics, publishable in an international journal out of it.

The paper was submitted to SIAM Journal on Numerical Analysis, on March 2, 2005, and resubmitted with a correction of a serious error, on March 4, 2005. This correction has also been made in the thesis.

4.3 On the contribution of Anders Muszta

The paper is joint work of Anders Muszta, Martin Richter, J.M.P. Albin and Bjarne Astrup Jensen. The roles of J.M.P. Albin and Bjarne Astrup Jensen have been advisory, including some typographical work on the presentation, while it is Anders Muszta and Martin Richter that have contributed with the actual content of the paper, on equal terms. On a more detailed level, it is Martin that have contributed with the "raw material" to the paper, from an economical background, while it is Anders that has made a paper

of it, that can be submitted to a scientific journal, including remaking of virtually all proofs, by agreement with Martin Richter, J.M.P. Albin and Bjarne Astrup Jensen, to make it a joint publication.

Bibliography

[1] Burrage, K., Burrage, P.M. and Tian, T. (2004). Numerical methods for strong solutions of stochastic differential equations: an overview. *Proc. Roy. Soc. London Ser. A* **460** 373-402.

[2] Chan, K., Koralyi, G., Longstaff, F. and Sanders, A. (1992). An empirical comparison of alternative models of the short-term interest rate. *J. Finance* **47** 1209-1227.

[3] Gyöngy, I. (1998). A note on Euler approximations. *Potential Anal.* **8** 205-216.

[4] Karatzas, I. and Shreve, S.E. (1991). *Brownian Motion and Stochastic Calculus*. Springer, Berlin.

[5] Kloeden, P.E. and Platen, E. (1992). *Numerical Solution of Stochastic Differential Equations*. Springer, Berlin.

[6] Muszta, A. (2003). Recent advances in the numerical solutions of stochastic differential equations. Licentiate thesis, Department of Mathematics, Chalmers University of Technology.

`http://www.math.chalmers.se/~palbin/licmuszta.ps`

[7] Platen, E. (1999). An introduction to numerical methods for stochastic differential equations. *Acta Numer.* **8** 197-246.

[8] Richter, M. (2004). A study of stochastic differential equations with volatility induced stationarity. Unpublished manuscript.

[9] Schuëller, G.I., Ed. (1997). A state-of-the-art report on computational stochastic mechanics. *Probabilistic Engineering Mechanics* **12** 197-321.

PAPER I

RECENT ADVANCES IN THE NUMERICAL SOLUTION OF STOCHASTIC DIFFERENTIAL EQUATIONS*

ANDERS MUSZTA† AND J. M. P. ALBIN‡

Abstract. This is a survey of numerical methods to obtain approximate strong solutions of stochastic differential equations. The focus is on methods that have emerged the last decade, that are not documented in monographs.

Key words. adaptive method, Euler method, geometric integration, Lie group method, Milstein method, non-linear stochastic differential equation, stochastic differential equation, stochastic Taylor expansion, stochastic delay equation, strong approximation, symplectic integration, theta method, waveform relaxation

AMS subject classifications. 60H10, 60H35, 60J60, 65C30, 68U20.

Introduction. This overview article gives the reader the main ideas, of some recent techniques in strong numerical solutions of stochastic differential equations (`sde`), driven by Wiener processes. Many of these techniques are so new that only a few articles have been written on them.

Mainly, techniques that have appeared from around 1992 are considered, since this was the year of publication of Kloeden and Platen [29]. However, for completeness, a few of the most basic older methods are described as well. Note that what is new in the field of numerical solution of `sde`, often has been employed for some time in the numerical solution of ordinary differential equations (`ode`), as many of the latter techniques have proven themselves adaptable to `sde`.

Targeted readers include novel researchers in numerical `sde`, as well as probabilists and mathematicians in general, who will get an economical overview of recent developments in the field, and a guide for selection of further literature to study. Readers with an applied interest in `sde`, will hopefully be able to use the article as an aid, to choose candidate numerical methods.

It has been a goal to keep the treatment as uncomplicated as possible. In order to make the article accessible to a wide audience, short introductions are given to the most vital concepts in connection with `sde`, as well as to some argubly non-basic mathematical tools that are used.

There have appeared other surveys in recent years on strong numerical solution of `sde`: In 1999 Platen published an overview [49] in which he present methods based on the stochastic Taylor formula from Kloeden and Platen [29] by . The overview is a condensed version of some of the topics covered in the book.

In 2004 Burrage, Burrage and Tian published an overview [6]. They also focus on methods based on the Stochastic Taylor Formula, and on implicit methods.

Compared to these survey articles, our survey focuses on adaptive methods, the stocahstic theta method, methods that attempt to preserve invariants of the origial equation (Lie group methods and Symplectic methods), methods designed for large systems of `sde` (Waveform relaxation methods), methods for non-linear `sde`, and methods for `sde` with memory: stochastic delay differential equations (`sdde`).

*Recived March 2005.

†Department of Mathematics, Chalmers University of Technology, 412 96 Göteborg, Sweden. (muszta@math.chalmers.se).

‡Department of Mathematics, Chalmers University of Technology, 412 96 Göteborg, Sweden. (palbin@math.chalmers.se), supported by The Swedish Research Council grant no. 621-2003-5214.

Contents

1 Prerequisites: Stochastic integration and `sde` 2
- 1.1 Measurability, filtrations and the Wiener process 2
- 1.2 Stochastic integrals 3
- 1.3 Stochastic differential equations 4
- 1.4 Itô formula 5

2 Strong approximations of `sde` 5
- 2.1 Strong approximations 6
- 2.2 Stochastic Taylor expansion 7
- 2.3 The Euler method 8
- 2.4 The Milstein method 8

3 Non-linear `sde` 8

4 The stochastic theta method 9

5 Stochastic delay differential equations 10

6 Waveform relaxation 11
- 6.1 Waveform relaxation for `ode` 11
- 6.2 Waveform relaxation for `sde` 11
- 6.3 Waveform relaxation for a stochastic sine-cosine system 13

7 Adaptive methods 14

8 Symplectic methods 15
- 8.1 Prerequisites: Symplectic transformations and Hamiltonian systems 16
- 8.2 Symplectic Runge-Kutta methods 16
- 8.3 Symplectic methods for `sde` 17

9 Lie group methods 18
- 9.1 Prerequisites: Lie groups 19
- 9.2 The Runge-Kutta-Munthe-Kaas scheme 20
- 9.3 The Magnus formula 20
- 9.4 The Magnus formula applied to `sde` 21
- 9.5 The Kunita formula 22
- 9.6 Applying the Kunita formula 24

Bibliography 25

1. Prerequisites: Stochastic integration and `sde`. The material in this chapter is intended for easy reference, and to fix the notation. See e.g., Karatzas and Shreve [28] as a basic reference.

1.1. Measurability, filtrations and the Wiener process. The definition of the Wiener process, in the context of `sde`, requires some care:

DEFINITION 1.1. *An $\mathbb{R}^d$-valued stochastic process $\mathbb{X} = \{\mathbb{X}_t\}_{t\geq 0} = \{((X_1)_t, \ldots, (X_d)_t)\}_{t\geq 0}$, defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, is* measurable *if*

- $\mathbb{X} : \Omega \times [0,\infty) \to \mathbb{R}^d$ *is* $\mathcal{F} \times \mathcal{B}([0,\infty))$*-measurable.*

DEFINITION 1.2. *A* filtered probability space *is a probability space* $(\Omega, \mathcal{F}, \mathbf{P})$*, equipped with a non-decreasing family of* σ*-algebras* $\{\mathcal{F}_t\}_{t\geq 0} \subset \mathcal{F}$*, a* filtration.

DEFINITION 1.3. *A filtration* $\{\mathcal{F}_t\}_{t\geq 0}$ *is* right-continuous*, if*

- $\forall t \geq 0 : \ \mathcal{F}_t = \bigcap_{\epsilon>0} \mathcal{F}_{t+\epsilon}$.

DEFINITION 1.4. *An filtered probability space* $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbf{P})$ *is* augmented*, if* $(\Omega, \mathcal{F}, \mathbf{P})$ *is complete, and* $\mathcal{F}_0$ *contains all* $\mathbf{P}$*-null sets of* $\mathcal{F}$*.*

DEFINITION 1.5. *An* $\mathbb{R}^d$*-valued (stochastic) process,* $\{\mathbb{X}_t\}_{t\geq 0}$*, on a filtered probability space* $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbf{P})$*, is* adapted *if* X_t *is* $\mathcal{F}_t$*-measurable for* $t \geq 0$*.*

DEFINITION 1.6. *A (standard)* $\mathbb{R}$*-valued* Wiener process*, is a continuous and adapted process* $W = \{W_t\}_{t\geq 0}$*, defined on a filtered probability space* $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbf{P})$*, starting at* $W_0 = 0$*, such that such that*

- $\forall s, t \geq 0 : \ W_t - W_s$ *is independent of (events in)* $\mathcal{F}_s$*;*
- $\forall s, t \geq 0 : \ W_t - W_s$ *is zero-mean normal distributed with variance* $t - s$*.*

A (standard) $\mathbb{R}^d$-valued *Wiener process* $\mathbb{W} = \{\mathbb{W}_t\}_{t\geq 0}$, is a process whose components $\{(W_1)_t\}_{t\geq 0}, \ldots, \{(W_d)_t\}_{t\geq 0}$ are independent Wiener processes.

Henceforth, a right-continuous and augmented filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbf{P})$ is assumed given, on which stochastic processes and random variables that feature are defined.

1.2. Stochastic integrals. Stochastic integration gives sense to the stochastic integral, $\int_0^t f_s \, dg_s$ or $\int_0^t f_s \circ dg_s$, of an adapted and measurable $\mathbb{R}$-valued processes $f = \{f_t\}_{t\geq 0}$, with respect to an adapted and continuous $\mathbb{R}$-valued process $g = \{g_t\}_{t\geq 0}$. The integral can be defined as a Riemann-Stieltjes (`RS`) integral if f has bounded variation, by integration by parts. The challenge lies in a useful construction for more general f.

Integrals of $\mathbb{R}^d$- or matrix $\mathbb{R}_{d|d}$-valued processes, with respect to $\mathbb{R}$- or $\mathbb{R}^d$-valued processes, are defined by (multiplying and) considering components.

The construction is based on first integrating *simple processes* $f_t = \sum_{k=1}^{\infty} c_{k-1} \, I_{[t_{k-1}, t_k)}(t)$, for some partition $0 = t_0 < t_1 < \cdots < t_k \uparrow \infty$ of $[0,\infty)$, and some sequence $\{c_k\}_{k=0}^{\infty}$ of random variables with $c_k \in \mathcal{F}_{t_k}$. Picking $p_k \in [0,1]$, and denoting $q_k = 1 - p_k$, a candidate integral is then given by

$$\sum_{k=1}^{n} (p_k c_{k-1} + q_k c_k)(g_{t_k \wedge t} - g_{t_{k-1} \wedge t}) = \sum_{k=1}^{\infty} (p_k f_{t_{k-1}} + q_k f_{t_k})(g_{t_k \wedge t} - g_{t_{k-1} \wedge t}). \quad (1.1)$$

The `RS` integral of a not necessarily simple f, is the unique limit of the sum (1.1), as $\max_{k\geq 1}(t_k - t_{k-1}) \to 0$, regardless of the choice of $p_k \in [0,1]$. But for the stochastic integral, the limit depends on that choice.

The Itô integral. of a simple f, is the integral with $p_k = 1$ in (1.1)

$$\int_0^t f_s \, dg_s = \sum_{k=1}^{\infty} f_{t_{k-1}} (g_{t_k \wedge t} - g_{t_{k-1} \wedge t}), \quad t \geq 0: \quad (1.2)$$

If the sum in (1.2) converges in probability (i.e., in measure) as $\max_{k\geq 1}(t_k - t_{k-1}) \to 0$, for a not necessarily simple f, then the limit is the *Itô integral* of f.

The Itô integral with respect to a Wiener process $\int_0^t f_s\,dW_s$, $t \geq 0$, exists when

$$\mathbf{P}\left\{\forall t \in [0,\infty) : \int_0^t f_s^2\,ds < \infty\right\} = 1. \tag{1.3}$$

In that case the integral is a continuous process with probability 1.

The Stratonovich integral, or *Fisk-Stratonovich integral*, of a simple f, is the integral with $p_k = \frac{1}{2}$ in (1.1)

$$\int_0^t f_s \circ dg_s = \sum_{k=1}^{\infty} \frac{f_{t_{k-1}} + f_{t_k}}{2}(g_{t_k \wedge t} - g_{t_{k-1}\wedge t}), \quad t \geq 0: \tag{1.4}$$

If the sum in (1.4) converges in probability, as $\max_{k\geq 1}(t_k - t_{k-1}) \to 0$, for a not necessarily simple f, then the limit is the *Stratonovich integral* of f.

The Stratonovich integral with respect to a Wiener process $\int_0^t f_s \circ dW_s$, $t \geq 0$, exists when f is a continuous *local martingale*, meaning that f is an Itô integral $f_t = \int_0^t \hat{f}_s\,d\hat{W}_s$, $t \geq 0$, for some $\hat{f}$ satisfying (1.3), and some Wiener process $\hat{W}$.

1.3. Stochastic differential equations.

Definition 1.7. *Given a constant $\mathbf{x}_0 \in \mathbb{R}^d$, and measurable functions*

$$a : [0,\infty) \times \mathbb{R}^d \to \mathbb{R}^d \qquad \text{and} \qquad b : [0,\infty) \times \mathbb{R}^d \to \mathbb{R}_{d|d}, \tag{1.5}$$

the following equation is a multidimensional Itô `sde` *(of diffusion type):*

$$d\mathbb{X}_t = a(t,\mathbb{X}_t)\,dt + b(t,\mathbb{X}_t)\,d\mathbb{W}_t, \quad t \geq 0, \qquad \mathbb{X}_0 = \mathbf{x}_0. \tag{1.6}$$

For $d = 1$, (1.6) becomes a one dimensional Itô `sde` *(of diffusion type):*

$$dX_t = a(t,X_t)\,dt + b(t,X_t)\,dW_t, \quad t \geq 0, \qquad X_0 = x_0, \tag{1.7}$$

Given a constant $\mathbf{x}_0 \in \mathbb{R}^d$, and measurable functions

$$a : \mathbb{R}^d \to \mathbb{R}^d \qquad \text{and} \qquad b : \mathbb{R}^d \to \mathbb{R}_{d|d}, \tag{1.8}$$

the following equation is an autonomous multidimensional Itô `sde`*:*

$$d\mathbb{X}_t = a(\mathbb{X}_t)\,dt + b(\mathbb{X}_t)\,d\mathbb{W}_t, \quad t \geq 0, \qquad \mathbb{X}_0 = \mathbf{x}_0. \tag{1.9}$$

For $d = 1$, (1.9) becomes an autonomous one dimensional Itô `sde`*:*

$$dX_t = a(X_t)\,dt + b(X_t)\,dW_t, \quad t \geq 0, \qquad X_0 = x_0. \tag{1.10}$$

Solutions to `sde`:s are called *diffusion processes*, and come in two kinds:

Definition 1.8. *Given a Wiener process $\mathbb{W}$, a measurable and adapted $\mathbb{R}^d$-valued process $\mathbb{X}$, is a* strong solution *to the multidimensional Itô* `sde` *(1.6), if*

$$\mathbf{P}\left\{\forall t \in [0,\infty) : \mathbb{X}_t = \mathbf{x}_0 + \int_0^t a(s,\mathbb{X}_s)\,ds + \int_0^t b(s,\mathbb{X}_s)\,d\mathbb{W}_s\right\} = 1. \tag{1.11}$$

Note that, in view of (1.3) (and with obvious notation), (1.11) requires that

$$\mathbf{P}\Big\{\forall t \in [0,\infty) : \int_0^t \|a(s,\mathbb{X}_s)\|\,ds + \int_0^t \|b(s,\mathbb{X}_s)\|^2\,ds < \infty\Big\} = 1.$$

DEFINITION 1.9. *The* `sde` (1.6) *has a* unique strong solution*, if any two strong solutions* $\mathbb{X}$ *and* $\mathbb{Y}$ *are* indistinguishable*, that is,*

$$\mathbf{P}\Big\{\forall t \in [0,\infty) : \mathbb{X}_t = \mathbb{Y}_t\Big\} = 1.$$

The `sde` (1.6) has a strong solution, that is unique, under suitable *Lipschitz* or *Hölder* type conditions, in the space variable, together with *linear growth* bounds, on the coefficients (1.5).

DEFINITION 1.10. *A right-continuous and augmented filtered probability space* $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbf{P})$*, together with a Wiener process* $\mathbb{W}$ *and an adapted* $\mathbb{R}^d$*-valued process* $\mathbb{X}$*, is a* weak solution *to the* `sde` (1.6)*, if* (1.11) *holds.*

The filtered probability space is a part of a weak solution, but given from the beginning for a strong solution. This seemingly minor difference is important.

We do not cover weak solutions in this article, as there exist a recent oversight of weak methods already, Schuëller [53], to which we do not have much to add.

A *multidimensional Stratonovich* `sde` (of diffusion type) is given by:

$$d\mathbb{X}_t = a(t,\mathbb{X}_t)\,dt + b(t,\mathbb{X}_t)\circ d\mathbb{W}_t, \quad t \geq 0, \qquad \mathbb{X}_0 = \mathbf{x}_0.$$

1.4. Itô formula. The following cornerstone result corresponds to the fundamental theorem of calculus, together with the change of variable formula and chain rule, in calculus:

THEOREM 1.11. (ITÔ FORMULA) *If* $\mathbb{X}$ *satisfies the* `sde` (1.6)*, and* $f : [0,\infty) \times \mathbb{R}^d \to \mathbb{R}$ *has continuous second order partial derivatives, then we have*

$$f(t,\mathbb{X}_t) = f(0,\mathbf{x}_0) + \int_0^t \Big(\frac{\partial}{\partial s} + \mathcal{A}\Big) f(s,\mathbb{X}_s)\,ds + \int_0^t (\nabla_{\mathbf{x}} f \cdot b)(s,\mathbb{X}_s)\,d\mathbb{W}_s, \qquad (1.12)$$

where $\mathcal{A}$ *is the* generator *of the* `sde`

$$\mathcal{A}f(t,\mathbf{x}) = \sum_{i=1}^d a_i(t,\mathbf{x})\frac{\partial f(t,\mathbf{x})}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^d (bb^T)_{ij}(t,\mathbf{x})\frac{\partial^2 f(t,\mathbf{x})}{\partial x_i x_j} \qquad (1.13)$$

(with b^T *denoting the transpose of* b*), and*

$$(\nabla_{\mathbf{x}} f \cdot b)(t,\mathbf{x}) = \left(\frac{\partial f(t,\mathbf{x})}{\partial x_1}, \ldots, \frac{\partial f(t,\mathbf{x})}{\partial x_d}\right) b(t,\mathbf{x}).$$

In the *Stratonovich formula*, for functions of solutions to Stratonovich `sde`, the Itô differential $d\mathbb{W}$ should be replaced with a Stratonovich differential $\circ d\mathbb{W}$ in (1.12), and the second derivatives should be omitted in (1.13). In other words, it is the formula one expects from the intuition of calculus.

2. Strong approximations of `sde`. As compared to the literature on weak solutions, much more is published on approximative strong solutions of `sde`, because of their close relation to `ode`.

2.1. Strong approximations. See Kloeden and Platen [29] as a basic reference for material in this section.

Let $\{\overline{\mathbb{X}}_{t_k}\}_{k=0}^n$ be approximations of the solution $\{\mathbb{X}_{t_k}\}_{k=0}^n$ to the `sde` (1.6), for

$$0 = t_0 < t_1 < \ldots < t_{n-1} < t_n = T \qquad \text{of} \qquad [0,T]. \tag{2.1}$$

DEFINITION 2.1. *A stochastic process* $\overline{\mathbb{X}} = \{\overline{\mathbb{X}}_t\}_{t\geq 0}$, *that is right continuous with left limits, is called a* strong approximation *of the* `sde` (1.6), *if*

$$\overline{\mathbb{X}}_{t_k} = f(\overline{\mathbb{X}}_{t_0}, \ldots, \overline{\mathbb{X}}_{t_{k-1}}, t_0, \ldots, t_{k-1}, \mathbb{W}_{t_0}, \ldots, \mathbb{W}_{t_k})$$

for some function f *making* $\overline{\mathbb{X}}_{t_k}$ $\mathcal{F}_{t_k}$*-measurable, for* $k \in \{0, n\}$.

Since errors accumulate, a measure of the accuracy of the approximation is obtained by comparing the final approximation $\overline{\mathbb{X}}_T$ with $\mathbb{X}_T$.

DEFINITION 2.2. *A strong approximation* $\overline{\mathbb{X}}$ converges strongly *to a solution* $\mathbb{X}$ *of the* `sde` (1.6), *if*

$$\mathbf{E}\big[\|\overline{\mathbb{X}}_T - \mathbb{X}_T\|\big] \to 0 \qquad as \qquad \max_{1\leq k\leq n}(t_{k+1} - t_k) \to 0.$$

The approximation is of strong order $p > 0$, *if*

$$\exists C > 0: \quad \mathbf{E}\big[\|\overline{\mathbb{X}}_T - \mathbb{X}_T\|\big] \leq C\Big(\max_{1\leq k\leq n}(t_{k+1} - t_k)\Big)^p.$$

The approximation is of mean-square order $p > 0$, *if*

$$\exists C > 0: \quad \sqrt{\mathbf{E}\big[\|\overline{\mathbb{X}}_T - \mathbb{X}_T\|^2\big]} \leq C\Big(\max_{1\leq k\leq n}(t_{k+1} - t_k)\Big)^p.$$

We close this section by citing a famous result on mean-square approximation:

THEOREM 2.3. (WONG AND ZAKAI [56], [57], IKEDA AND WATANABE [23]) *Consider the partition* (2.1) *of* $[0,T]$, *and define*

$$\overline{\mathbb{W}}_t = \mathbb{W}_{t_{k-1}} + \frac{t - t_{k-1}}{t_k - t_{k-1}}\big(\mathbb{W}_{t_k} - \mathbb{W}_{t_{k-1}}\big), \quad t \in [t_{k-1}, t_k], \quad k = 1, \ldots, n.$$

Let $\overline{\mathbb{X}}$ *solve the autonomous multidimensional* `sde`

$$d\overline{\mathbb{X}}_t = a(\overline{\mathbb{X}}_t)\,dt + b(\overline{\mathbb{X}}_t)\,d\overline{\mathbb{W}}_t, \quad t \in [0,T], \qquad \overline{\mathbb{X}}_0 = \mathbf{x}_0. \tag{2.2}$$

Further, let $\mathbb{X}$ *solve the autonomous multidimensional Stratonovich* `sde`

$$d\mathbb{X}_t = a(\mathbb{X}_t)\,dt + b(\mathbb{X}_t) \circ d\mathbb{W}_t, \quad t \in [0,T], \qquad \mathbb{X}_0 = \mathbf{x}_0.$$

If $a \in C^1(\mathbb{R}^d, \mathbb{R}^d)$ *and* $b \in C_B^2(\mathbb{R}^d, \mathbb{R}_{d|d})$, *then we have*

$$\mathbf{E}\Big[\max_{t\in[0,T]}\|\overline{\mathbb{X}}_t - \mathbb{X}_t\|^2\Big] \to 0 \qquad as \qquad \max_{1\leq k\leq n}(t_k - t_{k-1}) \to 0.$$

REMARK 2.4. Since $\overline{\mathbb{W}}$ has bounded variation, (2.2) is viewed in the `RS` sense.

2.2. Stochastic Taylor expansion. Stochastic Taylor expansions were introduced by Platen [48], and a thorough discussion can be found in Kloeden and Platen [29]. See also Pettersson [47].

Let X denote a solution to the autonomous one dimensional `sde` (1.10). If the coefficients a and b in (1.8) are two times continuously differentiable, then the Itô formula (1.12) gives the following representations of $a(X_t)$ and $b(X_t)$:

$$\begin{aligned} a(X_t) &= a(X_s) + \int_s^t (L_0 a)(X_r)\, dr + \int_s^t (L_1 a)(X_r)\, dW_r \\ b(X_t) &= b(X_s) + \int_s^t (L_0 b)(X_r)\, dr + \int_s^t (L_1 b)(X_r)\, dW_r \end{aligned} \tag{2.3}$$

Here the operators L_0 and L_1 are given by

$$L_0 = a\,\frac{d}{dx} + \frac{1}{2}\,b^2\,\frac{d^2}{dx^2} \qquad \text{and} \qquad L_1 = b\,\frac{d}{dx}. \tag{2.4}$$

Inserting (2.3) into the integral representation of a solution to (1.10), we get

$$X_t = X_s + \int_s^t a(X_r)\, dr + \int_s^t b(X_r)\, dW_r = X_s + \sum_{i=0}^{1} c_i(X_s) \int_s^t dW_r^{(i)} + R_1, \tag{2.5}$$

where the principal form of the remainder R_1 is given by

$$R_1 = \int_s^t \int_s^{r_1} \sum_{i,j=0}^{1} (L_j c_i)(X_{r_2})\, dW_{r_2}^{(j)} dW_{r_1}^{(i)}.$$

Here we use the notation $c_0 = a$, $c_1 = b$, $dW_r^{(0)} = dr$ and $dW_r^{(1)} = dW_r$.

Applying L_0 and L_1 to (2.3) in turn, if the coefficients a and b are sufficiently differentiable, we get a representation of X_t in terms of iterated integrals:

$$X_t = X_s + \sum_{i=0}^{1} c_i(X_s) \int_s^t dW_r^{(i)} + \sum_{i,j=0}^{1} (L_j c_i)(X_s) \int_s^t \int_s^{r_1} dW_{r_2}^{(j)} dW_{r_1}^{(i)} + R_2, \tag{2.6}$$

where the remainder R_2 is given by

$$R_2 = \int_s^t \int_s^{r_1} \int_s^{r_2} \sum_{i,j,k=0}^{1} (L_j L_k c_i)(X_{r_3})\, dW_{r_3}^{(k)} dW_{r_2}^{(j)} dW_{r_1}^{(i)}.$$

For a and b smooth enough, the expansions (2.5) and (2.6) can be carried on, to get remainders of increasingly high order. The expansion can be extended to multidimesional `sde` in an obvious manner. Expansions of this type, by iterated integrals, are called *Itô-Taylor expansions*. *Stratonovich-Taylor expansions* are analogous, except that they use Stratonovich integrals instead of Itô integrals.

The building blocks in Itô-Taylor expansions are iterated integrals, of the form

$$I_t^{(i_1,\ldots,i_n)} = \int_0^t \int_0^{r_1} \cdots \int_0^{r_{n-1}} dW_{r_n}^{(i_n)} \ldots dW_{r_2}^{(i_2)} dW_{r_1}^{(i_1)}, \quad i_1, \ldots, i_n \in \{0, 1\}.$$

With the *Hermite polynomials*

$$\mathcal{H}_n(x, y) = \sqrt{n!}\, y^{n/2} H_n\Big(\frac{x}{\sqrt{y}}\Big) \qquad \text{where} \qquad H_n(x) = (-1)^n\, \mathrm{e}^{x^2}\, \frac{d^n}{dx^n} \mathrm{e}^{-x^2},$$

the iterated Itô integrals can be calculated as, see e.g., Karatzas and Shreve [28],

$$I_t^{(1,\dots,1)} = \mathcal{H}_n(W_t, t)/(n!). \tag{2.7}$$

For multidimensional `sde` stochastic Taylor expansions involve iterated integrals with different components of a multidimensional Wiener process. See Milstein [38] and Ryden and Wiktorsson [50] on approximation of iterated integrals.

2.3. The Euler method. The simplest numerical scheme, based on stochastic Taylor expansion, is a stochastic version of the Euler iteration for numerical solution of `ode`. It uses the expansion (2.5) with the approximation $R_1 \approx 0$, which gives

$$\overline{X}_{t_k} = \overline{X}_{t_{k-1}} + a(\overline{X}_{t_{k-1}})\, h + b(\overline{X}_{t_{k-1}})\, (W_{t_k} - W_{t_{k-1}}), \quad k = 1, \dots, n. \tag{2.8}$$

The *Euler approximation* $\overline{X} = \{\overline{X}_{t_k}\}_{k=0}^n$, suggested by Maruyama [36]. Kloeden and Platen [29] show that it is of strong order $\frac{1}{2}$.

Euler approximation is attractive, since it is quite amenable to theoretical analysis. From a computational point of view, it is very simple to implement, but requires a finer partition (2.1), to give accurate approximations, than do higher order methods.

2.4. The Milstein method. The *Milstein approximation* differs from the Euler approximation, in that it uses one of the integrals from the second sum in (2.6):

$$X_t \approx X_s + a(X_s)\int_s^t dr + b(X_s)\int_s^t dW_r + (L_1 b)(X_s)\int_s^t \int_s^{r_1} dW_{r_2}\, dW_{r_1}.$$

The idea is that, since $dW_t^2 = dt$, in a sense, the error in resulting approximation should be $o(dt)$, instead of $o(\sqrt{dt})$, as for the Euler-Maryuama approximation,.

Using an equidistant partition (2.1) of $[0, T]$, and using Theorem 2.7, the Milstein approximation $\{\overline{X}_{t_k}\}_{k=0}^n$, of a solution X to the autonomous one dimensional `sde` (1.10), at the nodes $\{X_{t_k}\}_{k=0}^n$, is given by the scheme

$$\overline{X}_{t_{k+1}} = \overline{X}_{t_k} + a(\overline{X}_{t_k})\, h + b(\overline{X}_{t_k})(W_{t_{k+1}} - W_{t_k}) + \frac{b(\overline{X}_{t_k})b'(\overline{X}_{t_k})}{2}\big((W_{t_{k+1}} - W_{t_k})^2 - h\big). \tag{2.9}$$

Kloeden and Platen [29] prove that this scheme is of strong order 1.

3. Non-linear `sde`. Virtually all numerical methods for `sde` require that the coefficients satisfy global Lipschitz conditions, and have linear growth. However, recently there has been some progress in weakening these assumptions.

Consider the `sde` (1.6). Assume that the coefficients a and b in (1.5) are locally bounded and locally Lipscitz continuous, and that, with obvious notation,

$$\exists\, M > 0 : \forall\, x \in \mathbb{R}^d : \quad \langle x, a(t,x)\rangle + \operatorname{trace}(bb^T) \leq M(|x|^2 + 1).$$

Consider the partition (2.1), with uniform stepsize $h = t_{k+1} - t_k$, together with the Euler type approximating process $\overline{\mathbb{X}}$, defined by

$$d\overline{\mathbb{X}}_t = a(t, \overline{\mathbb{X}}_{\kappa(t)})\, dt + b(t, \overline{\mathbb{X}}_{\kappa(t)})\, d\mathbb{W}_t \quad \text{where} \quad \kappa(t) = t_k \quad \text{for } t \in [t_k, t_{k+1}).$$

Gyöngy [16] shows that there exists a random variable η such that, for $\gamma < 1/2$,

$$\mathbf{P}\Big\{ \sup_{t\in[0,T]} |\mathbb{X}_n(t) - \mathbb{X}(t)| \leqslant \eta h^\gamma \Big\} = 1.$$

Higham, Mao and Stuart [19] consider the `sde` (1.9). They consider three numerical schemes for (1.9), based on the partition (2.1), with uniform stepsize $h = t_{k+1} - t_k$:

$$\begin{cases} \overline{X}_{t_{k+1}} = \overline{X}_{t_k} + a(\overline{X}_{t_k})h + b(\overline{X}_{t_k})(W_{t_{k+1}} - W_{t_k}) & \text{(Euler)} \\ \begin{cases} \overline{X}_{t_{k+1}} = \overline{X}^*_{t_k} + b(\overline{X}^*_{t_k})(W_{t_{k+1}} - W_{t_k}) \\ \overline{X}^*_{t_k} \ = \overline{X}_{t_k} + a(\overline{X}^*_{t_k})h \end{cases} & \text{(Split-step backward Euler)} \\ \overline{X}_{t_{k+1}} = \overline{X}_{t_k} + a(\overline{X}_{t_{k+1}})h + b(\overline{X}_{t_k})(W_{t_{k+1}} - W_{t_k}) & \text{(Backward Euler).} \end{cases} \tag{3.1}$$

Let $\overline{X}_t$ be the continuous process, interpolating the values $\overline{X}_{t_k}$ in (3.1), given by

$$\overline{X}_t = \overline{X}_{t_k} + a(\overline{X}_{t_k})(t - t_k) + b(\overline{X}_{t_k})(W_t - W_{t_k}) \quad \text{for } t \in [t_k, t_{k+1}). \tag{3.2}$$

Higham, Mao and Stuart [19], Theorem 4.4, show that, if a is $\mathcal{C}^1$, satisfies a one-sided Lipschitz condition [see (6.5) below], and has at most polynomial growth, and if b is $\mathcal{C}^1$ and satisfies a global Lipschitz condition, and if for every $p \geq 1$,

$$\mathbf{E}\Big\{ \sup_{t\in[0,T]} |X_t|^p \Big\} < \infty, \quad \mathbf{E}\Big\{ \sup_{t\in[0,T]} |\overline{X}_t|^p \Big\} < \infty \quad \text{and} \quad \mathbf{E}\Big\{ \sup_{0\leq k\leq n} |\overline{X}_{t_k}|^p \Big\} < \infty, \tag{3.3}$$

then the continuous Euler approximation, given by (3.1) and (3.2), satisfies

$$\mathbf{E}\Big\{ \sup_{t\in[0,T]} |\overline{X}_t - X_t|^2 \} = O(h) \quad \text{as } h \downarrow 0. \tag{3.4}$$

Under the same conditions, except that (3.3) is no longer needed, Higham et al., Theorems 4.7 and 5.3, show that the continuous split-step backward Euler method, and the continuous backward Euler method, given by (3.1) and (3.2), satisfy (3.4).

In addition, Higham et al., Theorem 2.2, show that (3.3) for some $p > 2$, together with local Lipschitz conditions for the coefficients, imply that $\sup_{t\in[0,T]} |\overline{X}_t - X_t| \to 0$ in mean-square as $h \downarrow 0$.

4. The stochastic theta method. Consider the `sde` (1.10). Let $0 = t_0 < t_1 < \ldots$ be a uniform partition of $[0, \infty)$, with stepsize $h = t_{n+1} - t_n$. The *stochastic theta method*, as presented by Higham [18], is a semi-implicit method for computing approximations, $\overline{X}_{t_n}$ to X_{t_n}, given an $\theta \in [0, 1]$, as

$$\overline{X}_{t_{n+1}} = \overline{X}_{t_n} + \big((1-\theta)a(\overline{X}_{t_n}) + \theta a(\overline{X}_{t_{n+1}})\big)h + b(\overline{X}_{t_n})(W_{t_{n+1}} - W_{t_n}). \tag{4.1}$$

For $\theta = 0$, (4.1) is the Euler method (2.8), while $\theta = 1$ gives the implicit Euler method. Bryden and Higham [3] notice that the theta method has mean-square order 1/2.

One reason for using the stochastic theta method is preservation of *asymptotic* properties. By this we understand that we want approximations $\overline{X}_{t_n}$ to share some asymptotic property with the exact solution.

Approximations preserving the limit $\lim_{t\to\infty} |X(t)|$ are called *asymptotically stable* (*A-stable*), and those preserving $\lim_{t\to\infty} \mathbf{E}\{|X(t)|^2\}$ *mean-square stable*. See Kloeden and Platen [29] on more information on stability.

The key question is for which stepsizes h the numerical method (4.1) for (1.10) is stable. For an answer, Higham [18] studies the test equation

$$dX_t = \lambda X_t\, dt + \mu X_t\, dW_t, \tag{4.2}$$

where $\lambda, \mu \in \mathbb{C}$. This test equation is selected to show features of more complicated problems, but yet being simple enough to enable a theoretical analysis.

If $\mathbf{P}\{X_0 \neq 0\} = 1$, then the solution to (4.2) has the asymptotic properties

$$\lim_{t\to\infty} \mathbf{E}\{|X(t)|^2\} = 0 \quad \Leftrightarrow \quad \mathtt{Re}(\lambda) + |\mu|^2/2 < 0,$$
$$\mathbf{P}\Big\{ \lim_{t\to\infty} |X(t)| = 0 \Big\} = 1 \quad \Leftrightarrow \quad \mathtt{Re}(\lambda - \mu^2/2) < 0.$$

Higham characterizes for which h the approximation (4.1) shares these asymptotics.

The theta method can also be applied to the Milstein scheme (2.9) of (4.2)

$$\overline{X}_{t_{n+1}}$$
$$= \overline{X}_{t_n} + \big((1-\theta)\lambda\overline{X}_{t_n} + \theta\lambda\overline{X}_{t_{n+1}}\big)h + \mu\overline{X}_{t_n}(W_{t_{n+1}} - W_{t_n}) + \frac{\mu^2}{2}\overline{X}_{t_n}\big((W_{t_{n+1}} - W_{t_n})^2 - h\big).$$

In the fashion of Higham [18], Bryden and Higham [3] study stability of this sheme.

5. Stochastic delay differential equations. A *stochastic delay differential equation* (`sdde`) is an `sde` of non-diffusion type

$$dX_t = a(X_t, X_{t-\tau})\,dt + b(X_t, X_{t-\tau})\,dW_t, \quad X_t = \Psi(t) \quad \text{for } t \in [-\tau, 0]. \tag{5.1}$$

Buckwar [4] shows that, under the usual global Lipschitz continuity and linear growth conditions on the coefficients a and b in (5.1), the Euler scheme

$$\overline{X}_{t_{n+1}} = \overline{X}_{t_n} + a(\overline{X}_{t_n}, \overline{X}_{t_{n-N_\tau}})h + b(\overline{X}_{t_n}, \overline{X}_{t_{n-N_\tau}})(W_{t_{n+1}} - W_{t_n}),$$

is of mean-square order 1. Here N_τ is an integer such that $\tau = N_\tau h$, and $h = t_{n+1} - t_n$ the uniform stepsize.

Baker and Buckwar [1] consider a special case of (5.1); the stochastic pantograph equation

$$dX_t = (a_1 X_t + a_2 X_{qt})\,dt + (b_1 + b_2 X_t + b_3 X_{qt})\,dW_t,$$

where $q \in (0,1)$ and $a_1, a_2, b_1, b_2, b_3 \in \mathbb{R}$. They use a theta method that is implicit in the drift and explicit in the diffusion part, to find a numerical approximation for this equation, which they demonstrate is mean-square order $1/2$.

Küchler and Platen [30] consider an `sdde` driven by more than one Wiener process

$$dX_t = a(X_t, X_{t-\tau})\,dt + \sum_{p=1}^{m} b_p(X_t, X_{t-\tau})\,dW^p(t), \quad X_t = \Psi(t) \quad \text{for } t \in [-\tau, 0].$$

They construct an Euler method of mean-square order $1/2$ and a Milstein-type method of mean-square order 1. To facilitate the analysis, they suggest a novel idea of expressing the solution of the `sdde` as a sequence of multi-dimensional `sde`.

Mao and Sabanis [35] find an Euler method for an `sdde`, with linear growth conditions on the coefficients, but only requiring *local Lipschitz* conditions. They also consider a variable Lipschitz continuous delay $\tau(t)$. Their Euler method is of mean-square order $\gamma/2$, when $\Psi(t)$ is Hölder continuous of order $\gamma \in (0,1]$.

Hu, Mohammed and Yan [22] construct a Milstein method for the `sdde`

$$dX_t = a(t, X_{t-\tau_1}, \ldots, X_{t-\tau_k})\,dt + b(t, X_{t-\pi_1}, \ldots, X_{t-\pi_m})\,dW_t,$$

assuming globally Lipschitz coefficients a and b with linear growth. They show that the method is of mean-square order 1, using Malliavin calculus (see e.g., Nualart [45]), and the anticipating stochastic calculus of Nualart and Pardoux [46].

6. Waveform relaxation. We first describe the idea of waveform relaxationa in the ode context that it originates from, and then carry it over to sde.

The next example illustrates two aspects of an ode indicating that it may be useful to employ waveform relaxation methods: Stiffness and weakly interacting subsystems.

EXAMPLE 6.1. Consider the following two dimensional ode:

$$\begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix}' = \begin{pmatrix} -1 & 0 \\ 0 & -100 \end{pmatrix} \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix}, \quad t \geq 0, \qquad \begin{pmatrix} y_1(0) \\ y_2(0) \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \tag{6.1}$$

When solving this equation numerically, a relatively small time step is required to capture the quick variations of y_2. For y_1 this is a waste, with its slower clock.

However, the ode can be broken up into two separate, non-interacting odes:

$$y_1'(t) = -y_1(t), \quad y_1(0) = 1 \qquad \text{and} \qquad y_2'(t) = -100\, y_2(t), \quad y_2(0) = 1.$$

Hence a numerical method that computes the components of the system (6.1) in parallel can be used, with time steps individually adapted to the components.

Of course, the fact that the function y_2 decreases much faster than y_1 origins in that there are two widely separated eigenvalues, $\lambda_1 = -1$ and $\lambda_2 = -100$, of the coefficient matrix of the sde: A system of ode, $\mathbf{y}_t' = \mathbf{A}\mathbf{y}$, is called *stiff* if there are eigenvalues λ_i and λ_j of the coefficient matrix $\mathbf{A}$ such that $|\lambda_i| << |\lambda_j|$.

6.1. Waveform relaxation for ode. Consider the multidimensional ode

$$\mathbf{y}_t' = \mathbf{f}(t, \mathbf{y}_t), \quad t \in [0, T] \ \text{ and } \ \mathbf{y}_t \in \mathbb{R}^d. \tag{6.2}$$

Let $\{\mathbf{y}_t^{(n)}\}_{n=1}^N$ be a sequence of approximate solutions to (6.2). The elements of this sequence are exact solutions to a system of ode, obtained by means of a function, $\mathcal{T} : [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$.

$$(\mathbf{y}^{(n)})_t' = \mathcal{T}(t, \mathbf{y}_t^{(n-1)}, \mathbf{y}_t^{(n)}), \quad \mathbf{y}_0^{(n)} = \mathbf{y}_0. \tag{6.3}$$

The function $\mathcal{T}(t, \mathbf{x}, \mathbf{z})$ is chosen so that $\mathcal{T}(t, \mathbf{z}, \mathbf{z}) = \mathbf{f}(t, \mathbf{z})$ and the functional matrix $\partial_{\mathbf{z}}\mathcal{T}$ is block-diagonal.

Two examples of of the *iteration function* $\mathcal{T}(t, \mathbf{x}, \mathbf{z})$ are

$$\mathcal{T}_k(t, \mathbf{x}, \mathbf{z}) = f_k(t, x_1, \ldots, x_d) \qquad \text{(Picard)}$$
$$\mathcal{T}_k(t, \mathbf{x}, \mathbf{z}) = f_k(t, z_1, \ldots, z_k, x_{k+1}, \ldots, x_d) \quad \text{(Gauss-Seidel)}$$

Picard schemes give decoupled systems that can be solved in parallel, while for Gauss-Seidel schemes the approximate solutions $\mathbf{y}^{(n)}$ are computed sequentially.

High-dimensional ode are used, for example, for VLSI (Very Large Scale Integrated) circuits. It was in this context the waveform relaxation method was developed, by Lelarasmee [34]. See Vandewalle [55] for a text book presentation. Some current research on waveform relaxation for ode focus on acceleration of convergence. See e.g., Leimkühler [33] and Janssen and Vandewalle [27].

6.2. Waveform relaxation for sde. Introduce a splitting $\mathbb{R}^d = \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_n}$ of the d-dimensional sde (1.6)

$$d((\mathbb{X}_{(j)})_t) = f_{j0}(t, \mathbb{X}_t)dt + \sum_{k=1}^{d} f_{jk}(t, \mathbb{X}_t)\, d(W_k)_t, \quad j = 1, \ldots, n. \tag{6.4}$$

For (1.6) to have a unique solution, we need conditions on the functions.

Let $\mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_n}$ for some $d_1, \dots, d_n \in \mathbb{N}$.

A function $f_0 = (f_{10}, \dots, f_{n0}) : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ is said to be *uniformly one-sided Lipschitz continuous* if $\forall t \in [0,T]$, $\forall k \in \{1, \dots, n\}$, $\exists \hat{L}^{(1)}_{k0}, \dots, \hat{L}^{(n)}_{k0} > 0$, $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d$,

$$\Big(f_{k0}(t, \mathbf{x}) - f_{k0}(t, \mathbf{y}),\, \mathbf{x}_{(k)} - \mathbf{y}_{(k)} \Big)_{d_k} \leqslant \sum_{i=1}^{n} \hat{L}^{(i)}_{k0} ||\mathbf{x}_{(i)} - \mathbf{y}_{(i)}||^2_{d_i} \tag{6.5}$$

A function $f_0 = (f_{10}, \dots, f_{n0}) : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ is said to be *globally Lipschitz continuous* if $\forall t \in [0,T]$, $\forall k \in \{1, \dots, n\}$, $\forall r \in \{1, \dots, d\}$, $\exists L^{(1)}_{kr}, \dots, \hat{L}^{(n)}_{kr} > 0$, $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d$,

$$\Big(f_{kr}(t, \mathbf{x}) - f_{kr}(t, \mathbf{y}),\, \mathbf{x}_{(k)} - \mathbf{y}_{(k)} \Big)_{d_k} \leqslant \sum_{i=1}^{n} L^{(i)}_{kr} ||\mathbf{x}_{(i)} - \mathbf{y}_{(i)}||^2_{d_i} \tag{6.6}$$

A function $f : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ is said to be *uniformly anti-coercive* if

$$\exists C \in \mathbb{R} : \forall t \in [0,T],\, \forall \mathbf{x} \in \mathbb{R} : \quad (f(t, \mathbf{x}), \mathbf{x})_d \leqslant C(1 + ||\mathbf{x}||^2_d).$$

The space to which the exact solution belongs, along with the approximations to be discussed below, is denoted $\mathcal{E}^p$, for some $p \geqslant 1$ and is defined as follows:

$$\mathcal{E}^p = \Big\{ \mathbb{X} \in \mathcal{C}\big([0,T], \mathbb{R}^d\big) : \|\mathbb{X}\|_{\mathcal{E}^p} < \infty \Big\} \quad \text{where} \quad \|\mathbb{X}\|_{\mathcal{E}^p} = \mathbf{E}\Big[\sup_{t \in [0,T]} \|\mathbb{X}_t\|^p_d \Big].$$

The following theorem is due to Schneider and Schurz [52].

THEOREM 6.2. *Pick an $p \geqslant 2$. Let the functions f_{i0} of the coefficient function a in* (6.4) *be uniformly one-sided Lipschitz continuous as well as uniformly anti-coercive. Let the functions f_{ij}, $i > 0$, of the coefficient matrix b in* (6.4) *be globally Lipschitz continuous and bounded. If $\mathbf{E}\{||\mathbf{x}^{(i)}_0||^p_{d_i}\} < \infty$ for $k \in \{1, \dots, n\}$, then the system* (6.4) *has a unique, $\mathcal{F}_t$-adapted and continuous solution in $\mathcal{E}^p$.*

Next we come to the question of whether a waveform-relaxed approximate solution to (6.4) converges to the exact solution in $\|\cdot\|_{\mathcal{E}^p}$. We mention the Picard and the Gauss-Seidel waveform relaxation schemes.

DEFINITION 6.3. *The* Picard scheme *for* (6.4) *is the following iterative scheme to compute an approximation $\mathbb{X}^{(\cdot)}_t$ to the exact solution $\mathbb{X}_t$:*

$$\begin{aligned} (\mathbb{X}^{(m)}_{(j)})_t = (\mathbf{x}_{(1)})_0 &+ \int_0^t f_{j0}(s, (\mathbb{X}^{(m-1)}_{(1)})_s, \dots, (\mathbb{X}^{(m-1)}_{(n)})_s)\, ds \\ &+ \sum_{k=1}^{d} \int_0^t f_{jk}(s, (\mathbb{X}^{(m-1)}_{(1)})_s, \dots, (\mathbb{X}^{(m-1)}_{(n)})_s)\, d(W_k)_s. \end{aligned} \tag{6.7}$$

DEFINITION 6.4. *The* Gauss-Seidel *scheme for* (6.4) *is the following iterative scheme to compute an approximation $\mathbb{X}^{(\cdot)}_t$ to the exact solution $\mathbb{X}_t$:*

$$\begin{aligned} (\mathbb{X}^{(m)}_{(j)})_t = (\mathbf{x}_{(1)})_0 &+ \int_0^t f_{j0}(s, (\mathbb{X}^{(m)}_{(1)})_s, \dots, (\mathbb{X}^{(m)}_{(j-1)})_s, (\mathbb{X}^{(m-1)}_{(j)})_s \dots, (\mathbb{X}^{(m-1)}_{(n)})_s)\, ds \\ &+ \sum_{k=1}^{d} \int_0^t f_{jk}(s, (\mathbb{X}^{(m)}_{(1)})_s, \dots, (\mathbb{X}^{(m)}_{(j-1)})_s, (\mathbb{X}^{(m-1)}_{(j)})_s \dots, (\mathbb{X}^{(m-1)}_{(n)})_s)\, d(W_k)_s. \end{aligned} \tag{6.8}$$

Recall that the *spectral radius* $\rho(\mathbf{A})$ of a matrix $\mathbf{A}$ is the largest modulus of any of the eigenvalues of A

Let $C_p > 0$ be the constant featuring in the *Burkholder-Davies-Gundy inequality*

$$C_p^{2/p} \leqslant \frac{p(p-1)}{2}\Big(\frac{p}{p-1}\Big)^p. \tag{6.9}$$

The following theorem is due to Schneider and Schurz [52].

THEOREM 6.5. *Let the functions* $\{f_{r0}\}_{r=1}^n$ *and* $(f_{rk})_{r,k=1,\dots,n}$ *satisfy the assumptions of Theorem 6.2.*

Consider the matrix $\hat{\mathbb{L}} = (\hat{l}_{jm})$ *of transformed Lipschitz coefficients*

$$(\hat{l}_{jm})^p = pT\sum_{k=1}^n \max\Big(0,\, \hat{L}_{k0}^{(m)} + \frac{n(p-1)}{2}\sum_{r=1}^d (L_{kr}^{(m)})^2\Big) + p\sqrt{T}C_p\sum_{r=1}^d\sqrt{n\sum_{k=1}^n (L_{kr}^{(m)})^2} \tag{6.10}$$

of the functions $\{f_{r0}\}_{r=1}^n$ *and* $(f_{rk})_{r,k=1,\dots,n}$, *from* (6.5) *and* (6.6). *If the spectral radius* $\rho(\hat{\mathbb{L}}) < 1$, *then the Picard scheme* (6.7) *converges in* $\mathcal{E}^p$.

Consider the matrix $\tilde{\mathbb{L}}$, *the general definition of which is too cumbersome to give here [but the two dimensional matrix is given in* (6.12) *below]. If the spectral radius* $\rho(\tilde{\mathbb{L}}) < 1$, *then the Gauss-Seidel scheme* (6.7) *converges in* $\mathcal{E}^p$.

The idea of the proofs of Theorems 6.2 and 6.5 are to show that the exact and approximate solutions are contractive maps from $\mathcal{E}^p$ to $\mathcal{E}^p$, in the fashion of proofs of fixed-point theorems for Banach spaces.

6.3. Waveform relaxation for a stochastic sine-cosine system. Consider a sine-cosine system with a random peturbation:

$$d\mathbb{X}_t = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\mathbb{X}_t\,dt + \begin{pmatrix} 0 & a \\ 0 & 0 \end{pmatrix}\mathbb{X}_t\,dW_t\,, \quad \mathbb{X}_0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad t \in [0,T]. \tag{6.11}$$

Approximate solutions to (6.11) based on the Gauss-Seidel scheme (6.8) are denoted $\{\mathbf{X}_t^{(n)}\}_{n=1}^N$, and given by

$$\begin{pmatrix} f_{10}(t,\mathbf{x}) & f_{11}(t,\mathbf{x}) \\ f_{20}(t,\mathbf{x}) & f_{21}(t,\mathbf{x}) \end{pmatrix} = \begin{pmatrix} x_2 & ax_2 \\ -x_1 & 0 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

The first two iterations of this scheme produces the following result.

$$\begin{cases} (X_1^{(1)})_t = t + (\Xi_1^{(1)})_t; \\ (X_2^{(1)})_t = 1 - \frac{t^2}{2!} + (\Xi_2^{(1)})_t; \end{cases} \Rightarrow \begin{cases} (X_1^{(2)})_t = t - \frac{t^3}{3!} + (\Xi_1^{(2)})_t; \\ (X_2^{(2)})_t = 1 - \frac{t^2}{2!} + \frac{t^4}{4!} + (\Xi_2^{(2)})_t \end{cases},$$

where the stochastic variables $(\Xi_i^{(j)})_t$ are defined by

$$\begin{aligned}
(\Xi_1^{(1)})_t &= aW_t, \qquad (\Xi_2^{(1)})_t = -a\int_0^t W_s\,ds; \\
(\Xi_1^{(2)})_t &= aW_t - \frac{a}{2}\int_0^t s^2\,dW_s - a^2\int_0^t\int_0^s W_r\,dr\,dW_s - a\int_0^t\int_0^s W_r\,dr\,ds; \\
(\Xi_2^{(2)})_t &= -a\int_0^t W_s\,ds + \frac{a}{2}\int_0^t\int_0^s r^2\,dW_r\,ds \\
&\quad + a\int_0^t\int_0^s\int_0^r W_u\,du\,dr\,ds + a^3\int_0^t\int_0^s\int_0^r W_u\,du\,dW_r\,ds.
\end{aligned}$$

The one-sided Lipschitz coefficients of f_{10} are computed as

$$\big(f_{10}(t,\mathbf{x}) - f_{10}(t,\mathbf{y}),\, x_1 - y_1\big)_{d_1} = (x_2 - y_2)(x_1 - y_1) \leqslant \frac{1}{2}|x_1 - y_1|^2 + \frac{1}{2}|x_2 - y_2|^2.$$

Thus $\hat{L}_{10}^{(1)} = \frac{1}{2} = \hat{L}_{10}^{(2)}$. Similar computations give $\hat{L}_{20}^{(1)} = \frac{1}{2} = \hat{L}_{20}^{(2)}$ for f_{20}.

The global Lipschitz coefficients of f_{11} and f_{21} are respectively $L_{11}^{(1)} = 0$, $L_{11}^{(2)} = |a|$ and $L_{21}^{(1)} = 0 = L_{21}^{(1)}$. The matrix $\tilde{\mathbb{L}}$ governing the convergence of the stochastic Gauss-Seidel scheme (6.8) can be computed as

$$\tilde{\mathbb{L}} = \begin{pmatrix} \hat{l}_{11} & \hat{l}_{12} \\ \hat{l}_{21}\hat{l}_{11} & \hat{l}_{22} + \hat{l}_{21}\hat{l}_{12} \end{pmatrix} = \begin{pmatrix} 2T & 2T(1+a^2) + 2\sqrt{2T}C_2|a| \\ 4T^2 & 2T(1+2T)(1+a^2) + (4T+2)\sqrt{2T}C_2|a| \end{pmatrix}, \tag{6.12}$$

where $\hat{l}_{ij}$ are computed using (6.10), with $p = 2$:

$$\hat{\mathbb{L}} = \begin{pmatrix} \hat{l}_{11} & \hat{l}_{12} \\ \hat{l}_{12} & \hat{l}_{22} \end{pmatrix} = \begin{pmatrix} 2T & 2T(1+a^2) + 2\sqrt{2T}C_2|a| \\ 2T & 2T(1+a^2) + 2\sqrt{2T}C_2|a| \end{pmatrix}.$$

The spectral radius of $\tilde{\mathbb{L}}$ is

$$\rho(\tilde{\mathbb{L}}) = 2T(2+a^2) + 2\sqrt{2T}C_2|a| \leqslant 2T(2+a^2) + 8\sqrt{2T}|a|,$$

where we have used the estimate (6.9) to get $C_2 \leqslant 4$. According to Theorem 6.5 we have convergence of the Gauss-Seidel scheme on the set

$$M = \{(\xi,\eta) \in [0,\infty) \times [0,\infty) : (\xi+4)^2 + \eta < 17\},$$

where $\xi = \sqrt{2Ta^2}$ and $\eta = 4T$, corresponding to the condition $\rho(\tilde{\mathbb{L}}) < 1$.

Note that we have considered convergence in $\mathcal{E}^2$ in this example.

7. Adaptive methods. The numerical approximation of one dimensional `sdes` driven by one dimensional Wiener processes W by adaptive methods make use of the Brownian bridge process. The corresponding problem for multidimensional `sde` is considerably more complex, and in fact, only the two-dimensional can be handled, as to date. The problem in higher dimensions is to simulate high dimensional iterated integrals, of many Wiener processes, conditional on the size of the incement of those Wiener processes, over the interval of integration.

The paper by Lamba [31] considers time-stepping based on both the size of the diffusion coefficient *and* the drift coefficient.

Instead of giving details of the adaptive techniques we stop at an outline, and refer the reader to the original articles for more information. Papers to be discussed: Gaines and Lyons [13] and [12], two papers by Hofmann, Müller-Grönbach and Ritter [20] and [21], Mauthner [37], Newton [43] and [44], and a paper by Lehn [32] comparing Mauthner and one of the papers by Hofmann, Müller-Grönbach and Ritter.

Mauthner: Two Runge-Kutta methods, one for large steps and the other for small steps. Simulates conditional multiple Ito integrals. This is the first time this is done. It is needed when working with high order adaptive numerical methods. Clark and Cameron [9] have shown that methods of order higher that 0.5 require iterated Itô integrals. And adaptive method has to be able to compute refinements to the iterated Ito integrals appearing in the method.

Newton: The adaptive time points are stopping times. They are the times when the Wiener process goes from a level $k\delta$ to $(k+1)\delta$ or $(k-1)\delta$. The Euler and

the Milstein scheme are contructed and investigated with respect to this adaptive partition scheme. The Euler scheme is of mean square order 0.5 whereas the Milstein scheme is of mean square order 1.

More precisely: Consider the multidimensional `sde`

$$d\mathbb{X}_t = a(t, \mathbb{X}_t)\, dt + b(t, \mathbb{X}_t)\, dW_t, \quad t \geq 0, \qquad \mathbb{X}_0 = \mathbf{x}_0.$$

Note that the Wiener process W is one dimensional. Fix an arbitrary $\delta > 0$ and define a sequence $\{\tau_n\}_{n=1}^{N}$ as follows:

$$\tau_{n+1} = \tau_n + \inf\{t \geqslant 0 : |W_{\tau_n + t} - W_{\tau_n}| = \delta\}, \quad n \in \{0, 1, \ldots, N\}, \quad \tau_0 = 0.$$

The methods mentioned so far, have assumed a fixed (non-random) partition (2.1), to compute approximate solutions of the `sde` in question. However, in the one dimensional `sde` (1.7), for example, the evolution of the solution X, is coupled with the driving noise W, through the diffusion coefficient b.

Hofmann, Müller-Grönbach and Ritter [21] suggest that, if the diffusion coefficient b is big, then a small step-size should be used, to capture the influence the noise W has on the solution X. But for b small, a small step-size is a waste of computational resources. An *adaptive partition* exploits the varying coupling between X and W, to improve the efficiency of the numerical method.

To find the next step length, one has to know b at the next time. This requires knowledge of X at that time. To cope with this, one uses a simple, computationally inexpensive explicit method, for example, the Euler method, to get approximations of upcoming values of X. Those values are then used in the main numerical scheme, to choose the right step size according to the size of b.

This method of predicting a future value, by a simple method, and then using that value to a correct a main scheme, is called a *predictor-corrector* method.

Hofmann, Müller-Grönbach and Ritter derive a lower bound on the error, that holds for any method that uses finitely many values of the Wiener process. Their adaptive scheme, using the the diffusion coefficient b as step size regulator, is then shown to achieve this lower bound.

Gaines and Lyons [13] adjusts the step size according to the size of the local error, as the solution is computed. The rationale is that local errors should contribute equally to the global error. They use methods that reduce the standard deviation of local errors, to become negligible compared with their mean, arguing that asymptotically efficient numerical schemes have this property.

Since the step size is continuously adjusted, the driving Wiener process has to be generated dynamically, as the solution is computed. Gaines and Lyons use a binary tree to store consecutive refinements of the path of the Wiener process.

Eriksson, Johnson and Logg [11] propose small stabilizing explicit steps, instead of an implicit method, for `ode`, and illustrate the efficiency of this idea on stiff problems. It seems that such results have not been used for `sde`, were numerical methods often are implicit. Similar to the work of Ericsson, Johnson and Logg are the works by Gear and Kevrekidis [14] and [15], albeit they investigate acceleration of convergence for stiff problems, rather than adaptive methods. The reason we mention these papers is that it may be worthwhile to consider adapting the techniques to `sde`.

8. Symplectic methods. If a numerical method for a *Hamiltonian* `ode` is sought, with a high accuracy over a long period of time, then a *symplectic numerical method* is a good candidate. Such methods replace the original Hamiltonian `ode` with a numerical scheme, that solves another approximating Hamiltonian `ode`.

Symplectic integration, which preserves volume, has similarities with *geometric integration*, which preserves the Lie group structure. See e.g., Jacob [26] and Sanz-Serna and Calvo [51] on the underlying mathematical and numerical theory. The connection to `sde` is through Milstein and Tretyakov [40].

8.1. Prerequisites: Symplectic transformations and Hamiltonian systems.

Symplectic transformation. Put simply, A smooth transformation $\mathbf{f_t} : \mathbb{R}^{\mathbf{n}} \to \mathbb{R}^{\mathbf{n}}$, for $t \geq 0$, is *symplectic*, if the Jacobian is time invariant.

Hamiltonian systems. A *Hamiltonian* multidimensional `ode` is given by

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i} \qquad \text{and} \qquad \frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}, \qquad i = 1, \ldots, d, \tag{8.1}$$

for some function $H = H(t, \mathbf{p}, \mathbf{q})$, the *Hamiltonian* of the `ode`. The *propagator* of the Hamiltonian `ode` is the transformation $\Phi : (\mathbf{p_0}, \mathbf{q_0}) \to (\mathbf{p}, \mathbf{q})$.

Example 8.1. A point mass m hanging in a weight-less wire, oscillates around its point of equilibrium. If q is the angular displacement of the wire, and p the angular momentum of the point mass, then the system has Hamiltonian

$$H(t, p, q) = \frac{p^2}{2m} + \frac{kq^2}{2} \qquad (d = 1). \tag{8.2}$$

The Hamiltonian `ode`, associated with the Hamiltonian (8.2) through (8.1), is

$$\frac{dp}{dt} = -kq \qquad \text{and} \qquad \frac{dq}{dt} = \frac{p}{m} \tag{8.3}$$

With $\omega = \sqrt{k/m}$, the solution of (8.3) can be expressed by the propagator

$$\begin{pmatrix} p_t \\ q_t \end{pmatrix} = \Phi_t \begin{pmatrix} p_0 \\ q_0 \end{pmatrix} \qquad \text{where} \qquad \Phi_t = \begin{pmatrix} \cos \omega t & -m\omega \sin \omega t \\ (m\omega)^{-1} \sin \omega t & \cos \omega t \end{pmatrix}.$$

Preservation of phase space area can be checked using differential 2-forms

$$dp \wedge dq = |\Phi_t| \, dp_0 \wedge dq_0 = (\cos^2 \omega t + \sin^2 \omega t) \, (dp_0 \wedge dq_0) = dp_0 \wedge dq_0.$$

Propagators of Hamiltonian `ode`:s coincide with symplectic transforms. This is the background for *symplectic numerical methods*, which are numerical methods that preserve the symplectic structure of Hamiltonian `ode`:s. So a *symplectic numerical method*, is an exact solution to a perturbed version of the original Hamiltonian `ode`. This idea is called *backward error analysis* in numerical linear algebra, see e.g., Hairer, Lubich and Wanner [17].

8.2. Symplectic Runge-Kutta methods. With an equidistant partition (2.1) of $[0, T]$, an m-stage *Runge-Kutta* scheme, to numerically solve the multidimensional `ode` (6.2) (see Butcher [8]), is given by

$$\mathbf{y}_{t_{k+1}} = \Psi_{t_{k+1}, t_k} \mathbf{y}_{t_k} = \mathbf{y}_{t_k} + \sum_{i=0}^{m} b_i h f(t_k + c_i h, \theta_i), \quad \theta_i = \mathbf{y}_{t_k} + \sum_{j=0}^{m} a_{ij} h f(t_k + c_j h, \theta_j). \tag{8.4}$$

It turns out that a Runge-Kutta method is symplectic, if its tableau satsifies $b_i a_{ij} + b_j a_{ji} - b_i b_j = 0$ and $c_i = \sum_{j=1}^{m} a_{ij}$ for $i, j \in \{1, \ldots, m\}$.

8.3. Symplectic methods for `sde`. We illustrate symplectic methods for `sde`, by means of two examples:

EXAMPLE 8.2. Consider the following (version of) a two dimensional `sde`:

$$\begin{aligned} dP &= f(t,P,Q)\,dt + \sum_{i=1}^{d} \sigma_i(t,P,Q)\,d(W_i)_t, \qquad P_0 = p_0, \\ dQ &= g(t,P,Q)\,dt + \sum_{i=1}^{d} \gamma_i(t,P,Q)\,d(W_i)_t, \qquad Q_0 = q_0. \end{aligned} \tag{8.5}$$

This `sde` is symplectic, if there exist functions $\{H_i(t,p,q)\}_{i=0}^{d}$ such that

$$f = -\frac{\partial H_0}{\partial q}, \quad g = \frac{\partial H_0}{\partial p} \qquad \text{and} \qquad \sigma_i = -\frac{\partial H_i}{\partial q}, \quad \gamma_i = \frac{\partial H_i}{\partial p}, \quad i = 1,\dots,d.$$

For the special case of (8.5), when the coefficients satisfy

$$\sigma_i(t,p,q) = \sigma_i(t), \quad \gamma_i(t,p,q) = \gamma_i(t), \quad f(t,p,q) = f(t,q), \quad g(t,p,q) = g(p).$$

Milstein, Repin and Tretyakov [39] propose an explicit symplectic method. They show that the method is of *mean-square order* 1. For $d = 1$, with an equidistant partition (2.1) of $[0,T]$, their sheme takes the following form:

$$\begin{aligned} P_{t_{k+1}} &= P_{t_k} + f\big(t_{k+1}, Q_{t_k} + hg(P_{t_k})\big)\,h + \sigma(t_k)\,(W_{t_{k+1}} - W_{t_k}) \\ Q_{t_{k+1}} &= Q_{t_k} + \qquad\quad g(P_{t_k})\,h \qquad\qquad + \gamma(t_k)\,(W_{t_{k+1}} - W_{t_k}) \end{aligned}$$

EXAMPLE 8.3. Consider the following two dimensional `sde`:

$$d\mathbb{X}_t = \mathbb{X}_t\,dt + \begin{pmatrix} \sigma & 0 \\ 0 & \gamma \end{pmatrix} d\mathbb{W}_t. \tag{8.6}$$

With an equidistant partition (2.1), the propagator for (8.6) is given by

$$\mathbb{X}_{t_{k+1}} = \begin{pmatrix} \cos h & \sin h \\ -\sin h & \cos h \end{pmatrix} \mathbb{X}_{t_k} + u_k, \qquad \mathbb{X}_0 = \mathbf{0},$$

where

$$u_k = \begin{pmatrix} \sigma \displaystyle\int_{t_k}^{t_{k+1}} \cos(t_{k+1}-s)\,dW_1(s) + \gamma \int_{t_k}^{t_{k+1}} \sin(t_{k+1}-s)\,dW_2(s) \\ -\sigma \displaystyle\int_{t_k}^{t_{k+1}} \sin(t_{k+1}-s)\,dW_1(s) + \gamma \int_{t_k}^{t_{k+1}} \cos(t_{k+1}-s)\,dW_2(s) \end{pmatrix}.$$

Milstein [38] proposes the following symplectic method of mean-square order 2:

$$\overline{\mathbb{X}}_{t_{k+1}} = \begin{pmatrix} \cos h & \sin h \\ -\sin h & \cos h \end{pmatrix} \overline{\mathbb{X}}_{t_k} + \frac{1}{h} \begin{pmatrix} \sigma \sin h & 2\gamma \sin^2 \frac{1}{2} h \\ -2\sigma \sin^2 \frac{1}{2} h & \sigma \sin h \end{pmatrix} (\mathbb{W}_{t_{k+1}} - \mathbb{W}_{t_k}). \tag{8.7}$$

For comparison, the following Euler scheme is mean-square order 1:

$$\overline{\mathbb{X}}_{t_{k+1}} = \begin{pmatrix} 1 & h \\ -h & 1 \end{pmatrix} \overline{\mathbb{X}}_{t_k} + \frac{1}{h} \begin{pmatrix} \sigma & 0 \\ 0 & \gamma \end{pmatrix} (\mathbb{W}_{t_{k+1}} - \mathbb{W}_{t_k}). \tag{8.8}$$

Figures 8.2 and 8.1 show that the symplectic method (8.7) is better at approximating the solution over a long time interval, than is the Euler approximation (8.8). Figures 8.3 and 8.4 show that the symplectic method preserves phase space area, which the Euler approximation does not.

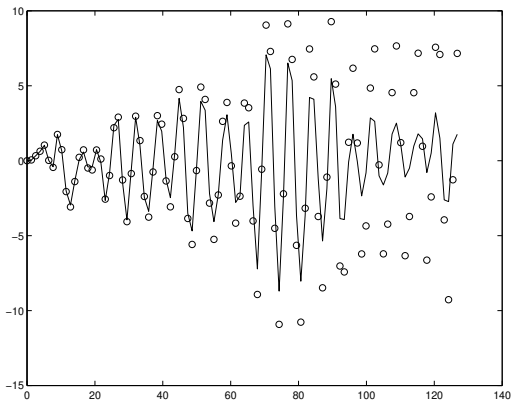

FIG. 8.1. *The Euler method(∘)(8.8) compared with the exact solution (solid), $\sigma = \gamma = 1$.*

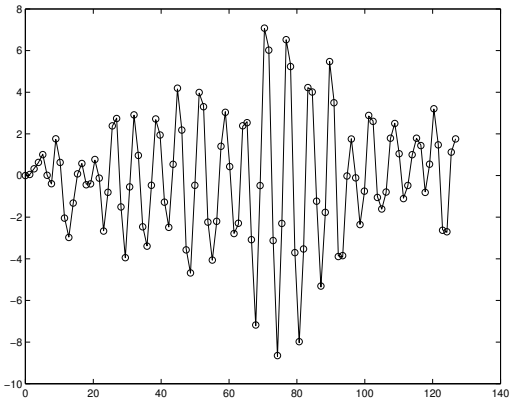

FIG. 8.2. *The symplectic method (8.7) (∘) compared with the exact solution (solid), $\sigma = \gamma = 1$.*

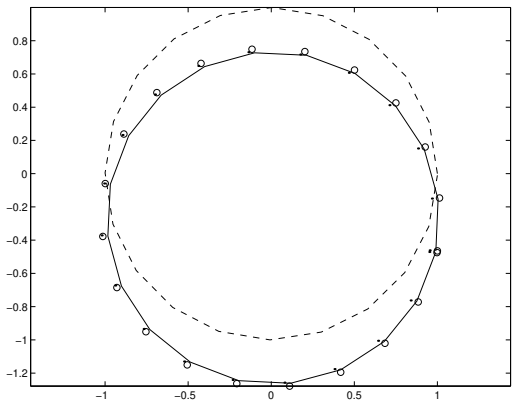

FIG. 8.3. *Phase space after 0.2 seconds, of symplectic approximation (dotted graph), Euler approximation (dotted circles), and exact solution (solid circle).*

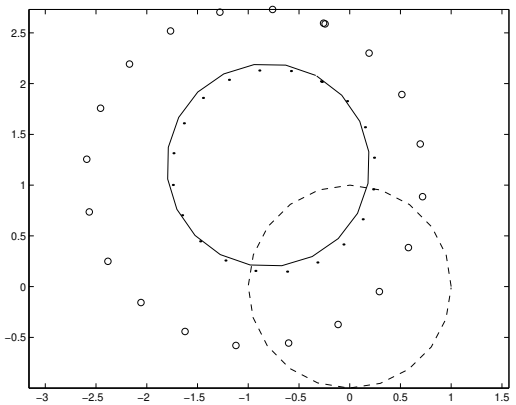

FIG. 8.4. *Phase space after 5 seconds, of symplectic approximation (dotted graph), Euler approximation (dotted circles), and exact solution (solid circle).*

9. Lie group methods. Lie group methods can be used to design numerical methods for `ode` evolving on differentiable manifolds, so that the numerical solution evolves on the same manifold, if that manifold is a Lie group.

EXAMPLE 9.1. Consider the two dimensional `ode`

$$\mathbf{y}'(t) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \mathbf{y}(t), \quad t \geq 0, \qquad \mathbf{y}(0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{9.1}$$

The exact solution of this system is $\mathbf{y}_t = (\sin t \quad \cos t)^T$ on the unit circle.

An explicit Euler method, using an equidistant partition (2.1) of $[0, T]$, might be what first comes to mind:

$$\mathbf{y}_{t_{k+1}} = \mathbf{y}_{t_k} + \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \mathbf{y}_{t_k} (t_{k+1} - t_k) = \begin{pmatrix} 1 & h \\ -h & 1 \end{pmatrix} \mathbf{y}_{t_k}, \quad \mathbf{y}_{t_0} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{9.2}$$

This method gives approximations that do not lie on the unit circle.

For $f(x) = x + o(x)$ as $x \to 0$, the following method has the same first order properties as (9.2), but evolves on the circle, as does the exact solution:

$$\mathbf{y}_{t_{k+1}} = \exp\left\{ f(t_{k+1} - t_k) \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \right\} \mathbf{y}_{t_k}, \qquad \mathbf{y}_{t_0} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

9.1. Prerequisites: Lie groups.

Matrix Lie group. A real *matrix Lie group* is a smooth subset of $\mathbb{R}_{d|d}$, that is closed under matrix products and matrix inversion.

Matrix Lie algebra. The *matrix Lie algebra* $\mathfrak{g}$, of a matrix Lie group $\mathcal{G}$, is the linear subspace of $\mathbb{R}_{d|d}$, consisting of all matrices of the form $A = \rho'(0)$, for $\rho(\cdot) \in \mathcal{G}$ a smooth curve with $\rho(0) = \mathbf{1}$. The *commutator* is the matrix commutator $[A, B] = AB - BA$.

Differential equation evolving on a matrix Lie group. If $\mathcal{G}$ is a matrix Lie group, and $\mathfrak{g}$ an associated matrix Lie algebra, then a differential equation evolving on $\mathcal{G}$ is an equation of the form

$$Y_t' = A(t, Y_t)Y_t, \quad t \in [0, T], \qquad Y_0 = y \in \mathcal{G}. \tag{9.3}$$

Here A is a map $A : \mathbb{R} \times \mathcal{G} \to \mathfrak{g}$.

For small $t \geq 0$, the solution of (9.3) is given by $Y_t = \exp(\Theta_t)Y_0$, where $\Theta_t \in \mathfrak{g}$ satisfies the differential equation

$$\Theta_t' = \operatorname{dexp}_{\Theta_t}^{-1} A(t, Y_t)), \qquad \Theta_0 = \mathbf{0}. \tag{9.4}$$

The operators $\exp(\cdot)$, $\operatorname{dexp}_{(\cdot)}(\cdot)$ and $\operatorname{dexp}_{(\cdot)}^{-1}(\cdot)$ are defined below.

The *exponential map*, $\exp : \mathfrak{g} \to \mathcal{G}$, is defined by

$$\exp(A) = \sum_{k=0}^{\infty} \frac{A^k}{k!}.$$

The *adjoints* $\operatorname{ad}_A^{(k)}(B)$ are defined by

$$\operatorname{ad}_A^{(k+1)}(B) = [A, \operatorname{ad}_A^{(k)}(B)] \qquad \text{where} \qquad \operatorname{ad}_A^{(0)}(B) = B.$$

The *differential* $\operatorname{dexp} : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$, of the exponential exp is defined by

$$\operatorname{dexp}_A(B) = \frac{\exp(\operatorname{ad}_A(B)) - \mathbf{1}}{\operatorname{ad}_A(B)} = \sum_{k=0}^{\infty} \frac{1}{(k+1)!} \operatorname{ad}_A^{(k)}(B).$$

One motivation for this definition, is the following important formula:

$$\frac{d}{dt} \exp(A_t) = \operatorname{dexp}_{A_t}(A_t') \exp(A_t).$$

The map dexp can be inverted, by Taylor expanding $x/(\mathrm{e}^x - 1)$,

$$\frac{x}{\mathrm{e}^x - 1} = \sum_{k=0}^{\infty} \frac{B_k}{k!} x^k,$$

where $\{B_k\}_{k=0}^{\infty}$ are the so called *Bernoulli numbers*

$$\operatorname{dexp}_A^{-1}(B) = \sum_{k=0}^{\infty} \frac{B_k}{k!} \operatorname{ad}_A^{(k)}(B) = B - \frac{1}{2}[A, B] + \frac{1}{12}[A, [A, B]] + \ldots. \tag{9.5}$$

A numerical solution of (9.4), can be had by a Runge-Kutta type scheme:

9.2. The Runge-Kutta-Munthe-Kaas scheme. When (8.4) is applied to the equation (9.4), the resulting numerical scheme is called the *Runge-Kutta-Munthe-Kaas* (RKMK) scheme:

$$\begin{aligned} \mathbf{y}_{t_{n+1}} &= \exp\left(\sum_{i=0}^{m} b_i \operatorname{dexp}_{\Theta_i}^{-1}\left(hA(t_n + c_i h, \exp(\Theta_i)\mathbf{y}_{t_n})\right)\right)\mathbf{y}_{t_n} \\ \Theta_i &= \sum_{j=0}^{m} a_{ij} \operatorname{dexp}_{\Theta_j}^{-1}\left(hA(t_n + c_j h, \exp(\Theta_j)\mathbf{y}_{t_n})\right) \end{aligned} \tag{9.6}$$

The application of the scheme (9.6) requires an approximation for the operator dexp^{-1}. One way to get this is to truncate the expansion (9.5). If the RKMK scheme is to be of the same order as the underlying Runge-Kutta scheme, the truncation has to be of that order. For high-order methods, this implies the use of several commutators from (9.5), giving complex numerical methods, see Crouch and Grossman [10]. A way to reduce the number of commutators is to use so called *free Lie algebras*. For this and more, see the survey article on Lie group methods by Iserles, Munthe-Kaas, Nøsett and Zanna [24].

An important fact is that, even when dexp^{-1} is approximated, the resulting numerical approximation, obtained from (9.6), will stay in the same Lie group as the exact solution. This is the main idea with Lie group methods.

9.3. The Magnus formula. Previously, we saw that the `ode`

$$\mathbf{y}_t' = A_t \mathbf{y}_t, \quad \mathbf{y}_0 \in \mathcal{G},$$

has solution

$$\mathbf{y}_t = \exp(\Theta_t)\mathbf{y}_0 \qquad \text{where} \qquad \Theta_t' = \operatorname{dexp}_{\Theta}^{-1}(A_t). \tag{9.7}$$

One way to approach the solution to (9.7), is by *Picard iteration*:

$$\Theta_t^{(k)} = \int_0^t \operatorname{dexp}_{\Theta_s^{(k-1)}}^{-1}(A_s)\, ds, \qquad \Theta_t^{(0)} = 0.$$

In a suitably small neighbourhood of 0, this scheme converges $\lim_{k\to\infty} \Theta_t^{(k)} = \Theta_t$.

The solution Θ to (9.7) can be represented by a so called *Magnus expansion*, or the *Magnus formula*. This is a sum of iterated integrals and commutators

$$\Theta_t = \sum_{k=0}^{\infty} H_t^{(k)}, \tag{9.8}$$

where

$$\begin{aligned} H_t^{(0)} &= \int_0^t A_{s_1}\, ds_1 \\ H_t^{(1)} &= -\frac{1}{2}\int_0^t \left[\int_0^{s_1} A_{s_2}\, ds_2, A_{s_1}\right] ds_1 \\ H_t^{(2)} &= \frac{1}{12}\int_0^t \left[\int_0^{s_1} A_{s_2}\, ds_2, \left[\int_0^{s_1} A_{s_2}\, ds_2, A_{s_1}\right]\right] ds_1 \\ &\quad + \frac{1}{4}\int_0^t \left[\int_0^{s_1}\left[\int_0^{s_2} A_{s_3}\, ds_3, A_{s_2}\right] ds_2, A_{s_1}\right] ds_1 \\ &\ \ \vdots \end{aligned} \tag{9.9}$$

By Moan [42], if the Lie algebra g is equipped with a norm, $\|\cdot\|$, the Magnus expansion (9.8) converges, in this norm, for $t \geq 0$ such that

$$\int_0^t \|A_s\|\, ds \leq \int_0^{2\pi} \frac{ds}{4 + s[1 - \cot(s/2)]} \approx 1.089.$$

A more transparent form of the expansion can be obtained by so called *rooted binary trees*, from *graph theory*, making numerical calculations and recursive computation of consecutive terms in the expansion more tractable. We do not go into this here, but refer to Iserles and Nørsett [25] together with Butcher [8].

The Magnus expansion (9.8) is used to get approximative solutions to (9.4) by truncation. Truncations cannot be done anywhere in the expansion: The rooted tree approach gives information on where and on the order of the resulting approximations.

EXAMPLE 9.2. The first two possible truncations satisfy

$$\Theta_t \approx \int_0^t A_{s_1}\, ds_1 \qquad \text{Order 2}$$

$$\Theta_t \approx \int_0^t A_{s_1}\, ds_1 - \frac{1}{2}\int_0^t \left[\int_0^{s_1} A_{s_2}\, ds_2, A_{s_1}\right] ds_1 \qquad \text{Order 4}$$

After truncation the integrals can be approximated by so called *numerical quadrature*, which is how far we go here. For more on this, see e.g., Shampine [54].

9.4. The Magnus formula applied to `sde`. Burrage and Burrage [5] apply the Magnus formula to a Stratonvich `sde`. This seems to be the first time the Magnus formula is used in a stochastic context.

EXAMPLE 9.3. Burrage and Burrage consider a two dimensional `sde`

$$d\mathbb{X}_t = G_0\mathbb{X}_t\, dt + G_1\mathbb{X}_t \circ d(W_1)_t + G_2\mathbb{X}_t \circ d(W_2)_t, \quad t \in [0,1], \quad \mathbb{X}_0 = \begin{pmatrix}1\\1\end{pmatrix}, \quad (9.10)$$

where

$$G_0 = \begin{pmatrix} -\frac{9}{10} & 0 \\ \frac{1}{4} & -\frac{1}{2}\end{pmatrix}, \qquad G_1 = \begin{pmatrix}\frac{3}{4} & 0 \\ 0 & -\frac{3}{4}\end{pmatrix} \qquad \text{and} \qquad G_2 = \begin{pmatrix} 0 & \frac{9}{10} \\ \frac{9}{10} & 0\end{pmatrix}.$$

Since G_0, G_1 and G_2 do not commute, the Magnus expansion is needed for good numerical approximations of (9.10): The solution to (9.10) is

$$\mathbb{X}_t = \exp(\Theta_t)\mathbb{X}_0 = \exp(\Theta_t),$$

with Θ given by (9.8) and (9.9). Further,

$$A_t\, dt = G_0\, dt + G_1 \circ d(W_1)_t + G_2 \circ d(W_2)_t.$$

Hence the first term in the Magnus expansion (9.9) is

$$H_t^{(0)} = \int_0^t A_s\, ds = G_0 t + G_1(W_1)_t J_1 + G_2(W_2)_t.$$

The second term in the expansion is calculated with somewhat more effort:

$$H_t^{(1)} = [G_0, G_1](J_{10} - J_{01}) + [G_0, G_2](J_{20} - J_{02}) + [G_1, G_2](J_{21} - J_{12}),$$

where $[G_i, G_j] = G_i G_j - G_j G_i$ and

$$J_{ij} = \int_0^t \int_0^{s_1} dW_{s_2}^{(i)} \circ dW_{s_1}^{(j)}, \qquad dW_t^{(i)} = \begin{cases} dt \quad , & i = 0 \\ d(W_1)_t, & i = 1 \\ d(W_2)_t, & i = 2 \end{cases}$$

Using only $H^{(0)}$ and $H^{(1)}$, Burrage and Burrage [5] construct a numerical scheme based on the Magnus expansion, and compare it to the Euler scheme. The results are depicted in Figure 9.1 and Figure 9.2.

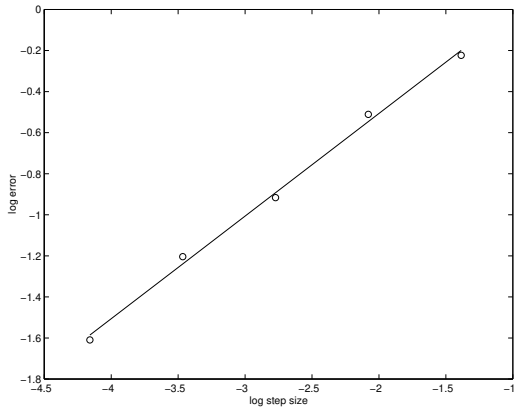

FIG. 9.1. *The logarithm of the global error versus that of the step size of the Euler scheme (o) follows a line with slope 0.5, suggesting that it is strong order 0.5.*

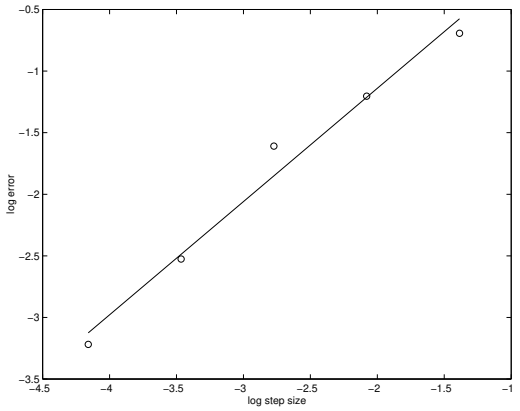

FIG. 9.2. *The logarithm of the global error versus that of the step size of the Magnus expansion (o) follows a 45^0 line, suggesting that it is strong order 1.*

The iterated Stratonovich integrals, that feature in Example 9.3, has to be simulated. One way to do this is described by Ryden and Wiktorsson [50].

The integrals appearing in (9.9) are interrelated, see Kloeden and Platen [29] for a discussion. Simulations of high order iterated integrals are involved, but can be made more tractable, using these interrelations. See Milstein [38].

9.5. The Kunita formula. Consider a Stratonovich `sde`, driven by a one dimensional Wiener-process:

$$d\mathbb{S}_t = a(\mathbb{S}_t)\, dt + b(\mathbb{S}_t) \circ dW_t, \quad t \geq 0, \qquad \mathbb{S}_0 = \mathbf{s}, \tag{9.11}$$

where the solution $\mathbb{S}$ evolves in a d-dimensional connected C^∞-manifold $\mathcal{M}$:

$$\mathbb{S} : [0, \infty) \to \mathcal{M} \qquad \text{and} \qquad a, b \in C^\infty(\mathcal{M}, \mathbb{R}^d).$$

Introduce the operators P_0 and P_1:

$$P_0 = \sum_{i=1}^{d} a_i \frac{\partial}{\partial s_i} \qquad \text{and} \qquad P_1 = \sum_{i=1}^{d} b_i \frac{\partial}{\partial s_i}.$$

A *multi-index* $J = (j_1, \ldots, j_m)$, of *length* $|J| = j_1 + \ldots + j_m$, is divided into shorter multi-indices, $J = (J_1, \ldots, J_q)$, such that each J_k contains the same element. Picking $k_1 < k_2 < \ldots < k_l = q$, a further sub-division

$$\Delta J = (J_1, \ldots, J_{k_1}, J_{k_1+1}, \ldots, J_{k_2}, \ldots, J_{k_{l-1}+1}, \ldots, J_{k_l}),$$

such that each J_{k_i} contains at most two elements, is a *single divided* multi-index of J, if $|J_{k_i}| = 1$ for all i, and otherwise a *double divided* multi-index of J.

If the Lie algebra $L(P_0, P_1)$ generated by the operators P_0 and P_1 is p-nilpotent, then, by a result of Kunita (see [41]), the solution of the `sde` (9.11) is given by

$$\mathbb{S}_t = \mathrm{e}^{Q_t}\mathbb{S}_0 = \mathrm{e}^{Q_t}\mathbf{s}, \quad Q_t = \sum_{i=0}^{1} W_t^{(i)} P_i + \sum_{J:2\le|J|\le p} \sum_{\Delta J}^{*} c_{\Delta J} W_t^{(\Delta J)} P_J,$$

where $\sum_{\Delta J}^{*}$ is the sum over all single and double divided multi-indices of J, and

$$P_J = [[\cdots[P_{j_1}, P_{j_2}], \ldots], P_{j_m}]$$
$$W_t^{(\Delta J)} = \int_0^t \int_0^{t_1} \cdots \int_0^{t_{l-1}} dW_{t_l}^{(J_{k_l})} \circ \ldots \circ dW_{t_1}^{(J_1)}$$
$$W_t^{(J_k)} = \begin{cases} W_t^{(i_k)} & J_k = \{i_k\} \\ t & J_k = \{i_k, i_k\} \\ 0 & J_k = \{0, 0\} \end{cases}$$

The coefficients $c_{\Delta J}$ are computed from a complicated formula, that is not given here.

If $L(P_0, P_1)$ is not p-nilpotent, for some p, then Q_t is given by an infinite sum

$$Q_t = \sum_{i=0}^{1} W_t^{(i)} P_i + \sum_{J:2\le|J|} \sum_{\Delta J} c_{\Delta J} W_t^{(\Delta J)} P_J,$$

provided that it converges. (Note the absence of p in the second summation.)

For $L(P_0, P_1)$ finite dimensional, Arous [2] showed that there exists a stopping time τ, with respect to the filtration $\mathcal{G}_t = \sigma(\mathbb{S}_s : s \in [0,1])$, such that

$$\mathbf{P}\left\{\forall t \le \tau : \; Q_t = \sum_{i=0}^{1} W_t^{(i)} P_i + \sum_{J:2\le|J|} \sum_{\Delta J} c_{\Delta J} W_t^{(\Delta J)} P_J\right\} = 1.$$

Kunita's formula implies that the solution, $\mathbb{S}_t = \mathrm{e}^{Q_t}\mathbf{s}$, to (9.11) can be expressed $\mathbb{S}_t = \phi(1)$, where $\phi : [0,1] \to \mathcal{M}$ satisfies the following `ode`:

$$\frac{d}{dr}\phi(r) = Q_t\phi(r), \qquad \phi(0) = \mathbf{s}.$$

9.6. Applying the Kunita formula. To calculate an approximative solution, consider an equidistant partition (2.1) of $[0,T]$. The numerical approximation $\overline{\mathbb{S}}$, of the solution $\mathbb{S}$ to (9.11), is

$$\overline{\mathbb{S}}_{t_{k+1}} = \mathrm{e}^{\overline{Q}_{t_{k+1},t_k}}\overline{\mathbb{S}}_{t_k}, \qquad \overline{\mathbb{S}}_{t_0} = \mathbf{s}.$$

where $\overline{Q}$ is an approximation (truncation) of Q, for example,

$$\overline{Q}_{t_{k+1},t_k} = hP_0 + (W_{t_{k+1}} - W_{t_k})P_1. \tag{9.12}$$

The Baker-Campbell-Hausdorff formula states the following relation:

$$\mathrm{e}^A\mathrm{e}^B = \exp\left\{A + B + \frac{1}{2}[A,B] + \frac{1}{12}([A,[A,B]] + [B,[B,A]] + \ldots)\right\}.$$

The exponential e^{A+B} can be approximated by $\mathrm{e}^{A}\mathrm{e}^{B}$, according to this formula. This approximation is exact when $[A,B]=0$.

In conclusion, one numerical scheme for approximating the solution $\mathbb{S}$, reads:

$$\overline{\mathbb{S}}_{t_{k+1}} = \mathrm{e}^{hP_0}\mathrm{e}^{(W_{t_{k+1}}-W_{t_k})P_1}\overline{\mathbb{S}}_{t_k}, \qquad \overline{\mathbb{S}}_{t_0}=\mathbf{s}.$$

The following example illustrates the use of Kunita's formula:

EXAMPLE 9.4. Consider the one dimensional Bessel-type Stratonovich `sde`

$$dS_t = S_t\,dt + 2\sqrt{S_t}\circ dW_t, \quad t\geq 0, \qquad S_0=s. \tag{9.13}$$

Here the operators P_0 and P_1 are given by

$$P_0 = s\frac{d}{ds} \qquad \text{and} \qquad P_1 = 2\sqrt{s}\frac{d}{ds}.$$

The Lie algebra $L(P_0,P_1)$ is finite dimensional, since

$$[P_0,P_1] = P_0P_1 - P_1P_0 = s\frac{d}{ds}(2\sqrt{s}\frac{d}{ds}) - 2\sqrt{s}\frac{d}{ds}(s\frac{d}{ds}) = -\sqrt{s}\frac{d}{ds} = -\frac{P_1}{2}.$$

With the truncation (9.12), e^{hP_0} can be computed in the following way:

$$z_h = \mathrm{e}^{hP_0}z_0 = \phi_1,$$

where the function ϕ satisfies the `ode`

$$\phi' = hP_0\phi = h\phi\frac{d\phi}{d\phi} = h\phi, \quad t\in[0,1], \qquad \phi_0 = z_0.$$

The solution is $\phi_\tau = z_0\mathrm{e}^{\tau h}$, which gives $\mathrm{e}^{hP_0}s = s\mathrm{e}^{h}$.

As for $\mathrm{e}^{(W_{t_{k+1}}-W_{t_k})P_1}$, it can be computed in the following way:

$$z_t = \mathrm{e}^{(W_{t_{k+1}}-W_{t_k})P_1}z_0 = \phi_1,$$

where the function ϕ satisfies the `ode`

$$\phi' = (W_{t_{k+1}}-W_{t_k})P_1\phi = (W_{t_{k+1}}-W_{t_k})2\sqrt{\phi}\frac{d\phi}{d\phi} \quad t\in[0,1], \qquad \phi_0=z_0.$$

This gives $\phi_\tau = \left(\tau(W_{t_{k+1}}-W_{t_k})+\sqrt{z_0}\right)^2$, so that

$$\mathrm{e}^{(W_{t_{k+1}}-W_{t_k})P_1}s = \left(\tau(W_{t_{k+1}}-W_{t_k})+\sqrt{s}\right)^2.$$

We get the following scheme, which is of mean-square order 1, by Misawa [41]:

$$\overline{S}_{t_{k+1}} = \mathrm{e}^{h}\left((W_{t_{k+1}}-W_{t_k})+\sqrt{\overline{S}_{t_k}}\right)^2, \qquad \overline{S}_{t_0}=s. \tag{9.14}$$

The solution S to the `sde` (9.13), can be shown to be non-negative. The numerical scheme (9.14) shares this property. This is not the case for e.g., the Euler approximation, which is of mean-square order 0.5:

$$\overline{S}_{t_{k+1}} = \overline{S}_{t_k}h + 2\sqrt{\overline{S}_{t_k}}(W_{t_{k+1}}-W_{t_k}). \tag{9.15}$$

Since the exact solution of (9.13) cannot be obtained in closed form, to compare the numerical solution based on the Kunita formula with that by the Euler method, a high order approximation, based on a stochastic Taylor expansion, is used as a substitute for the exact solution. This approximation is of mean-square order 1.5, and was suggested by Kloeden and Platen[29]:

$$\overline{S}_{t_{k+1}} = \overline{S}_{t_k} + (\overline{S}_{t_k}+1)\,h + 2\sqrt{\overline{S}_{t_k}}\,(W_{t_{k+1}} - W_{t_k}) + \big((W_{t_{k+1}} - W_{t_k})^2 - h\big) + 2\sqrt{\overline{S}_{t_k}} \int_{t_k}^{t_{k+1}} \int_{t_k}^{r_1} dW_{r_2} dr_1 + \sqrt{\overline{S}_{t_k}} \int_{t_k}^{t_{k+1}} \int_{t_k}^{r_1} dr_2\, dW_{r_1} + \frac{1}{2}(\overline{S}_{t_k}+1)h^2. \tag{9.16}$$

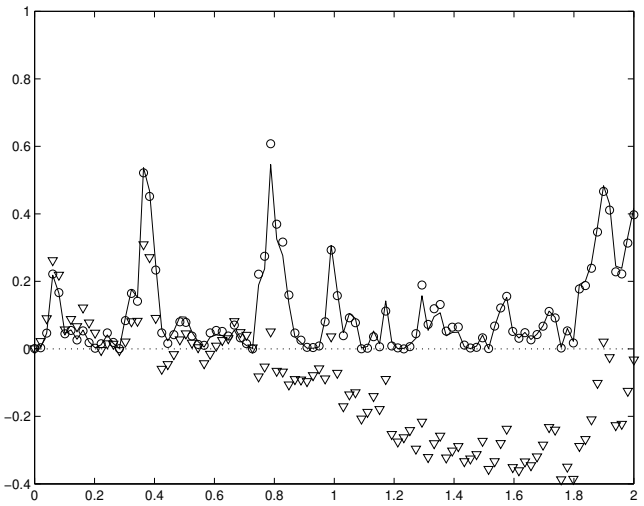

FIG. 9.3. *The dotted graph (o) depicts an approximative solution calculated by the scheme* (9.14)*, the solid graph (-) an approximation by the scheme* (9.16)*, and (▽) the Euler approximation* (9.15)*, of the* `sde` (9.13)*, with s=0.01.*

REFERENCES

[1] C. T. H. Baker and E. Buckwar, *Continuous Θ-methods for the stochastic panthograph equation*, Electron. Trans. Numer. Anal., 11 (2000), pp. 131–151.

[2] G. Ben-Arous, *Flots et series de Taylor stochastiques* (in french), Probab. Theory Related Fields, 81 (1989), pp. 29–77.

[3] A. Bryden and D. J. Higham, *On the boundedness of asymptotic stability regions for the stochastic theta-method*, University of Strathclyde Mathematics Research Report, 2002.

[4] E. Buckwar, *Introduction to the numerical analysis of stochastic delay differential equations*, J. Comput. Appl. Math., 125 (2000), pp. 297–307.

[5] K. Burrage and P. M. Burrage, *High strong order methods for non-commutative stochastic ordinary differential equation systems and the Magnus formula*, Phys. D, 133 (1999), pp. 34–48.

[6] K. Burrage, P. M. Burrage and T. Tian, *Numerical methods for strong solutions of stochastic differential equations: and overview*, Proc. Roy. Soc. London Ser. A, 460 (2004), pp. 373–402.

[7] P. M. Burrage, R. Herdiana and K. Burrage, *Adaptive stepsize based on control theory for stochastic differential equations*, J. Comput. Appl. Math., 170 (2004), pp. 317–336.

[8] J. C. Butcher, *The Numerical Analysis of Ordinary Differential Equations*, Wiley, New York, 1987.

[9] J. M. C. Clark and R. J. Cameron, *The maximum rate of convergence of discrete approximations for stochastic differential equations*, in Stochastic Differential Systems, B. Grigelionis, ed., Springer, Berlin, 1980, pp. 162–171.

[10] P. E. Crouch and R. Grossman, *Numerical integration of ordinary differential equations on manifolds*, J. Nonlinear Sci., 3 (1993), pp. 1–33.

[11] K. Eriksson, C. Johnson, and A. Logg, *Explicit time-stepping for stiff ODEs*, SIAM J. Sci. Comput., 25 (2003/04), pp. 1142-1157.

[12] J. G. Gaines and T. J. Lyons, *Random generation of stochastic area integrals*, SIAM J. Appl. Math., 54 (1994), pp. 1132–1146.

[13] J. G. Gaines and T. J. Lyons, *Variable step size control in the numerical solution of stochastic differential equations*, SIAM J. Appl. Math., 57 (1997), pp. 1455–1484.

[14] C. W. Gear and I. G. Kevrekidis, *Projective methods for stiff differential equations: Problems with gaps in their eigenvalue spectrum*, SIAM J. Sci. Comput., 24 (2003), pp. 1091–1106.

[15] C. W. Gear and I. G. Kevrekidis, *Telescopic projective methods for parabolic differential equations*, J. Comput. Phys., 187 (2003), pp. 95–109.

[16] I. Gyöngy, *A note on Euler approximations*, Potential Anal., 8 (1998), pp. 205–216.

[17] E. Hairer, C. Lubich and G. Wanner, *Geometric numerical integration. Structure-preserving algorithms for ordinary differential equations*, Springer, Berlin, 2002.

[18] D. Higham, *Mean-square stability and asymtotic stability of the stochastic theta method*, SIAM J. Numer. Anal., 38 (2000), pp. 753–769.

[19] D. J. Higham, X. Mao and A. Stuart, *Strong convergence of numerical methods for nonlinear stochastic differential equations*, SIAM J. Numer. Anal., 40 (2002) pp. 1041–1063.

[20] N. Hofmann, T. Müller-Grönbach and K. Ritter, *Step size control for the uniform approximation of systems of stochastic differential equations with additive noise*, Ann. Appl. Probab., 10 (2000), pp. 616–633.

[21] N. Hofmann, T. Müller-Grönbach, and K. Ritter, *The optimal discretization of stochastic differential equations*, J. Complexity, 17 (2001), pp. 117–153.

[22] Y. Hu, S-E. A. Mohammed and F. Yan, *Discrete-time approximations of stochastic delay equations: The Milstein scheme*, Ann. Probab., 32 (2004), pp. 265–314.

[23] N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Proceses*, North Holland, 1981.

[24] A. Iserles, H. Z. Munthe-Kaas, S. P. Nøsett, and A. Zanna, *Lie-group methods*, Acta Numer., 9 (2000), pp. 215–365.

[25] A. Iserles and S. P. Nørsett, *On the solution of linear differential equations in Lie groups*, R. Soc. Lond. Philos. Trans. Ser. A Math. Phys. Eng. Sci., 357 (1999), pp. 983–1020.

[26] B. Jacob, *Linear Algebra*, Freeman, New York, 1990.

[27] J. Janssen and S. Vandewalle, *Convolution-based Chebyshev acceleration of waveform relaxation methods*, Technical report, Katholieke Universiteit Leuven, Dept. Computerwetenschappen, 2001.

[28] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, Springer, Berlin, 1991.

[29] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer, Berlin, 1992.

[30] U. Küchler and E. Platen, *Strong discrete time approximation of stochastic differential equations with time delay*, Math. Comput. Simulation, 54, (2000), pp. 189–205.

[31] H. Lamba, *An adaptive timestepping algorithm for stochastic differential equations*, J. Comput. Appl. Math., 161 (2003), pp. 417–430.

[32] J. Lehn , A. Rössler and O. Schein, *Adaptive schemes for the numerical solution of SDE:s - a comparison*, J. Comput. Appl. Math., 138 (2002), pp. 297–308.

[33] B. Leimkuhler, *Timestep acceleration of waveform relaxation*, SIAM J. Numer. Anal., 35 (1998), pp. 31–50.

[34] E. Lelarasmee, *The waveform relaxation method for the time domain analysis of large scale non-linear systems*, PhD thesis, University of California, Berkeley, 1982.

[35] X. Mao and S. Sabanis, *Numerical solution of stochastic differential delay equations under local Lipshitz condition*, J. Comput. Appl. Math., 151 (2003), pp. 215–227.

[36] G. Maruyama, *Continuous Markov processes and stochastic equations*, Rend. Circ. Mat. Palermo, 4 (1955), pp. 48–90.

[37] S. Mauthner, *Step size control in the numerical solution of stochastic differential equations*, J. Comput. Appl. Math., 100 (1998), pp. 93-109.

[38] G. A. Milstein, *Numerical Integration of Stochastic Differential Equations*, Cambridge, Cambridge, 1990.

[39] G. N. Milstein, Yu. M. Repin, and M. V. Tretyakov, *Symplectic integration of Hamiltonian systems with additive noise*, SIAM J. Numer. Anal., 39 (2002), pp. 2066–2088.
[40] G. N. Milstein and M. V. Tretyakov, *Quasi-symplectic methods for Langevin-type equations*, IMA J. Numer. Anal., 23 (2003), pp. 593–626.
[41] T. Misawa, *Lie algebraic approach to numerical integration of stochastic differential equations*, SIAM J. Sci. Comput., 23 (2001), pp. 866–890.
[42] P. C. Moan, *Efficient approximation of Sturm-Liouville problems using Lie-group methods*, Technical Report NA11, DAMTP, University of Cambridge, 1998.
[43] N. J. Newton, *An efficient approximation for stochastic differential equations on the partition of symmetrical first passage times*, Stoch. Stoch. Rep., 29 (1990), pp. 227–258.
[44] N. J. Newton, *Variance reduction for simulated diffusions*, SIAM J. Appl. Math, 54 (1994), pp. 1780–1805.
[45] D. Nualart, *The Malliavin Calculus and Related Topics*, Springer, Berlin, 1995.
[46] D. Nualart and E. Pardoux, *Stochastic calculus with anticipating integrands*, Prob. Th. Rel. Fields, 78 (1988), pp. 535–581.
[47] R. Pettersson, *Stratonovich-Taylor expansion and numerical methods*, Stochastic Anal. Appl. 10 (1992), pp. 603–612.
[48] E. Platen, *Approximaton of Itô integral equations*, in Stochastic Differential Systems, B. Grigelionis, ed., Springer, Berlin, 1980, pp. 172–176.
[49] E. Platen, *An introduction to numerical methods for stochastic differential equations*, Acta Numer., 8 (1999), pp. 197–246.
[50] T. Ryden and M. Wiktorsson, *On the simulation of iterated Itô integrals*, Stochastic Process. Appl., 91 (2001), pp. 151–168.
[51] J. M. Sanz-Serna and M. P. Calvo, *Numerical Hamiltonian Problems*, Chapman-Hall, London, 1994.
[52] K. R. Schneider and H. Schurz, *Waveform relaxation methods for stochastic differential equations*, Preprint, Weierstrass-Institut für Angewandte Analysis und Stochastik, 2000.
[53] G. I. Schuëller, ed., *A state-of-the-art report on computational stochastic mechanics*, Probabilistic Engineering Mechanics, 12 (1997) pp. 197–321.
[54] L. F. Shampine, *Numerical Solution of Ordinary Differential Equations*, Chapman-Hall, London, 1994.
[55] S. Vanderwalle, *Parallel Multigrid Waveform Relaxation for Parabolic Problems*, Teubner, berlin, 1993.
[56] E. Wong and M. Zakai, *On the convergence of ordinary integrals to stochastic integrals*, Ann. Math. Statist., 36 (1965), pp. 1560–1564.
[57] E. Wong and M. Zakai, *On the relation between ordinary and stochastic differential equations*, Int. J. Engrg. Sci., 3 (1965), pp. 213–229.

PAPER II

Applied Probability Trust (6th March 2005)

ON VOLATITILTY INDUCED STATIONARITY FOR STOCHASTIC DIFFERENTIAL EQUATIONS

ANDERS MUSZTA,* *Chalmers University of Technology*

MARTIN RICHTER,** *Copenhagen Business School*

J.M.P. ALBIN,*** *Chalmers University of Technology*

BJARNE ASTRUP JENSEN,**** *Copenhagen Business School*

Abstract

This article deals with stochastic differential equations with volatility induced stationarity. We study of theoretical properties of such equations, as well as numerical aspects, together with a detailed study of three examples.

Keywords: CIR model; CKLS model; heavy-tailed SDE; hyperbolic SDE; local martingale; mean reversion; numerical methods; stochastic differential equation; time changed SDE; volatility induced stationarity

2000 Mathematics Subject Classification: Primary 60G44; 60H10; 60H35; 60J60; 65C30; 68U20. Secondary 60J65; 62P05; 65C40

1. Introduction

This article deals with stochastic differential equations (SDE) with volatility induced stationarity (*vis*). Although the *vis* effect is more or less well-known, misinterpretations

* Postal address: Mathematical Sciences, Chalmers University of Technology, 412 96 Gothenburg, Sweden. Email: muszta@math.chalmers.se

** Postal address: Department of Finance, Copenhagen Business School, 2000 Frederiksberg, Denmark. Email: mr.fi@cbs.dk

*** Postal address: Mathematical Sciences, Chalmers University of Technology, 412 96 Gothenburg, Sweden. Email and www: palbin@math.chalmers.se and http://www.math.chalmers.se/~palbin

**** Postal address: Department of Finance, Copenhagen Business School, 2000 Frederiksberg, Denmark. Email: ba.fi@cbs.dk

in the literature are common. We clerify matters by giving a general definition of *vis*, and explain why SDE with *vis* have a local martingale term that is not a martingale. We investigate the relation between *vis* and stationary moments and mean reversion.

SDE with *vis* feature to model, for example, interest rates and eletricity prices. But, such SDE can be difficult from a statistical point of view, so that basic estimation procedures fail. Hence computer simulations to evaluate statistics are important. As standard simulation schemes can fail for SDE with *vis*, we discuss alternative schemes.

We provide three examples of SDE with *vis*, that all are different in terms of their *vis* effect, and that require different simulation procedures. The main example is the CKLS model, which has been studied by many authors, but where we find some new features. The others are a class of hyperbolic SDE, and a class of heavy-tailed SDE.

This article is organized as follows: After the preliminary Section 2, we define *vis* in Section 3, and adress connections to boundaries for the diffusion and existence of moments. Section 4 deals with theoretical aspects of computer simulation of SDE with *vis*, together with properties of the local martingale term and mean reversion. In Section 5-7 we study the three examples that have been mentioned, of SDE with *vis*.

2. Preliminaries

In this section we set some notation that is required in later sections.

Let $\{W_t\}_{t\geq 0}$ be standard Brownian motion and $I=(l,r)$, $-\infty\leq l<r\leq\infty$ an open interval. For measurable functions $b: I\to\mathbb{R}$ and $\sigma: I\to[0,\infty)$, consider the SDE

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dW_t, \quad X_0=\zeta, \tag{2.1}$$

where ζ is a random variable with values in I that is independent of W. The law of X is denoted $\mathbf{P}^\zeta$, while the probability distribution of X_t is denoted $P_t(\zeta,\cdot)$. Further, π denotes the stationary distribution for X, when it exists.

A property holds *locally on* I if the property is true on all compact subsets of I.

Assumption 2.1. *The* drift b *is continuous and the* volatility σ *is (strictly) positive and locally Hölder continuous of order* $\frac{1}{2}$.

Given an $x_0\in I$, the *scale* function S and *speed* measure m are given by

$$S(x) = \int_{x_0}^{x} \exp\left\{-2\int_{x_0}^{y} \frac{b(z)}{\sigma(z)^2}\,dz\right\} dy \quad \text{and} \quad \frac{dm(x)}{dx} = \frac{2}{\sigma(x)^2 S'(x)} \quad \text{for } x \in I.$$

Feller's test for non-explosion requires the following function v to go to infinity

$$v(x) = \int_{x_0}^{x} S'(y)\left(\int_{x_0}^{y} dm(z)\right) dy = \int_{x_0}^{x} (S(x) - S(y))\, dm(y) \quad \text{for } x \in I\text{:}$$

Assumption 2.2. *We have* $v(l^+) = \lim_{x \downarrow l} v(x) = \infty$ *and* $v(r^-) = \lim_{x \uparrow r} v(x) = \infty$.

Assumptions 2.1 and 2.2 are sufficient for the existence of a unique strong solution of (2.1) that does not explode (see e.g., Karatzas and Shreve, 1991, Section 5.5).

Assumption 2.3. *We have* $S(l^+) = -\infty$, $S(r^-) = \infty$ *and* $m(I) < \infty$.

If we strengthen Assumption 2.2 to Assumption 2.3, the stationary solution to (2.1), with $\zeta \overset{\mathrm{d}}{=} \pi = m/m(I)$, is *ergodic* (see e.g., Rogers and Williams, 1987, p. 300), i.e.,

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t f(X_s)\,ds = \int_I f(x)\,dm(x) \quad \text{for measurable } f : I \to \mathbb{R},$$

and *β-mixing*, so that the β-mixing coefficient

$$\beta(t) = \int_I \|dP_t^x(y) - d\pi(y)\|_{\mathrm{TV}}\, d\pi(x)$$

satisfies $\lim_{t\to\infty} \beta(t) = 0$, where $\|\cdot\|_{\mathrm{TV}}$ is the total variation norm (see e.g., Doukhan, 1994, Sections 2.4 and 2.5, and Rogers and Williams, 1987, p. 303).

The *infinitesimal generator* $\mathcal{L}f = bf' + \frac{1}{2}\sigma^2 f''$ has a *spectral gap* if it has a strictly negative second largest (to 0) eigenvalue λ. This is so if and only if the solution to (2.1) is *ρ-mixing*, that is, the ρ-mixing coefficient

$$\rho(t) = \sup\left\{\frac{\|P_t f\|_{\mathbb{L}^2(I,\pi)}}{\|f\|_{\mathbb{L}^2(I,\pi)}} : f \in \mathbb{L}^2(I,\pi), \langle f, 1\rangle_{\mathbb{L}^2(I,\pi)} = 1\right\}$$

satisfies $\lim_{t\to\infty} \rho(t) = 0$, where $P_t f(x) = \int_I f(y)\, dP_t^x(y)$, as $\rho(t) = \mathrm{e}^{\lambda t}$, see Genon-Catalot, Jeantheau and Larédo (2000), Section 2.4. See Hansen, Scheinkman and Touzi (1998) and Genon-Catalot, Jeantheau and Larédo on more on spectral gaps.

3. Volatility induced stationarity

In this section we generalize the definition of *vis*, given by Conley, Hansen, Luttmer and Scheinkman (1997) for the special case of the CKLS model [see (5.1) below]. Throughout this section, we assume that Assumptions 2.1 and 2.3 hold.

Usually when describing the dynamics of an SDE, we split it into an ordinary differential equation (ODE) for the drift, and a diffusion white noise perturbation of the ODE. We expect the ODE to have a non-exploding solution that converges to the stationary mean of the SDE. This interpretation corresponds to an Euler approximation of the SDE. But when adding more terms to the Itô-Taylor expansion of the SDE (see Kloeden and Platen, 1995, Section 5.5), the picture changes, as this shows that the drift part and diffusion part interact.

The simplified interpretation of the SDE is often fruitful, but is some times incorrect. This lead Conley, Hansen, Luttmer and Scheinkman (1997) to define the concept of *vis* for the CKLS model. Building on their ideas, we make a general definition of *vis*:

Definition 3.1. The stationary solution to the SDE (2.1) has *vis* at l $[r]$, and we call l $[r]$ a *vis boundary*, if $S'(l^+) < \infty$ $[S'(r^-) < \infty]$. If $S'(l^+) > 0$ $[S'(r^-) > 0]$ we call l $[r]$ a *positive vis boundary* – otherwise it is a *null vis boundary*.

One may make a more general definition using lim inf and lim sup instead of limits.

Because of Assumption 2.3, we can only have *vis* boundaries at $\pm\infty$. In addition, since $m(I) < \infty$, to have *vis* at l $[r]$ it is necessary that $\sigma(l+) = \infty$ $[\sigma(r_-) = \infty]$.

For *vis* the mean reversion associated with stationarity comes from high volatility rather than from the drift. SDE with *vis* do not behave as white noise perturbed ODE.

Example 3.1. A local martingale in natural scale (i.e., zero drift) has *positive vis boundaries* at $\pm\infty$. A local martingale in natural scale with other boundaries than $\pm\infty$ cannot be stationary: An example of the latter is $dX_t = |X_t|dW_t$ with $X_t = 1$, for which $m(I) = m((0, \infty)) = \infty$.

For a diffusion in natural scale with *vis*, large values of the process give a high volatility and a small speed measure. Hence the diffusion moves away quickly from this area, in a direction that is locally unbiased, so that the process is a local martingale. But if the process moves towards an even higher value, then the volatility gets even larger and the speed measure even smaller, so that the time spent at large values is short: If the volatility gets large fast enough, it is possible to achieve stationarity. Still, the process can take very large values, during very short period of time. In addition, the fact that the process has a very high volatility for large values of the process, in

the described fashion, makes it likely that, within short, the process takes on a smaller value. This will prevent the local martingale from being a martingale!

Proposition 3.1. *Under Assumptions 2.1 and 2.3, if X is the solution to (2.1) with fixed initial value $\zeta = x$, then the local martingale $Y = S(X)$ is a martingale if and only if l and r are natural boundaries for X.*

Proof. We have entrance [natural] boundary for X at l and $[r]$, if and only if the local martingale $Y = S(X)$, with natural scale $S_Y(y) = y$ and speed measure $dm_Y(y) = 2/(\sigma(y)S'(y))dy$, has entrance [natural] boundary at $-\infty$ $[\infty]$, if and only if

$$\infty > \quad [\infty =] \quad \int_{-\infty}^{0} |x| \, dm_Y(x) \quad \left[\int_0^{\infty} x \, dm_Y(x) \right]$$

(see e.g., Karlin and Taylor 1981). By Arbib (1965), Theorem 3, X is a martingale if l and r are natural boundaries for X, so that $\pm\infty$ are natural boundaries for Y. Conversely, assume that e.g., ∞ is not a natural boundary for Y, so that $\int_0^{\infty} x \, dm_Y(x) < \infty$. Using Revuz and Yor (1999), p. 429, Exercise 3.18, considering the additive functionals $A_t = \int_0^t Y_s^+ \, ds$ and $C_t = t$ of Y, we then have

$$\lim_{t\to\infty} \mathbf{E}^y\{A_t\}/\mathbf{E}^y\{C_t\} = \lim_{t\to\infty} \frac{1}{t} \int_0^t \mathbf{E}^y\{Y_s^+\} \, ds = \int_0^{\infty} x \, dm_Y(x) \quad \text{for } y \in \mathbb{R}. \tag{3.1}$$

If Y is a martingale, then Y^+ is a submartingale, so that $\mathbf{E}^y\{Y_s^+\}$ is non-decreasing, and $\mathbf{E}^y\{Y_s^+\} \geq y$. But this contradicts (3.1) if we select $y > \int_0^{\infty} x \, dm_Y(x)$. □

We have the following interesting easy consequence of Proposition 3.1 and (3.1):

Corollary 3.1. *Under Assumptions 2.1 and 2.3, if X is the stationary solution to (2.1) and l and r are natural boundaries, then the stationary local martingale $Y = S(X)$ has infinite mean and is thus not a matingale.*

We have the following useful result:

Proposition 3.2. *Under Assumptions 2.1 and 2.3, a positive vis boundary l [r] is an entrance [a natural] boundary if and only if*

$$\infty > \quad [\infty =] \quad \int_l^{x_0} \frac{-x}{\sigma(x)^2} \, dx \quad \left[\int_{x_0}^{r} \frac{x}{\sigma(x)^2} \, dx \right]. \tag{3.2}$$

A null vis boundary l [r] is natural if the integral [integral] in (3.2) is infinite.

Proof. By definition, l [r] is an entrance [a natural] boundary if

$$\infty > \quad [\infty =] \quad \int_l^{x_0} \frac{S(x_0) - S(x)}{\sigma(x)^2 S'(x)}\, dx \quad \left[\int_{x_0}^{r} \frac{S(x) - S(x_0)}{\sigma(x)^2 S'(x)}\, dx\right]. \tag{3.3}$$

If l [r] is a positive vis boundary, then the criteria (3.2) follows from the fact that

$$\frac{-S(l^+)x}{2\sigma(x)^2} \leq \frac{S(x_0) - S(x)}{\sigma(x)^2 S'(x)} \leq \frac{-2S(l^+)x}{\sigma(x)^2} \qquad \left[\frac{S(r^+)x}{2\sigma(x)^2} \leq \frac{S(x) - S(x_0)}{\sigma(x)^2 S'(x)} \leq \frac{2S(r^+)x}{\sigma(x)^2}\right]$$

for $x > l$ small enough [$x < r$ large enough]. If l [r] is a null vis boundary, then we get

$$\frac{S(x_0) - S(x)}{\sigma(x)^2 S'(x)} \geq \frac{-S(l^+)x}{\sigma(x)^2} \qquad \left[\frac{S(x) - S(x_0)}{\sigma(x)^2 S'(x)} \geq \frac{S(r^+)x}{\sigma(x)^2}\right]$$

for $x > l$ small enough [$x < r$ large enough], so that the integral [integral] in (3.3) is infinite if the integral [integral] in (3.2) is infinite. □

Example 3.2. Given a constant $p > 1$, Assumptions 2.1 and 2.3 hold for the SDE

$$dX_t = (1 + |X_t|^p)\, dW_t, \quad X_0 = \zeta, \tag{3.4}$$

and $\pm\infty$ are positive *vis* boundaries (cf. Example 3.1). By Proposition 3.2, $\pm\infty$ are natural boundaries for $1 < p \leq 2$, and entrance boundaries for $p > 2$. Hence, solutions to (3.4) with fixed initial values $\zeta = x$ are martingales for $1 < p \leq 2$, and local martingales that are not martingales for $p > 2$, by Proposition 3.1.

A stationary local martingale with finite variance is either constant or not a martingale, see Bibby and Sørensen (1997), Lemma 2.1. Thus the stationary solution X is never a martingale. But it is uniformly integrable, by stationarity.

By Hansen, Scheinkman and Touzi (1998), Sections 4.1-4.2 and Example 4, we have a spectral gap and ρ-mixing for $p \geq 2$, but no spectral gap for $p < 2$.

For diffusions with drift the interpretation of *vis* is the same as for local martingales, as by the definition of *vis*, the volatility will dominate the drift close to *vis* boundaries.

If there is *vis* only at one boundary point, then the local martingale term of the solution to the SDE (2.1) will be biased, giving the solution a local martingale term with a decreasing or increasing mean, taking it away from the *vis* boundary.

There can be statistical problem for SDE with *vis*. For example, functions in the domain of the generator are important for statistical inference. But for SDE with *vis*, the only polynomials in the domain are constants, as functions f in the domain satisfy $(f'/S')(r^-) = (f'/S')(l^+) = 0$, see Hansen, Scheinkman and Touzi (1998), p. 10.

4. Simulation of SDE with *vis*

In this section we discuss computer simulation of SDE with *vis*. Such simulations can be difficult because standard requirements for convergence, such as Lipschitz conditions and linear growth for the drift and volatility, are not satisfied.

Let $\Delta W_n = W_{t_{n+1}} - W_{t_n}$ for equidistant times $0 = t_0 < t_1 < \ldots < t_N = T$ with spacing $\Delta > 0$. For $\theta_b, \theta_\sigma \in [0,1]$, a family of Euler schemes, starting at Y_0, are given by

$$Y_{n+1} = Y_n + \big(\theta_b \bar{b}(Y_{n+1}) + (1-\theta_b)\bar{b}(Y_n)\big)\,\Delta + (\theta_\sigma \sigma(Y_{n+1}) + (1-\theta_\sigma)\sigma(Y_n))\,\Delta W_n \quad (4.1)$$

for $n < N$: Here $\bar{b} = b - \theta_\sigma \sigma\sigma'$ is a correction term to ensure convergence to an Itô integral. For $\theta_b = \theta_\sigma = 0$ we get the *Euler* scheme, for $\theta_\sigma = 0$ the stochastic *theta* method, see Higham (2000), and for $\theta_b = \theta_\sigma = 1$ the *fully implicit Euler* scheme. The *backward Euler* scheme (BE) is the theta method with $\theta_b = 1$. The *split step backward Euler* scheme (SSBE), studied by Mattingly, Stuart and Higham (2002), is given by

$$Y_{n+1} = Y_n + \sigma(Y_n^*)\Delta W_n \quad \text{where} \quad Y_n^* = Y_n + b(Y_n^*)\Delta \quad \text{for } n < N.$$

For $I \subsetneq \mathbb{R}$ we prefer the fully implicit Euler scheme to the SSBE scheme, as the latter can move out of I. But the implicit Euler scheme has the drawback that the implicit equation may not have a unique solution: We discuss this issue in Section 5.

Higham, Mao and Stuart (2002), Theorem 2.2, prove uniform strong convergence of the Euler scheme, for locally Lipschitz drift and volatility, when the suprema of the true solution and the Euler scheme, started at a fixed $\zeta = x$, have moments of some order $p > 2$. With the additional assumptions that the drift is one-sided Lipschitz,

$$(x-y)(b(x)-b(y)) \le C\,|x-y|^2 \quad \text{for } x, y \in I, \text{ for some constant } C > 0,$$

and the volatility is globally Lipschitz, Higham, Mao and Stuart (2002) further show existence of all moments of the solution and of the SSBE approximation. Finally they examine convergence of implicit Euler schemes, which we will return to later.

Gyöngy (1998) shows uniform almost sure convergence of the Euler scheme assuming that the (continuous) drift is one-sided Lipschitz and the volatility locally Lipschitz.

An important issue, besides convergence of the numerical scheme, is to have an approximation with the same boundary properties as the true solution. Normally the

Euler scheme will induce a recurrent Markov chain on $\mathbb{R}$. However, for many SDE the probability to end up outside I is negligible already for moderately small Δ.

Stability of simulation schemes has been studied by Higham (2000), Mattingly, Stuart and Higham (2002) and Talay (2002). The next results show that, for many diffusions with *vis* boundaries at infinity, the Euler scheme can be transient.

Proposition 4.1. *Let $\{Y_n\}_{n\geq 0}$ be the Euler approximation of the SDE with (2.1) with σ strictly poitive. If there exist constants $p > 0$ and $\phi > 0$ such that*

$$\lim_{x\to\pm\infty} \frac{|x|^p\,|b(x)|}{|\sigma(x)|} = 0 \quad \textit{and} \quad \liminf_{x\to\pm\infty} \frac{|b(x)|}{|x|} \geq \phi, \tag{4.2}$$

then

$$\mathbf{P}\left\{\bigcap_{n\geq 0}\{|Y_{n+1}| \geq (1+\tilde{\phi})|Y_n|\}\right\} > 0 \quad \textit{for } \tilde{\phi} \in (0,\phi).$$

Proof. Given constants $c > 0$ and $y_0 > 0$, let

$$a_n = c \sup_{|x|\geq(1+\phi)^n y_0} \frac{|b(x)|}{|\sigma(x)|} \quad \text{for } n \geq 0. \tag{4.3}$$

By (4.2) there exist constants $c_1 > 0$ and $k > 0$ such that

$$\sup_{|x|\geq c_1} \frac{|x|^p\,|b(x)|}{|\sigma(x)|} < k.$$

Taking $y_0 \geq c_1$, this gives

$$\sum_{n=0}^{\infty} a_n = \sum_{n=0}^{\infty} \sup_{|x|\geq(1+\phi)^n y_0} \frac{|x|^p\,|b(x)|}{|\sigma(x)|}\,|x|^{-p} \leq \sum_{n=0}^{\infty} \sup_{|x|\geq(1+\phi)^n y_0} k\,|x|^{-p} = k\sum_{n=0}^{\infty}(1+\phi)^{-np} y_0$$

is finite. From this it is an elementary exercise to see that

$$\mathbf{P}\left\{\bigcap_{n\geq 0}\{|\Delta W_n| \geq a_n\}\right\} > 0. \tag{4.4}$$

And so it is enough to show that

$$|Y_{n+1}| \geq (1+\tilde{\phi})\,|Y_n| \quad \text{on} \quad \bigcap_{n\geq 0}\{|\Delta W_n| \geq a_n\}. \tag{4.5}$$

We now specify the choice of c and y_0 in (4.3): Pick $\tilde{\phi} \in (0,\phi)$ and put $c = 1+\Delta+2/\tilde{\phi}$. As $\sigma > 0$, the Euler scheme will hit any interval with positive probability. Therefore, we can without loss assume that the Euler scheme start at some $Y_0 = y_0$. Now choose $y_0 \geq c_1$ such that $|b(x)|\,/\,|x| \geq \tilde{\phi}$ for $|x| \geq y_0$ [cf. (4.2)].

First assume that $Y_{n+1}/Y_n \geq 1$, so that (4.5) reduces to $Y_{n+1}/Y_n - 1 \geq \tilde{\phi}$. Since

$$|\Delta W_n| \geq a_n \geq (1 + \Delta + \frac{2}{\tilde{\phi}})\frac{|b(Y_n)|}{|\sigma(Y_n)|},$$

(4.5) follows from the following calculation:

$$\frac{Y_{n+1}}{Y_n} - 1 = \frac{|\sigma(Y_n)|}{|Y_n|}\left|\frac{b(Y_n)}{\sigma(Y_n)}\Delta + \Delta W_n\right| \geq \frac{|\sigma(Y_n)|}{|Y_n|}\frac{|b(Y_n)|}{|\sigma(Y_n)|} \geq 2 + \tilde{\phi} \geq \tilde{\phi}.$$

For $Y_{n+1}/Y_n < 1$ we have

$$\begin{aligned}&\left|\frac{Y_{n+1}}{Y_n} - 1\right| \\ &= \left|\frac{b(Y_n)}{Y_n}\Delta + \frac{\sigma(Y_n)}{Y_n}\Delta W_n\right| \geq \left|\frac{\sigma(Y_n)}{Y_n}\right|\left(1 + \frac{2}{\tilde{\phi}}\right)\left|\frac{b(Y_n)}{\sigma(Y_n)}\right| = \left|\frac{b(Y_n)}{Y_n}\right|\left(1 + \frac{2}{\tilde{\phi}}\right) \geq 2 + \tilde{\phi},\end{aligned}$$

so that $Y_{n+1}/Y_n < -1$, since $Y_{n+1}/Y_n < 1$. From this in turn, we get (4.5) again:

$$\frac{|Y_{n+1}|}{|Y_n|} - 1 = \left|2 + \frac{b(Y_n)}{Y_n}\Delta + \frac{\sigma(Y_n)}{Y_n}\Delta W_n\right| \geq \tilde{\phi}. \quad \square$$

Many processes with *vis* are stationary local martingales. The following result shows that for these processes the Euler scheme can usually be transient.

Proposition 4.2. *Let $\{Y_n\}_{n\geq 0}$ be the Euler approximation of the SDE (2.1) with zero drift $b = 0$ and $\sigma > 0$. If*

$$\lim_{x \to \pm\infty} |x|^p / \sigma(x) = 0 \quad \textit{for some } p > 1, \tag{4.6}$$

then

$$\mathbf{P}\left\{\bigcap_{n\geq 0}\{|Y_{n+1}| \geq (1+\phi)Y_n\}\right\} > 0 \quad \textit{for } \phi > 0.$$

Proof. Pick a $\phi > 0$. As in the proof of Proposition 4.1 we may assume that the Euler scheme starts at some suitable $y_0 > 0$. Define

$$a_n = (2 + \phi) \sup_{|x| \geq (1+\phi)^n y_0} \frac{|x|}{\sigma(x)} \quad \text{for } n \geq 0.$$

From (4.6) it follows that $\sum_{n\geq 0} a_n < \infty$, so that (4.4) holds. And so the proposition follows from the methodology used for the proof of Proposition 4.1. $\square$

For the SDE in Propositions 4.1 and 4.2, instability starts when the Euler scheme is numerical large, making it oscillate between large positive and negative values.

For stationary diffusions in natural scale that are not too heavy tailed, the Euler scheme can be transient, as we have the following easy corollary to Proposition 4.2:

Corollary 4.1. *The Euler scheme is transient with positive probability, for a local martingale X given by (2.1), that admits a stationary distribution π such that*

$$\int_{-\infty}^{\infty} |x|^p \, d\pi(x) < \infty \quad \textit{for some } p > 1.$$

By the results of Gyöngy (1998), the Euler scheme converges almost surely as $\Delta \downarrow 0$. But for many SDE with *vis*, for any specific Δ, the Euler scheme may diverge.

Some times one can make a bijective transformation of X to another process with constant volatility and a drift that is one-sided Lipschitz: If Assumption 2.1 holds and the volatility in (2.1) has an absolutely continuous derivative, then the function $f(z) = \int^z 1/\sigma(x)\,dx$ is strictly increasing and $Z_t = f(X_t)$ satisfies the SDE

$$dZ_t = \left(\frac{b(f^{-1}(Z_t))}{\sigma(f^{-1}(Z_t))} - \frac{1}{2}\sigma'(f^{-1}(Z_t)) \right) dt + dW_t, \qquad Z_0 = f(\zeta). \tag{4.7}$$

If the drift in (4.7) is one-sided Lipschitz, then the BE and SSBE schemes will usually work well, see Mattingly, Stuart and Higham (2002).

5. The CKLS model

In Sections 5-7 we study three examples of SDE with various degrees of *vis*, and different sample path properties with different simulation and modelling problems.

For simulations we use the pseudo random number generator *ran2* from Flannery, Press, Teukolsky and Vetterling (1995). All simulations use the same underlying Brownian motion, so that they are started with the same seed $d = -1$. Even though we supply just a few figures, our numerical experience build on many more simulations.

Extending the *CIR* model of Feller (1951) and Cox, Ingersoll and Ross (1985), given constants $\alpha, \beta \in \mathbb{R}$ and $\sigma, \gamma > 0$, Chan, Koralyi, Longstaff and Sanders (1992) studied the *CKLS* model

$$dX_t = (\alpha + \beta X_t)\,dt + \sigma X_t^{\gamma}\,dW_t, \quad X_0 = \zeta > 0. \tag{5.1}$$

We first find boundaries, moments and the stationary distribution of the process. Then we study simulation problems. The CKLS model has a medium degree of *vis*.

5.1. Boundaries, ergodicity, moments and stationarity

We will consider the parameter values for which the SDE (5.1) has *vis*, which are

$$\{\tfrac{1}{2}<\gamma<1, \alpha>0, \beta=0\}\cup\{\gamma=1, \alpha>0, 0\le\beta<\tfrac{1}{2}\sigma^2\}\cup\{\gamma>1, \alpha>0\}\cup\{\gamma>1, \alpha=0, \beta>0\}.$$

For these parameters, Assumptions 2.1 and 2.3 hold, and we have ρ-mixing, see Hansen, Scheinkman and Touzi (1998). The boundary 0 is an entrance boundary for $\alpha > 0$ and natural for $\alpha = 0$. The *vis* boundary ∞ is null for $\gamma = 1$ and positive otherwise. By Proposition 3.2, ∞ is an entrance boundary for $\gamma > 1$ and natural for $\gamma \le 1$.

One way to explain the *vis* for the CKLS model is to consider the transformed process $Z_t = f(X_t)$, where $f(x) = x^{1-\gamma}$ $\gamma \ge \frac{1}{2}$, $\gamma \ne 1$. By Itô's lemma we have

$$dZ_t = \varphi(Z_t)\,dt + (1-\gamma)\sigma\,dW_t, \tag{5.2}$$

where

$$\varphi(z) = \alpha(1-\gamma)z^{\gamma/(\gamma-1)} + (1-\gamma)\beta z + \frac{1}{2}\gamma(\gamma-1)\sigma^2 z^{-1}. \tag{5.3}$$

Except for a constant, this is the transformation in (4.7).

For $\gamma > 1$, close to zero, the term z^{-1} in the drift pushes the process away from zero. For large values, the process is pushed down towards zero, by the term $z^{\gamma/(\gamma-1)}$ for $\alpha > 0$, and by the term z for $\alpha = 0$. For $\frac{1}{2} < \gamma < 1$, close to zero, the term $z^{\gamma/(\gamma-1)}$ pushes the process away from zero, but there is no strong downforce from high levels.

We find a formula for moments of the CKLS model below, as well as the stationary distribution, which can be used to check the fit of the CKLS model to data.

The stationary mean is infinite for $\gamma \le 1$: Moments of order $p < 2\gamma - 1$ exist for $\{\frac{1}{2} < \gamma < 1, \alpha > 0, \beta = 0\}$, and of order $p < 1 - 2\beta/\sigma^2$ for $\{\gamma = 1, \alpha > 0, 0 \le \beta < \frac{1}{2}\sigma^2\}$.

Proposition 5.1. *The local martingale part of the stationary solution is not a martingale.*

Proof. For $\gamma > 1$, we get this from that $\int_0^t X_s^\gamma\,dW_s$ has a decreasing mean over time, by Corollary 5.1 below. For $\gamma \le 1$, if $\sigma\int_0^t X_s^\gamma\,dW_s$ were a martingale, we would have

$$\mathbf{E}^\pi\left\{X_t - X_0 - \int_0^t \beta X_s\,ds\right\} = \alpha t + \mathbf{E}^\pi\left\{\int_0^t \sigma X_s^\gamma\,dW_s\right\} = \alpha t. \tag{5.4}$$

For $\beta = 0$, this contradicts stationarity, exactly as in Section 6.2 below. For $\beta \ne 0$, (5.4) with $t = 1$ and Fubini's theorem give $\mathbf{E}^\pi\{X_1 - X_0 - \beta X_{t_0}\} < \infty$ for almost

all $t_0 \in (0,1)$. Subtracting this from (5.4), and using Fubini's theorem again, we get $\mathbf{E}^\pi\{X_s - X_{t_0}\} < \infty$ for almost all $s \in (0,1)$. If some mean $\mathbf{E}^\pi\{X_s - X_{t_0}\} \neq 0$ we would get a linear behavior like (7.3), by recursion, which we reject, as in Section 6.2. And so, for almost all $s \in (0,1)$,

$$\alpha s = \mathbf{E}^\pi\left\{X_s - X_{t_0} - \int_0^s \beta(X_r - X_{t_0})\,dr + (1-\beta s)X_{t_0} - X_0\right\} = \mathbf{E}^\pi\{(1-\beta s)X_{t_0} - X_0\}.$$

But considering two distinct values of s, we get the contradiction $\mathbf{E}^\pi\{X_{t_0}\} < \infty$. □

Writing $\Gamma(\cdot)$ [$\Gamma(\cdot,\cdot)$] for the gamma function [incomplete gamma function], denote

$$M(p) = \sum_{k=0}^{\infty} \frac{1}{k!}\left[\frac{-\beta}{\sigma^2(\gamma-1)}\left(\frac{\sigma^2(2\gamma-1)}{2\alpha}\right)^{(2\gamma-2)/(2\gamma-1)}\right]^k \Gamma\left(k\frac{2\gamma-2}{2\gamma-1} + 1 - \frac{p}{2\gamma-1}\right),$$

$$M(p,x) = \sum_{k=0}^{\infty} \frac{1}{k!}\left[\frac{-\beta}{\sigma^2(\gamma-1)}\left(\frac{\sigma^2(2\gamma-1)}{2\alpha}\right)^{(2\gamma-2)/(2\gamma-1)}\right]^k \Gamma\left(k\frac{2\gamma-2}{2\gamma-1} + 1 - \frac{p}{2\gamma-1}, x\right).$$

Proposition 5.2. *Let $\gamma > 1$. For $p < 2\gamma - 1$ we have*

$$\mathbf{E}^\pi\{X_t^p\} = \begin{cases} \left(\dfrac{\sigma^2(2\gamma-1)}{2\alpha}\right)^{-p/(2\gamma-1)} \dfrac{M(p)}{M(0)} & \text{if } \alpha>0 \text{ and } \beta\in\mathbb{R}, \\ \left(\dfrac{\sigma^2(\gamma-1)}{\beta}\right)^{-p/(2\gamma-2)} \Gamma\left(\dfrac{2\gamma-1-p}{2\gamma-2}\right)\Big/\Gamma\left(\dfrac{2\gamma-1}{2\gamma-2}\right) & \text{if } \alpha=0 \text{ and } \beta>0, \end{cases} \tag{5.5}$$

while $\mathbf{E}^\pi\{X_t\} = \infty$ for $p \geq 2\gamma - 1$. The stationary distribution is given by

$$\mathbf{P}^\pi\{X_t > x\} = \begin{cases} \dfrac{1}{M(0)} M\left(0, \dfrac{2\alpha}{\sigma^2(2\gamma-1)}x^{1-2\gamma}\right) & \text{if } \alpha>0 \text{ and } \beta\in\mathbb{R}, \\ \Gamma\left(\dfrac{2\gamma-1}{2\gamma-2}, \dfrac{\beta}{\sigma^2(\gamma-1)}x^{2-2\gamma}\right)\Big/\Gamma\left(\dfrac{2\gamma-1}{2\gamma-2}\right) & \text{if } \alpha=0 \text{ and } \beta>0. \end{cases}$$

Proof. A non-normalized stationary density function is given by the speed measure

$$\frac{dm(x)}{dx} = 2\,x^{-2\gamma}\exp\left\{\frac{2\alpha}{\sigma^2(1-2\gamma)}x^{1-2\gamma} - \frac{\beta}{\sigma^2(\gamma-1)}x^{2-2\gamma}\right\},$$

so that $\int_0^\infty x^p\,dm(x) = \infty$ for $p \geq 2\gamma - 1$. We readily get the upper case in (5.5) from

$$\begin{aligned} &\int_0^\infty x^p\,dm(x) \\ &= \int_0^\infty 2\,x^{p-2\gamma}\exp\left\{\frac{2\alpha}{\sigma^2(1-2\gamma)}x^{1-2\gamma} - \frac{\beta}{\sigma^2(\gamma-1)}x^{2-2\gamma}\right\}dx \\ &= \frac{2}{2\gamma-1}\int_0^\infty y^{p/(1-2\gamma)}\exp\left\{\frac{2\alpha}{\sigma^2(1-2\gamma)}y\right\}\sum_{k=0}^{\infty}\frac{1}{k!}\left(\frac{-\beta}{\sigma^2(\gamma-1)}y^{(2-2\gamma)/(1-2\gamma)}\right)^k dy, \end{aligned}$$

and along the same lines, the lower case from

$$\int_0^\infty x^p\, dm(x) = \int_0^\infty 2\, x^{p-2\gamma} \exp\left\{ \frac{-\beta}{\sigma^2(\gamma-1)} x^{2-2\gamma} \right\} dx.$$

In the same fashion we calculate the stationary distribution function. □

Proposition 5.3. *Let $\gamma > 1$. We have*

$$\mathbf{E}^\pi\{X_t\} \in (0, -\alpha/\beta) \quad \textit{for } \beta < 0. \tag{5.6}$$

Further, for $\alpha > 0$ and $\beta \in \mathbb{R}$ we have

$$\lim_{\gamma\downarrow 1} \mathbf{E}^\pi\{X_t\} = \begin{cases} -\alpha/\beta & \textit{if } \beta<0 \\ \infty & \textit{if } \beta\geq 0 \end{cases} \quad \textit{and} \quad \lim_{\gamma\to\infty} \mathbf{E}^\pi\{X_t\} = \begin{cases} -\alpha/\beta & \textit{if } \beta<-\alpha \\ 1 & \textit{if } \beta\geq -\alpha \end{cases}, \tag{5.7}$$

while for $\alpha = 0$ and $\beta > 0$

$$\lim_{\gamma\downarrow 1} \mathbf{E}^\pi\{X_t\} = \begin{cases} \infty & \textit{if } \sigma^2 < 2\beta\, e \\ 0 & \textit{if } \sigma^2 \geq 2\beta\, e \end{cases} \quad \textit{and} \quad \lim_{\gamma\to\infty} \mathbf{E}^\pi\{X_t\} = 1. \tag{5.8}$$

Proof. By routine calculations, we have

$$\begin{aligned} M(1) &= \sum_{k=0}^\infty \frac{1}{k!} \left[\frac{-\beta}{\sigma^2(\gamma-1)} \left(\frac{\sigma^2(2\gamma-1)}{2\alpha} \right)^{(2\gamma-2)/(2\gamma-1)} \right]^k \Gamma\left((k+1) \frac{2\gamma-2}{2\gamma-1} \right), \\ M(0) &= 1 + \sum_{k=1}^\infty \frac{1}{(k-1)!} \left[\frac{-\beta}{\sigma^2(\gamma-1)} \left(\frac{\sigma^2(2\gamma-1)}{2\alpha} \right)^{(2\gamma-2)/(2\gamma-1)} \right]^k \left(\frac{2\gamma-2}{2\gamma-1} \right) \Gamma\left(k \frac{2\gamma-2}{2\gamma-1} \right). \end{aligned} \tag{5.9}$$

From this in turn, we readily obtain

$$M(0) = 1 - \frac{\beta}{\alpha} \left(\frac{\sigma^2(2\gamma-1)}{2\alpha} \right)^{-1/(2\gamma-1)} M(1).$$

Combining this with (5.5), we readily get (5.6) from

$$\mathbf{E}^\pi\{X_t\} = 1 \Big/ \left[\left(\frac{\sigma^2(2\gamma-1)}{2\alpha} \right)^{1/(2\gamma-1)} \frac{1}{M(1)} - \frac{\beta}{\alpha} \right]. \tag{5.10}$$

To prove (5.7) for $\beta < 0$ and $\gamma \downarrow 1$, by (5.10), it is enough to show that $\lim_{\gamma\downarrow 1} M(1) = \infty$. This in turn follows by inspection of the following version of (5.9):

$$M(1) = \sum_{k=0}^\infty \frac{1}{(k+1)!} \left[\frac{-\beta}{\sigma^2(\gamma-1)} \left(\frac{\sigma^2(2\gamma-1)}{2\alpha} \right)^{(2\gamma-2)/(2\gamma-1)} \right]^k \left(\frac{2\gamma-1}{2\gamma-2} \right) \Gamma\left((k+1) \frac{2\gamma-2}{2\gamma-1} + 1 \right).$$

To prove (5.7) for $\alpha > 0$, $\beta < -\alpha$ and $\gamma \to \infty$, by (5.10), it is enough to show that $\lim_{\gamma\to\infty} M(1) = \infty$. This in turn follows from the fact that, by routine calculations, term number k in the sum for $M(1)$ in (5.9) converges to $(-\beta/\alpha)^k$ as $\gamma \to \infty$.

To prove (5.7) for $\alpha > 0$, $\beta \geq 0$ and $\gamma \downarrow 1$, notice that $\mathbf{E}^\pi\{X_t\}$ is an non-decreasing function of α and β, by inspection of (4.7). And so it is enough to show $\lim_{\gamma\downarrow 1} \mathbf{E}^\pi\{X_t\} = \infty$ for $\beta = 0$. This in turn we get observing, by (5.5),

$$\mathbf{E}^\pi\{X_t\} = \left(\frac{\sigma^2(2\gamma-1)}{2\alpha}\right)^{-1/(2\gamma-1)} \Gamma\left(\frac{2\gamma-2}{2\gamma-1}\right) \quad \text{for } \alpha > 0 \text{ and } \beta = 0.$$

To prove (5.7) for $\alpha > 0$, $-\alpha \leq \beta < \alpha$ and $\gamma \to \infty$, it is enough to consider $-\alpha < \beta < \alpha$, by the monotonicity noted in the previous paragraph. This in turn we get from (5.10) and that term number k in the sum $M(1)$ goes to $(-\beta/\alpha)^k$ as $\gamma \to \infty$.

The fact that (5.7) holds for $\alpha > 0$, $\beta \geq \alpha$ and $\gamma \to \infty$ follows from (5.8): This is so because when we start with $\lim_{\gamma\to\infty} \mathbf{E}^\pi\{X_t\} = 1$ for $\alpha = 0$ and $\beta > 0$, and succesively increase α, by the noted monotonicity property together with (5.7) for $\alpha > 0$, $\beta < \alpha$ and $\gamma \to \infty$, we must have $\lim_{\gamma\to\infty} \mathbf{E}^\pi\{X_t\} = 1$ for all intermediate $\alpha \in [0, \beta]$.

It remains to prove (5.8). Here the limit when $\gamma \to \infty$ is immediate from (5.5), while the limit when $\gamma \downarrow 1$ follows from (5.5) together with Stirking's formula. □

As $\mathbf{E}^\pi\{X_t\} < -\alpha/\beta$ for $\alpha > 0$, $\beta < 0$ and $\gamma > 1$, Chan, Koralyi, Longstaff and Sanders (1992), Equation 2, misspecified the first moment conditions for their generalized method of moments approach: The reason is that they overlooked the *vis.*

Corollary 5.1. *Let $\gamma > 1$. We have*

$$\mathbf{E}^\pi\left\{\sigma \int_0^t X_s^\gamma dW_s\right\} = -\left(\alpha + \beta \int_0^\infty x \, d\pi(x)\right) t \equiv -d(\alpha, \beta, \sigma, \gamma)t, \tag{5.11}$$

where $d(\alpha, \beta, \sigma, \gamma) > 0$. Further, for $\alpha > 0$ and $\beta \in \mathbb{R}$ we have

$$\lim_{\gamma\downarrow 1} d(\alpha, \beta, \sigma, \gamma) = \begin{cases} 0 & \text{if } \beta < 0 \\ \infty & \text{if } \beta \geq 0 \end{cases} \quad \text{and} \quad \lim_{\gamma\to\infty} d(\alpha, \beta, \sigma, \gamma) = \begin{cases} 0 & \text{if } \beta < -\alpha \\ \alpha+\beta & \text{if } \beta \geq -\alpha \end{cases},$$

while for $\alpha = 0$ and $\beta > 0$

$$\lim_{\gamma\downarrow 1} d(\alpha, \beta, \sigma, \gamma) = \begin{cases} \infty & \text{if } \sigma^2 < 2\beta e \\ 0 & \text{if } \sigma^2 \geq 2\beta e \end{cases} \quad \text{and} \quad \lim_{\gamma\to\infty} d(\alpha, \beta, \sigma, \gamma) = \beta.$$

Conley, Hansen, Luttmer and Scheinkman (1997), p. 529, claim that a diffusion with a (positive) *vis* at infinity behaves, for large values, as a Brownian motion. But, for the CKLS model the scale function diminishes all the obtained large values when transforming back from the naturale scaled version, so that the process is different from the Brownian motion. And even if the local martingale is unbiased towards the direction, the speed of the clock will introduce a mean reverting bias. Thus, Conley, Hansen, Luttmer and Scheinkman (1997) give a misleading interpretation of *vis*.

Since we have an entrance *vis* boundary at infinity, the local martingale $\sigma \int_0^t X_s^\gamma \, dW_s$ will induce mean reversion for large values of the process, but display martingale behavior for smaller values, giving it a decreasing mean. For $\beta < 0$ the mean reversion comes from both *vis* and the drift. This is the reason that the stationary mean is always less than $-\alpha/\beta$. The value $d(\alpha, \beta, \sigma, \gamma)$ of the drift of the local martingale in (5.11) measures the size of the *vis*.

5.2. Strong approximations and simulations of the CKLS model

For $\gamma \leq 1$, Muszta, Richter, Albin and Astrup Jensen (2005), Theorem 2.1, gives uniform convergence in mean-square of the Euler scheme, see Section 7.1. For $\gamma > 1$, the Euler scheme breaks down with positive probability, by Proposition 4.1. For example, for $\gamma = 50$, the scheme breaks down almost immediately with $\Delta = 10^{-9}$.

Broze, Scaillet and Zakoïan (1995), pp. 220-221, prove that $\mathbf{P}\{Y_n^2 \to \infty\} > 0$ for $\gamma > 1$. We think their proof contains an error: As a first step, they show that

$$\mathbf{P}\{Y_{n+1}^2 \geq (\phi + 1) Y_n^2 \mid Y_n^2\} \geq 1 - d > 0 \quad \text{for } |Y_n| > M, \tag{5.12}$$

for some constants $M, \phi, d > 0$. To finish the proof, they claim that (5.12) and Petruccelli and Woolford (1984), p. 274, give $\mathbf{P}\{\lim_{n\to\infty} Y_n^2 = \infty | Y_0\} > 0$. Working through the details in Petruccelli and Woolford (1984), p. 274, the inequality

$$\mathbf{P}\{Y_{n+1}^2 \geq (\phi + 1) Y_n^2 \mid Y_n^2\} \geq 1 - c/Y_n^2 = 1 - d_n > 0 \quad \text{for } |Y_n| > M, \tag{5.13}$$

for some constant $c > 0$ (for a different model), is used to get the claimed convergence. However, the requirement in (5.13) is stronger than the requirement in (5.12)!

We now examine implicit Euler schemes for the CKLS model, see (4.1). As the non-Lipschitz part of the CKLS model is the volatility (not the drift), we need an implicit

scheme ($\theta_\sigma > 0$). The explicit part of the implicit scheme for the CKLS model is

$$ {}_x\tilde{Y}_n = {}_xY_n + (1-\theta_b)(\alpha + \beta_x Y_n - \theta_\sigma\gamma\sigma^2\, {}_xY_n^{2\gamma-1})\Delta + (1-\theta_\sigma)\sigma_x Y_n^\gamma \Delta W_n, $$

and the non-linear equation to be solved for the implicit part is $f({}_xY_{n+1}) = {}_x\tilde{Y}_n$, where

$$ f(y) = y - \theta_b(\alpha + \beta y - \theta_\sigma\gamma\sigma^2 y^{2\gamma-1})\Delta - \theta_\sigma\sigma y^\gamma \Delta W_n $$

(with x referring to the initial value). The issue whether $f(Y_{n+1}) = \tilde{Y}_n$ has a unique solution Y_{n+1} depends on the size of ΔW_n: The derivative f' has global minimum at

$$ \bar{y} = \left[\frac{\Delta W_n}{2\theta_b\sigma(2\gamma-1)\Delta}\right]^{1/(\gamma-1)} \quad \text{with} \quad f'(\bar{y}) = 1 - \theta_b\beta\Delta - \frac{\theta_\sigma\gamma}{4\theta_b(2\gamma-1)}\frac{(\Delta W)^2}{\Delta}. $$

The two roots $r_\pm$, $r_- < r_+$, are given by

$$ r_\pm = \left[\frac{\Delta W_n \pm \sqrt{d}}{2\theta_b\sigma(2\gamma-1)\Delta}\right]^{1/(\gamma-1)} \quad \text{where} \quad d = (\Delta W_n)^2 - 4(1-\theta_b\beta\Delta)\frac{\theta_b(2\gamma-1)}{\theta_\sigma\gamma}\Delta. \tag{5.14} $$

If $\theta_b\beta\Delta < 1$ and the discriminant is positive, then the roots $r_\pm$ are both positive [negative] if $\Delta W_n > 0$ [$\Delta W_n < 0$]. The asymptotic probability that f is not monotone is $\mathbf{P}\{\mathrm{N}(0,1) > 2\sqrt{\theta_b(2\gamma-1)/(\theta_\sigma\gamma)}\}$ as $\Delta \downarrow 0$. With $\theta_b = \theta_\sigma = 1$ and γ close to one, that probability is 2.3%, decreasing to 0.23% for large values of γ. For f not monotone, there are several options.

We suggest the following: If f is not monotone, then for $Y_n \in [f(r_+), f(r_-)]$ the solution $f(Y_{n+1}) = Y_n$ is not unique. To have the continuity $\tilde{Y}_{n+1} \to Y_{n+1}$ as $\tilde{Y}_n \to Y_n$, we must select the smallest of the possible solutions. One option is to use the full implicit scheme for $Y_n < f(r_+)$, where the solution is unique. Another option is to use the full implicit scheme for $Y_n < f(r_-)$, and in cases of more than one solution choose the smallest one. The motivation is that the solution of an SDE in the interval $[t_n, t_{n+1}]$ should depend only on the Brownian motion in the interval, together with the drift and volatility at X_{t_n}, so that this information should suffice to determine $X_{t_{n+1}}$.

The question remains what to do if $Y_n > f(r_+)$ [$Y_n > f(r_-)$]. We suggest that θ_b and θ_σ are adjusted so that $f(Y_{n+1}) = Y_n$ gets a unique solution. The simplest choice is to take an explicit step ($\theta_b = \theta_\sigma = 0$). In our test examples, the explicit step is applied very seldom, and only if $\Delta W_n > 0$ is large, at the same time as $Y_n > f(r_\pm)$. In practice we get a stable numerical scheme, even for very large values of γ.

We now describe yet another and fruitful approach to simulate the CKLS model, which is to simulate the transformed process (5.2). With a constant volatility the implicit schemes of interest reduce to the theta method; and for $\theta_b = 1$ the BE scheme,

$$ {}_z\tilde{Y}_n = {}_zY_n + (1-\theta_b)\varphi({}_zY_n)\Delta + \sigma(1-\gamma)\Delta W_n, $$

where φ is defined in equation (5.3). The non-linear equation to be solved is $g({}_zY_{n+1}) = {}_z\tilde{Y}_n$ where $g(z) = z - \theta_b\varphi(z)\Delta$. As $g : (0,\infty) \to (-\infty,\infty)$ invertible, the BE scheme defines a solution which can attain any positive value, exactly like the true solution.

The full implicit scheme is promising as the function φ in (5.3) is C^1 and one-sided Lipschitz. Since b has a polynomial behaviour at infinity, Mattingly, Stuart and Higham (2002), Theorem 5.3, shows that for the true solution and the continuous-time extension of the BE approximation, all moments of order $p > 2$ are uniformly bounded, and the approximation converges uniformly in L^2 with order $\frac{1}{2}$. By transforming back with the inverse function $f^{-1}({}_zY_n)$, $f^{-1}(y) = y^{1/(1-\gamma)}$ which is decreasing and convex, we get our benchmark for the true solution sample path of the CKLS model.

We will now consider some numerical examples of CKLS models.

In Figure 1 in Appendix A, the left panel shows the stationary mean as a function of $\gamma > 1$, in part illustrating the limit behavior from Proposition 5.3, while the right panel shows the stationary density for two sets of parameter values.

We illustrate the results from Corollary 5.1 in Figure 2.

Another way to measure the reversion back to the stationary level is by the spectral gap. Genon-Catalot, Jeantheau and Larédo (2000) give the upper bound $1/(8Cm((0,\infty))^2)$ for that gap, where C is the median of π. With the same parameters as in Figure 1, we display this bound in Figure 3. Even though the gap estimate may be crude, it shows strong exponential mean reversion for large values of γ. The reversion grows with β. As the stationary density moves to the right when β increases, this suggests that excursions from the mean become higher and shorter, like peaks.

Sample paths are different for the cases $\gamma < 1$ and $\gamma > 1$: For $\gamma > 1$ the process escapes the mean during short explosive peaks, while for $\gamma < 1$, the clock is slower.

Simulation of the CKLS model using the transformation (5.2). We now give two numerical examples of the CKLS model with $\gamma = 50$, giving an extremely high *vis.*

The other parameters are $\alpha = 1$, $\beta = \pm 1$ and $\sigma = 2$, to illustrate the difference between pure *vis* ($\beta = 1$) and a diffusion with stationarity from both *vis* and drift ($\beta = -1$).

In Figure 4 the left panels show two sample paths of the CKLS model using the full implicit Euler scheme $({}_zY_n)$, with $X_0 = 1$ and $\Delta = 10^{-9}$, and transformed back to the CKLS model, see (5.2). Note that for $\beta = 1$ the process has a lot of peaks. Both processes are reverting extremely fast back to the stationary level. The transformed processes (5.2) are plotted in the right panels, and they behave wildly!

The peaks for the processes in Figure 4 are very narrow. To see the behavior at peaks, we display two windows with volatile periods in Figure 5. Note that the excursions from the mean are very short. Also, the process in the right panel ($\beta = -1$) has a longer excursion than the one in the left panel ($\beta = 1$). This is the general picture, and is also confirmed by estimates for the spectral gap in Figure 3.

Albeit the simulation scheme converges as $\Delta \downarrow 0$, the CKLS model is hard to simulate for large γ. In contrast to the explicit scheme, which overestimates the peaks and ends up transient, the implicit scheme underestimates peaks, giving a downward bias. For example, the peaks close to 1.15 in Figure 5 reduce to 1.07 for the step size 10^{-5}.

Simulation of the CKLS model using modified implicit schemes. Consider the difference between a modified full implicit scheme $({}_xY_n)$ and the transformed scheme $f^{-1}({}_zY_n)$. The full implicit scheme for the transformed process converges by Higham, Mao and Stuart (2002). We trust the Euler approximation $({}_zY_n)$ and define the relative error for a modified implicit Euler scheme, $({}_xY_n)$, as the relative difference between the two implicit Euler schemes

$$\varepsilon_n = 2\frac{f^{-1}({}_zY_n) - {}_xY_n}{f^{-1}({}_zY_n) + {}_xY_n}, \quad n = 0, 1, \ldots, \qquad f^{-1}(z) = z^{\frac{1}{1-\gamma}}.$$

Note that $\varepsilon_0 = 0$ and that $\varepsilon_n > 0$ when a modified full implicit Euler scheme ${}_xY_n$ underestimates the true value. In most cases, the two schemes attain their extremes at the same times, but in rare cases there are differences.

We used the modified full implicit scheme with the left root. And so, for ${}_xY_n \leq f(r_-)$ we use the full implicit Euler scheme, while otherwise we take an explicit step.

The right panel of Figure 6 shows the relative error ε_n for $\beta = -1$. Remarkably, we did not take any explicit steps. Large errors arise around high peaks, when the

transformed process $({}_zY_n)$ is close to zero. Otherwise, the error is of magnitude 10^{-6}.

The left panel in Figure 6 shows $\beta = 1$. The bracketed numbers indicate the numbers of explicit steps applied around peaks. Except for at peaks, the relative error ε_n is about 10^{-6}, just as for $\beta = -1$. To illustrate this, we have in Figure 7 divided $[0, 1]$ into 1000 equally long subintervals, and evaluated the ε_n only at the largest value of the process in each subinterval. The figure is very similar to Figure 6.

To give a better picture of the problem around the peaks we have plotted the error process (ε_n) in Figure 8 with the same window as in Figure 5.

We may instead take explicit steps whenever ${}_xY_n$ is larger than $f(r_+)$, using $f(r_+)$ as the boundary for where to take implicit and explicit steps. This will not improve this significantly. The error using the root r_+ is plotted in the left panel of Figure 9.

Instead of full explicit steps, we could take $\theta_b = 1$ and $\theta_\sigma < 1$, still using $f(r_+)$ as the boundary for where to stop with implicit steps. For a "troubled" increments ΔW_n, we choose a θ_σ such that the discriminant is negative, see (5.14). Specifically, we take

$$\theta_b = 1 \quad \text{and} \quad \theta_\sigma = \frac{2(1 - \theta_b\beta\Delta)\theta_b(2\gamma - 1)\Delta}{\gamma(\Delta W_n)^2}.$$

For this θ_σ, the discriminant in (5.14) is zero for $d = 2\theta_\sigma$, while positive for $\theta_\sigma = 1$. Hence $\theta_\sigma \in (0, 1)$. The right panel of Figure 9 shows the relative error for the scheme.

All schemes seems to differ at the peaks, because at peaks the clock is so fast that, even with step size 10^{-9}, the process can change value up to 0.03 for a single step.

The case of Chan, Koralyi, Longstaff and Sanders (1992). As a last example we use the parameters from the paper by Chan, Koralyi, Longstaff and Sanders (1992),

$$\alpha = 0.0408,\ \beta = -0.5921,\ \sigma^2 = 1.6704,\ \gamma = 1.4999. \tag{5.15}$$

The data set in Chan, Koralyi, Longstaff and Sanders (1992) is the monthly yields of U.S. one-month Treasury Bills, 1964-89 (CRSP). They view the data as discretely observed data from the CKLS model, hence $\Delta = \frac{1}{12}$. The stationary mean is 0.06886 and $-\alpha/\beta = 0.06891$, so the true mean is in this case only slightly smaller than if no *vis* was present. Still this SDE displays different behavior than for the case $\frac{1}{2} \le \gamma < 1$.

In Figure 10 we have plotted the CKLS model (5.15) for 25 years (left panel) and 1000 years (right panel), with the latter showing the high peak characteristics of *vis*.

We decompose the process X into its drift part $X_0 + \int_0^t (\alpha + \beta X_s)\, ds$ and local martingale part $\sigma \int_0^t X_s^\gamma\, dW_s$. We start by simulating X_t as in Figure 10. Next we simulate the drift part and the local martingale part as

$$\int_{t_n}^{t_{n+1}} (\alpha + \beta X_s)\, ds \approx \alpha\Delta + \frac{\beta}{2}(X_{t_{n+1}} + X_{t_n})\Delta, \tag{5.16}$$

$$\sigma \int_{t_n}^{t_{n+1}} X_s^\gamma\, dW_s \approx X_{t_{n+1}} - X_{t_n} - \alpha\Delta - \frac{\beta}{2}(X_{t_{n+1}} + X_{t_n})\Delta. \tag{5.17}$$

If no *vis* effect were present, the local martingale part should have had zero mean, so that (5.16) and (5.17) were zero mean processes, under the stationary measure.

Figure 11 shows the drift and local martingale parts of the SDE started at 0.06886 (left panel), and the drift and the local martingal parts together with the solution started at 4, but moved different constant steps in the y direction to facilitate viewing (right panel). Note that the negative drift of the local martingale is recognizable only when the process is much larger than the stationary mean.

As long as the process is close to the stationary mean there is no recognizable drift effect (left panel of Figure 11) and hence, close to the mean, the reversion is controlled by the drift. In other words, the *vis* effect is only present for large values. As discussed earlier this is the case when β is dominating, $\beta < -\alpha$, which is the case here.

Endnotes for the CKLS example. To show that our numerical schemes capture the drift in the local martingale (5.17), we have included the new parameter set $\alpha = 1$, $\beta = 1$, $\sigma = 2$, and $\gamma = \frac{3}{2}$, with stationary mean 3.13 and *vis* effect $d(\alpha, \beta, \sigma, \gamma) = 4.13$.

Figure 12 displays the process with its local martingale and drift components. Since $\beta > 0$ there is no reversion from the drift. The stationary local martingale has mean $-4.13t$. For Figure 11 we argued that there is a close connection between peaks of the local martingale and of the process: Figure 12 shows an almost perfect match!

6. A hyperbolic diffusion with *vis*

Our second example is within the hyperbolic class of diffusions, see Bibby and Sørensen (2003). The actual SDE we study is taken from Bibby and Sørensen (1997):

$$dX_t = \sigma \exp\Big\{ \frac{1}{2}\Big(\alpha\sqrt{\delta^2 + (X_t - \mu)^2} - \beta(X_t - \mu)\Big)\Big\}\, dW_t, \quad X_0 = \zeta.$$

To satisfy Assumptions 2.1 and 2.3 the parameters must satisfy $\alpha > |\beta| \geq 0$, $\delta, \sigma > 0$ and $\mu \in \mathbb{R}$. Then the diffusion is also ρ-mixing, as $\pm\infty$ are entrance boundaries, see Hansen, Scheinkman and Touzi (1998), Section 4. There is a high degree of *vis*

The stationary solution has all moments, some of which are calculated by Bibby and Sørensen (2003). The stationary density is maximal at $\mu+\delta\beta/\sqrt{\alpha^2-\beta^2}$. As in Example 3.2, the stationary local martingale is uniformly integrable but not a martingale.

Denoting the spectral gap λ, ergodicity and ρ-mixing gives $\mathbf{E}^x\{\int_0^t \sigma(X_s)\,dW_s\} = O(e^{\lambda t})$ as $t \to \infty$ for $x \in \mathbb{R}$, while $\mathbf{E}^\pi\{\int_0^t \sigma(X_s)\,dW_s\} = 0$. Hence mean reversion is not due to bias of the local martingale, but that the clock is very quick at large values.

6.1. Time-changed simulation

By Corollary 4.1, the Euler scheme is transient with positive probability. This should then apply also to higher order explicit schemes, and from numerical tests we have seen that the Milstein and strong 1.5 schemes start to oscillate between large positive and negative values. However, Bibby and Sørensen (1997) report that the strong 1.5 scheme works for them!

The implicit scheme is problematic because the equation to be solved will not have a unique solution. The transformation method that was used for the CKLS model is not an option here, as the function f in (4.7) and its inverse have to be evaluated by time-consuming numerical methods, and the drift in (4.7) is not one-sided Lipschitz.

We will describe an alternative way to simulate a local martingale as a time-changed Brownian motion, which will be refered to as *time-changed simulation*. First we set up some notation (see e.g., Revuz and Yor, 1999, Chapter 5):

Let $[X]_t = \int_0^t \sigma^2(X_s)\,ds$ be the quadratic variation process. If T_t solves $[X]_{T_t} = t$, then $B_t = X_{T_t} - X_0$ is a Brownian motion: The idea is to simulate B and get X from

$$T_t = \int_0^t \frac{du}{\sigma^2(X_0 + B_u)} \quad \text{and} \quad X_{T_t} = X_0 + B_t.$$

Consider a grid $0 = t_0 < t_1 < \ldots < t_N = T$ with spacing Δ. Given T_{t_n}, we have

$$X_{T_{t_{n+1}}} = X_0 + B_{t_{n+1}} \quad \text{where} \quad T_{t_{n+1}} - T_{t_n} = \int_{t_n}^{t_{n+1}} \frac{du}{\sigma^2(X_0 + B_u)}. \tag{6.1}$$

For B close to the boundaries the speed measure gets small, so that $T_{t_{n+1}} - T_{t_n}$ is very small, which makes computations slow. We will now explain how to deal with this:

Pick intervals $(a_1, b_1) \subsetneq (a_2, b_2)$ and numbers $\varepsilon_1 > \varepsilon_2 > 0$ such that $\sigma^2(x)$ is strictly monotone outside (a_1, b_1) with $1/\sigma^2(a_1) = 1/\sigma^2(b_1) = \varepsilon_1$ and $1/\sigma^2(a_2) = 1/\sigma^2(b_2) = \varepsilon_2$. As long as $B_{t_n} \in (a_2 - X_0, b_2 - X_0)$ we calculate $X_{T_{t_{n+1}}}$ by (6.1). If B reaches the bound $a_2 - X_0$ $[b_2 - X_0]$ at the time $\tau = t_i$, then we stop the scheme (6.1) and start it again at the first time $\rho = t_j$ at which B returns to the bound $a_1 - X_0$ $[b_1 - X_0]$.

To approximate the integrals in (6.1), we use a trapezoid rule for regular steps from t_n to t_{n+1}, and for jumps with $\rho - \tau \leq 1$. For larger jumps we use a grid with spacing $\tilde{\Delta} = 0.1$. As we know the values of the Brownian motion at the endpoints, we simulate intermediate values according to a (time-scaled) Brownian bridge.

By the strong *vis*, the hyperbolic diffusion spends very little time at numerically large values, so that the excursions in (T_τ, T_ρ) should not influence "overall properties" of the solution significantly. Still, in some applications, these excursions could be important.

6.2. A numerical example

The parameter found by Bibby and Sørensen (1997) for stock price dynamics were

$$\alpha = 4.4875,\ \beta = -3.8412,\ \delta = 1.1949,\ \mu = 7.2915,\ \sigma = \sigma_0 = 0.0047.$$

This gives a stationary distribution with mean 4.1705 and variance 3.8943, which is plotted in the left panel of Figure 13 in Appendix A.

Simulations by the time-change scheme, with $\Delta = 10^{-5}$, $T = 1000$, $\varepsilon_1 = 0.1$, $\varepsilon_2 = 0.01$ and $X_0 = 4.1705$ (the stationary mean) are displayed to the right in Figure 13. There we also considered $\sigma = \sigma_1 = 0.047$, meaning that the clock runs 100 times faster.

7. A family of heavy tailed SDE

Given a constant $a < \frac{1}{3}$, our third SDE is given by

$$dX_t = 3X_t^a\, dt + 3X_t^{2/3}\, dW_t, \quad X_0 = \zeta > 0. \tag{7.1}$$

This SDE satisfies Assumptions 2.1 and 2.3, with speed measure given by

$$\frac{m(x)}{dx} = \frac{2}{9} x^{-4/3} \exp\left\{ \frac{2}{3a-1} x^{a-1/3} \right\} \quad \text{for } x > 0. \tag{7.2}$$

The boundary 0 is natural and ∞ a positive *vis* natural boundary. There is no spectral gap, see Hansen, Scheinkman and Touzi (1998), Section 4: The lack of ρ-mixing indicates that the process can in relatively long periods stay away from the stationary level.

Stationary moments of order $p < \frac{1}{3}$ exist, and are easily calulated from (7.2). The local martingale part $\int_0^t 3X_s^{2/3}\,dW_s$ of the stationary solution to (7.1) is not a martingale, because this would contradict stationarity in the following way:

$$\mathbf{E}^\pi\{X_t - X_0\} = \mathbf{E}^\pi\left\{\int_0^t 3X_s^a\,ds + \int_0^t 3X_s^{2/3}\,dW_s\right\} = 3t\int_0^\infty x^a\,d\pi(x) < \infty. \quad (7.3)$$

The stationary density attains its maximum at $x = 2^{3/(3a-1)}$. Notice that the process is pushed towards zero as a increases. Obviously, accurate simulations of (7.1) is an issue of both stability at infinity, and control of the boundary at zero.

7.1. Simulation techniques

Itô-Taylor scemes are strong uniformly convergent under linear growth and global Lipschitz drift and volatility, see Kloeden and Platen (1995), Section 10.6. For $a \geq 0$, we can have an Itô-Taylor scheme that is uniformly closer than δ to the true solution, with probability $1-\delta$, for any $\delta \in (0,1)$, by considering a modified SDE, as in the proof of Muszta, Richter, Albin and Astrup Jensen (2005), Theorem 2.1, see the next paragraph, picking $\Delta > 0$ as well as the parameter $\varepsilon > 0$ of that proof small enough.

In Muszta, Richter, Albin and Astrup Jensen (2005), Theorem 2.1, it is shown that, for a diffusion on $I = (0,\infty)$, satisfying Assumptions 2.1 and 2.3, and that has linear growth, it is enough with a global Lipschitz condition on any given interval $[y,\infty) \subset I$, $y > 0$, for the Euler scheme to be uniformly convergent in mean-square. The proof make use of Higham, Mao and Stuart (2002), Theorem 2.2, where the true solution is replaced with a modified "nicer" SDE for values in the interval $(0,\varepsilon)$, and the Euler scheme extended by symmetry to a scheme on $\mathbb{R}$, to establish the moment bounds required by Higham, Mao and Stuart, through some martingale inequalities.

By Muszta, Richter, Albin and Astrup Jensen (2005), Theorem 2.1, the Euler scheme is uniformly convergent in mean-square for $a \geq 0$.

For $a < 0$ explicit schemes are unstable, as they jump, from zero to high values from which there is little reversion. This in turn, is not a property displayed by the true solution. For example, for $a = -10$, we have noted instability for $\Delta = 10^{-9}$!

As for the CKLS model, we can use the transformation (4.7), with $f(x) = x^{1/3}$ and $\varphi(z) = z^{3a-2} - z^{-1}$ one-sided Lipschitz, so that $Z_t = f(X_t)$ satisfies

$$dZ_t = \varphi(Z_t)\,dt + dW_t, \quad Z_0 = f(\zeta) > 0. \quad (7.4)$$

With the notation of the CKLS example, the partial explicit step for (7.4) is

$$ {}_z\tilde{Y}_n = {}_zY_n + (1-\theta_b)\varphi({}_zY_n)\Delta + \Delta W_n. $$

The equation to be solved is $g({}_zY_{n+1}) = {}_z\tilde{Y}_n$, where $g(z) = z - \theta_b\varphi(z)\Delta$. Then we transform back with $f^{-1}(z) = z^3$, and the approximation of X_{t_n} is ${}_xY_n = f^{-1}({}_zY_n)$. By Higham, Mao and Stuart (2002), this scheme converges to the true solution.

7.2. A numerical example

The left panel in Figure 14 in Appendix A, shows the SDE (7.1) simulated in the interval $[0, 10]$ for $a = -10$, using the BE scheme ($\theta_b = 1$) with $\Delta = 10^{-6}$ and $X_0 = 0.935 -$ the maximum of the speed density. Notice the ability to escape from the value 0.935 during long periods. On the other hand, the right panel displays the process in the interval $[8, 10]$, and shows that it also stays close to 0.935 for long periods.

In Figure 15, everything is as in Figure 14, except that $a = \frac{1}{10}$ and the maximum of the speed density is at $X_0 = 0.05127$. Note that the behaviour of the process for large values (the left panel), is virtually identical to that when $a = -10$, illustrating that we have *vis*. On the other hand, the process spends much less time close to the maximum of the speed density, than when $a = -10$ (right panel).

8. Conclusion

We have studied a class of stationary and ergodic SDE, for which stationarity is ensured by a high volatility - *vis*. SDE with *vis* escape from the stationary level often, but in short periods if the *vis* is strong. Between these excursions, the process is very steady. As SDE with stationarity from mean reversion by the drift do not behave in this manner, SDE with *vis* are important additions to more conventional models.

The *vis* appears at the boundaries of the SDE. Either we have a local martingale part of the solutions that is not a martingale, with a mean reverting drift, or we have a martingale or non-martingale local martingale part that is mean reverting, simply because the clock runs quicker nere boundaries. We conclude that interpretations and statistical methods for SDE with *vis* in the literature have to be revised (see below).

To work with SDE with *vis* it is important to have simulation methods. We considered simulated of strong solutions. As virtually all traditional simulation schemes

may fail for SDE with *vis*, we considered alternative methods.

We gave three examples of SDE with a different degree of *vis*: The CKLS model has a medium degree of *vis*, where some but not all moments exist, and is or is not ρ-mixing. We simulated CKLS by transformation of the SDE and by modified implicit schemes. An SDE with a stationary distribution in the hyperbolic class is ρ-mixing with a high degree of *vis*, and all moments exist. These SDE were simulated by change of time. The class of heavy tailed SDE were not ρ-mixing, and had a weak *vis* with an infinite stationary mean. They were simulated by transformation of the SDE.

There are mistakes in the literature on inference for SDE with *vis*.

Appendix A. Figures

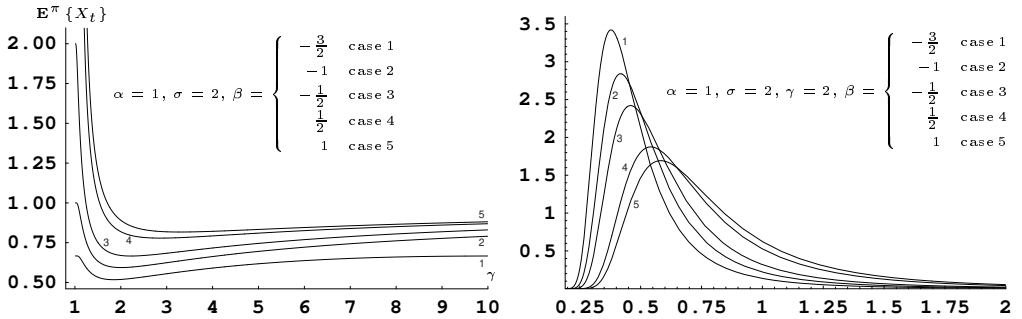

FIGURE 1: The stationary mean as a function of γ and examples of the shape of the stationary density

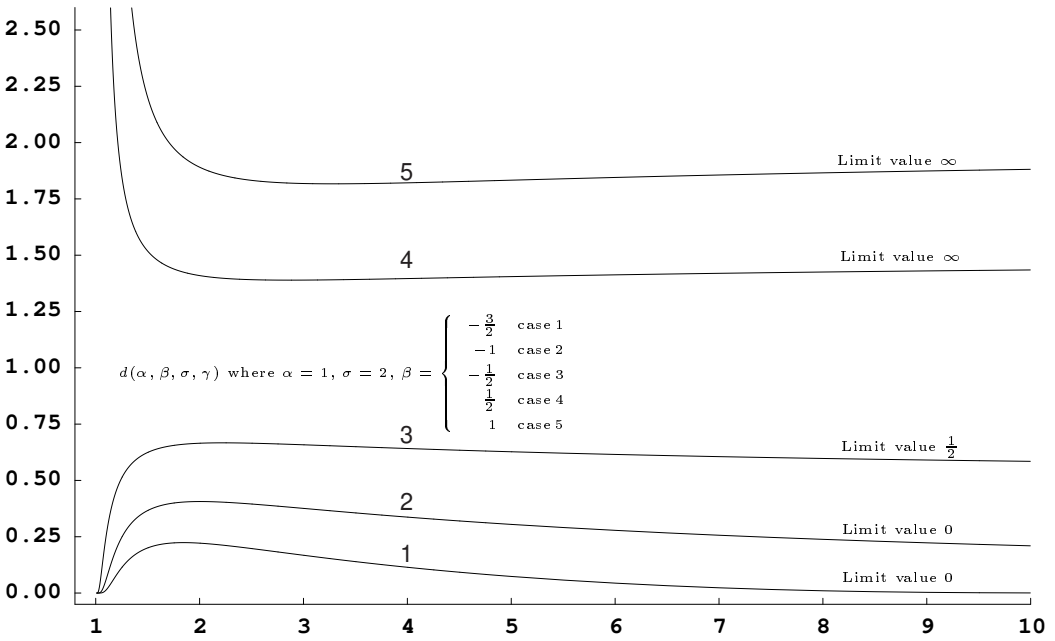

FIGURE 2: Minus the mean, $d(\alpha, \beta, \sigma, \gamma)$, (5.11), of the local martingale, (5.17)

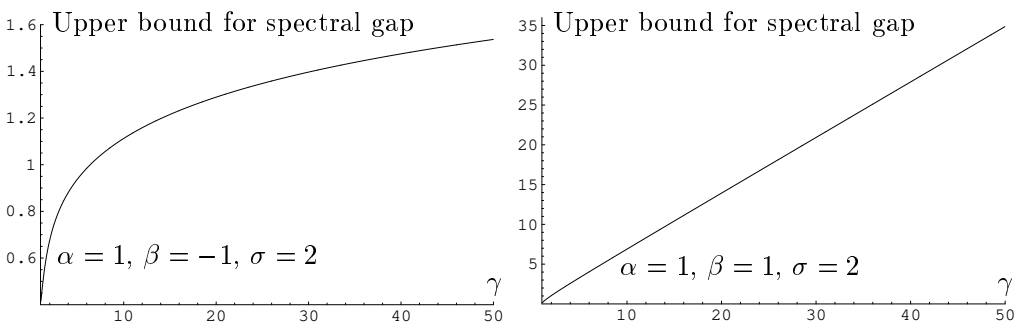

FIGURE 3: A upper bound for the spectral gap in the CKLS model

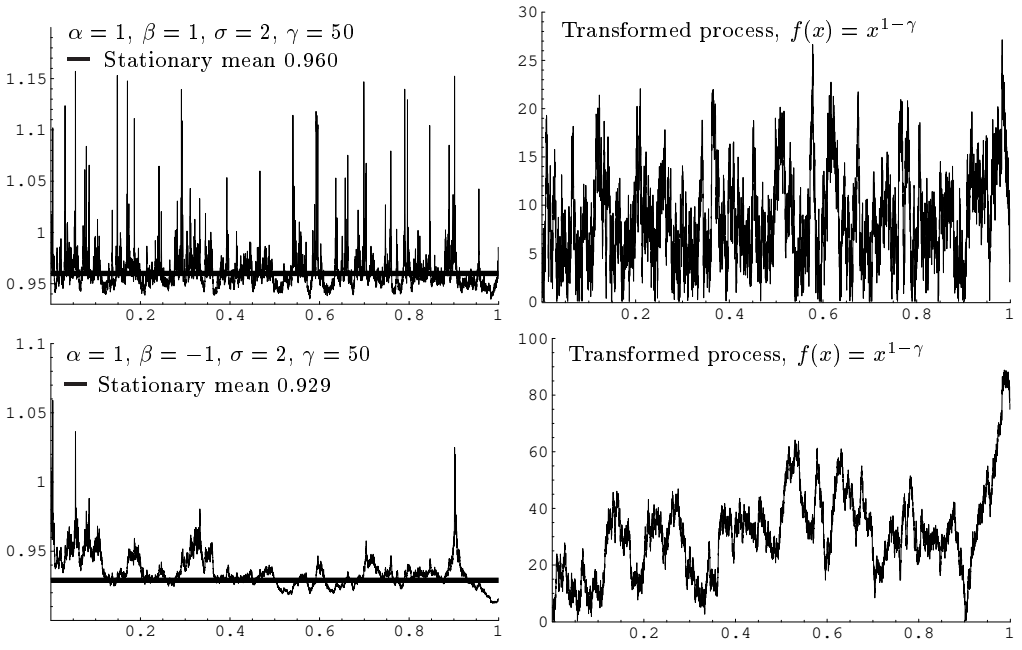

FIGURE 4: The CKLS model with $\Delta = 10^{-9}$ sample paths

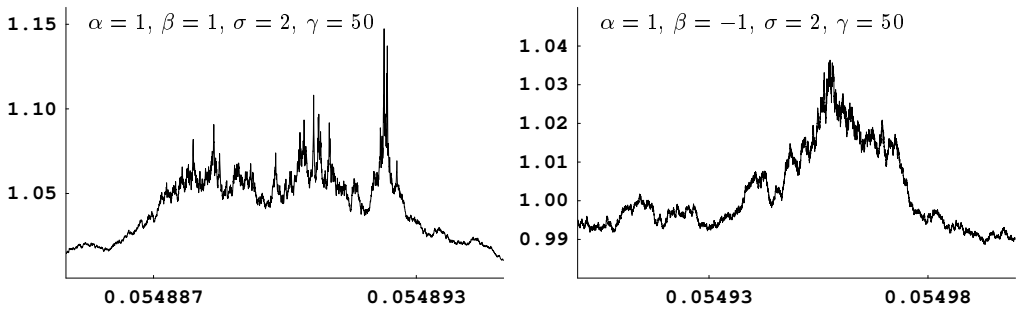

FIGURE 5: Two windows each covering one of the volatile periods for one of the processes in the left panel of Figure 4

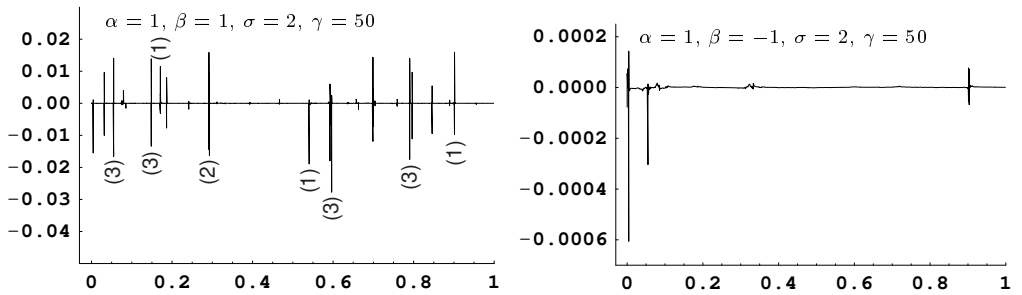

FIGURE 6: The relative error (ε_n) with $\Delta = 10^{-9}$ for the CKLS model, Figure 4

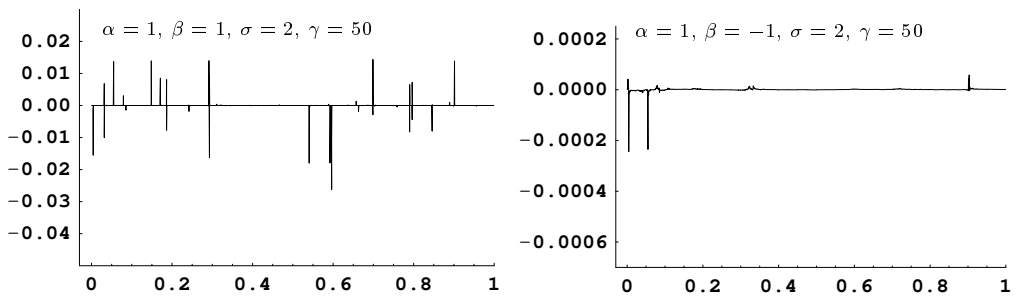

FIGURE 7: (ε_n) evaluated only at the extreme points for each subinterval

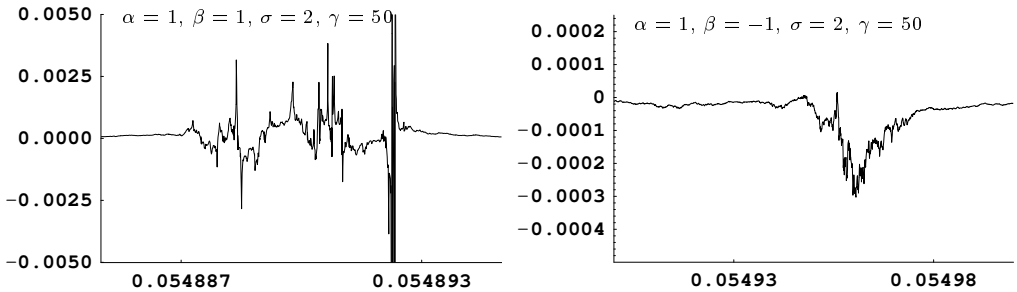

FIGURE 8: (ε_n) for the windows in Figure 5

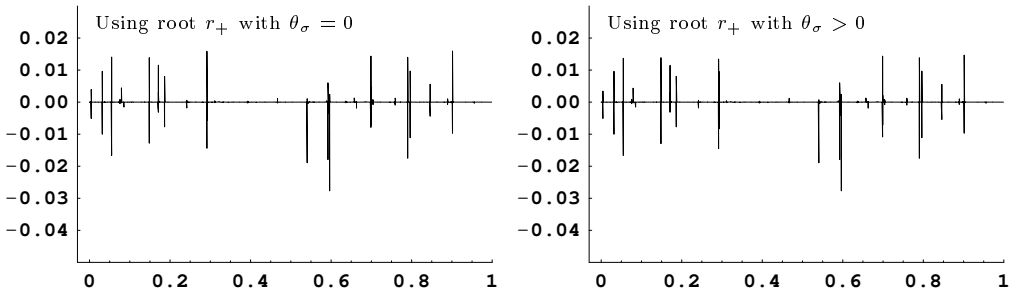

FIGURE 9: The error (ε_n), $\Delta = 10^{-9}$, for two modified schemes

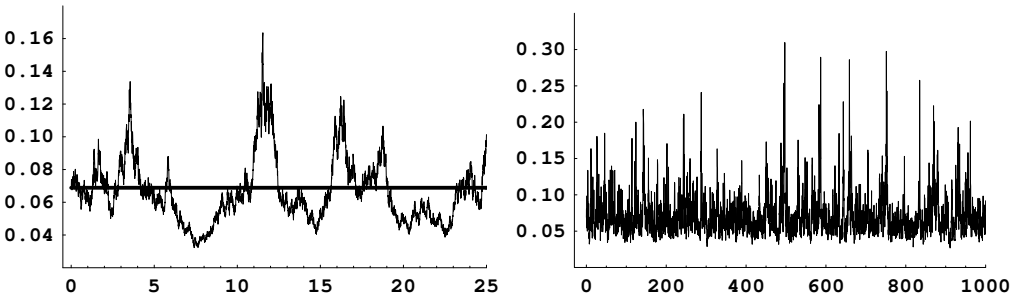

FIGURE 10: The CKLS model, $X_0 = 0.06886$, $\Delta = \frac{1}{12}10^{-5}$

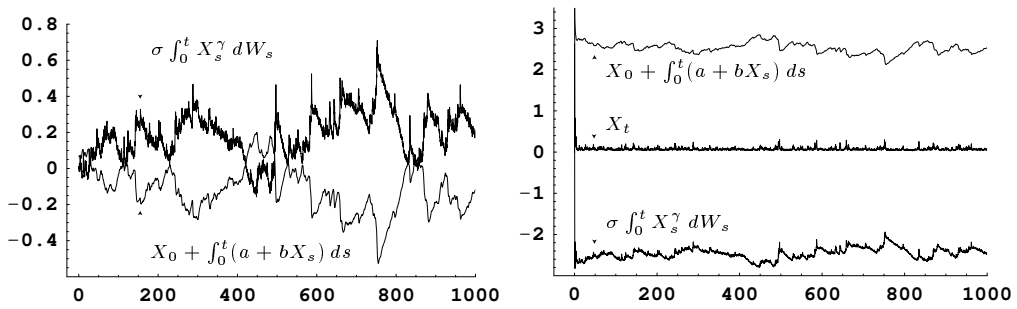

FIGURE 11: Decomposing X_t to its drift part and its local martingale part (5.17) for two initial values $X_0 = 0.06886$ (left panel) and $X_0 = 4$ (right panel)

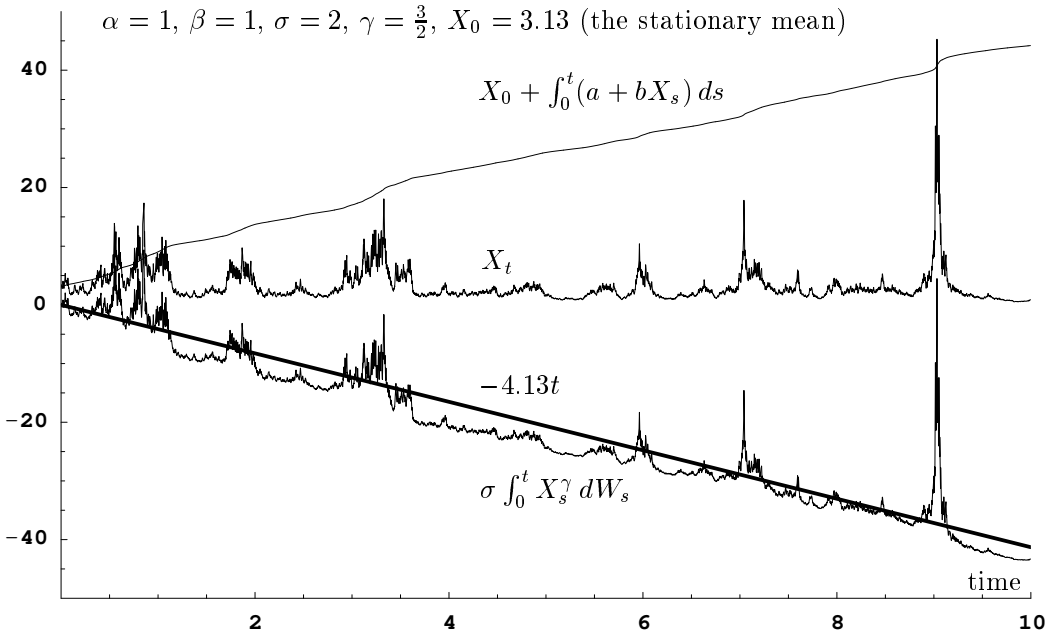

FIGURE 12: Decomposing X_t to its drift part and its local martingale part (5.17)

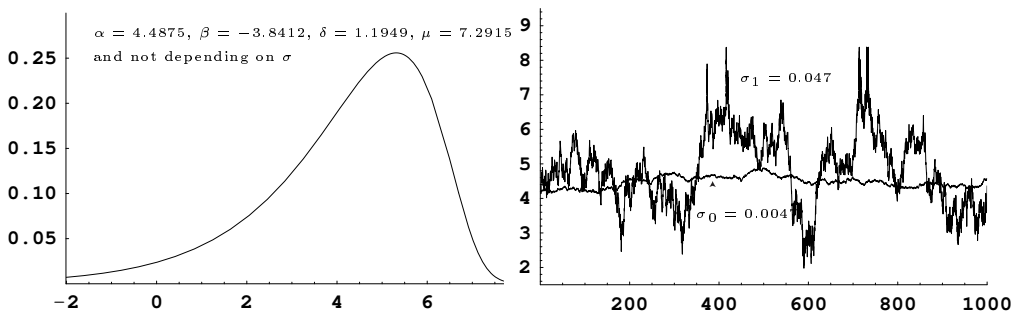

FIGURE 13: Hyperbolic diffusion: The stationary density and two sample paths with different speeds

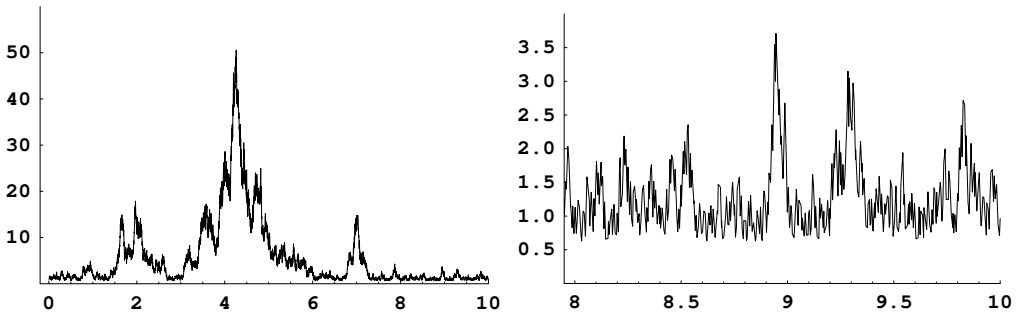

FIGURE 14: Heavy tailed diffusion with $a = -10$, $X_0 = 0.935$ and $\Delta = 10^{-6}$.

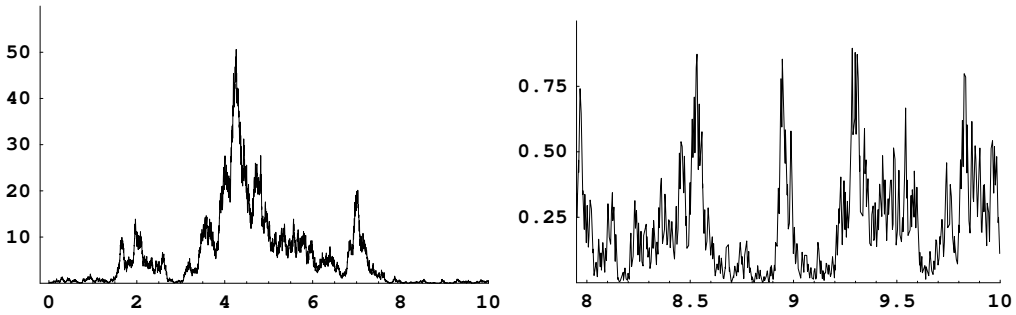

FIGURE 15: Heavy tailed diffusion with $a = \frac{1}{10}$, $X_0 = 0.05127$ and $\Delta = 10^{-6}$.

Acknowledgement

The third author, J.M.P. Albin, was supported by The Swedish Research Council contract no. 621-2003-5214.

References

[1] Arbib, M. A. (1965). Hitting and martingale characterizations of one-dimensional diffusions. *Z. Wahrsch. verw. Gebiete* **4**, 232–247.

[2] Bibby, B. M. and Sørensen, M. (1997). A hyperbolic diffusion model for stock prices. *Finance and Stochastics* **1**, 25–41.

[3] Bibby, B. M. and Sørensen, M. (2003). Hyperbolic processes in finance. In: Rachev, S. T. (ed.) *Handbook of Heavy Tailed Distributions in Finance*, 211–248. Elsevier, New York.

[4] Broze, L., Scaillet, O. and Zakoïan, J.-M. (1995). Testing for continuous-time models of the short-term interest rate. *J. Empirical Finance* **2**, 199–223.

[5] Chan, K., Koralyi, G., Longstaff, F. and Sanders, A. (1992). An empirical comparison of alternative models of the short-term interest rate. *J. Finance* **47**, 1209–1227.

[6] Conley, T. G., Hansen, L. P., Luttmer, E. G. and Scheinkman, J. A. (1997). Short-term interest rates as subordinated diffusions. *The Review of Financial Studies* **10**, 525–577.

[7] Cox, J., Ingersoll, J. and Ross, S. (1985). A theory of the term structure of interest rates. *Econometrica* **53**, 385–408.

[8] Doukhan, P. (1994). *Mixing. Properties and Examples.* Springer, Berlin.

[9] Feller, W. (1951), Two singular diffusion problems. *Ann. Math.* **54**, 173–182.

[10] Flannery, B. P., Press, W. H., Teukolsky, S. A. and Vetterling, W. T. (1995). *Numerical Recipes in C*, 2nd edn. Cambridge, Cambridge.

[11] Genon-Catalot, V., Jeantheau, T. and Larédo, C. (2000). Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli* **6**, 1051–1079.

[12] Gyöngy, I. (1998). A note on Euler approximations. *Potential Anal.* **8**, 205–216.

[13] Hansen, N. R. (2003). Geometric ergodicity of dicrete approximations of multivariate diffusions. *Bernoulli* **9**, 725–743.

[14] Hansen, L. P., Scheinkman, J. A. and Touzi, N. (1998). Spectral methods for identifying scalar diffusions. *J. Econometrics* **86**, 1–32.

[15] Higham, D. J. (2000). Mean-square and asymptotic stability of the stochastic Theta method. *SIAM J. Numer. Anal.* **38,** 753–769.

[16] Higham, D. J., Mao, X. and Stuart, A. (2002). Strong convergence of numerical methods for nonlinear stochastic differential equations. *SIAM J. Numer. Anal.* **40,** 1041–1063.

[17] Karatzas, I. and Shreve, S. E. (1991). *Brownian Morion and Stochastic Calculus*, 2nd edn. Springer, Berlin.

[18] Karlin, S. and Taylor, H. M. (1981). *A Second Course in Stochastic Processes.* Academic, New York.

[19] Kloeden, P. E. and Platen, E. (1995). *Numerical Solution of Stochastic Differential Equations.* Springer, Berlin.

[20] Mattingly, J., Stuart, A. and Higham, D. (2002). Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stochastic Process. Appl.* **101,** 185–232.

[21] Muszta, A., Richter, M., Albin, J. M. P. and Astrup Jensen, B. (2005). A note on the Lipschitz condition for Euler approximation of stochastic differential equations. Submitted to *SIAM J. Numer. Anal.*

[22] Petruccelli, J. D. and Woolford, S. W. (1984). A treshold AR(1) model, *J. Appl. Probab.* **21,** 270–286.

[23] Revuz, D. and Yor, M. (1999). *Continuous Martingales and Brownian Motion.* Springer, Berlin.

[24] Rogers, L. C. G. and Williams, D. (1987). *Diffusions, Markov Processes, and Martingales. Volume 2: Itô Calculus.* Wiley, New York.

[25] Talay, D. (2002). Stochastic Hamiltonian dissipative systems with non globally Lipschitz coefficients: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. *Markov Process. Related Fields* **8,** 163–198.

PAPER III

On Numerical Solution of Stochastic Differential Equations driven by Lévy Processes

Anders Muszta

Mathematical Sciences, Chalmers University of Technology, 412 96 Gothenburg, Sweden.

Abstract

This article presents a stochastic Taylor formula for the solution of a stochastic differential equation driven by a Lévy process. The formula give numerical schemes for the equation, for which we present mean-square order of convergence results.

Key words:
Convergence order, Lévy process, numerical method, Protter inequality, semimartingale, stochastic Taylor expansion

1991 MSC: 60G44, 60H10, 60H35, 60J30, 60J60, 60J65, 60J75, 65C30, 68U20

1 Introduction

In 1992, the by now standard work: Numerical Solution of Stochastic Differential Equations, by Peter E. Kloeden and Eckhard Platen was published.

Email address: muszta@math.chalmers.se (Anders Muszta).

The basis for the work of Kloeden and Platen is the Stochastic Taylor Formula, to represent a solution of an SDE driven by Wiener processes, as a sum of multiple stochastic integrals. Approximate solutions are constructed by retaining a subset of the full representation. Kloeden and Platen then proceed to find convergence properties of the approximations.

We derive a Stochastic Taylor Formula for SDE driven by Lévy processes, and get approximation schemes by retaining a subset of the full representation. Further, we find mean-square convergence properties of the approximations.

2 Prerequisites

We use the Einstein summation convention: If in an expression a letter appears as a subscript and as a superscript, then summation is carried out over that letter. If not obvious from the context, the range of the summation is indicated.

For example, $\sum_{i=1}^{n}\sum_{j=2}^{m} a_{ij}b^i c^j$ is written $a_{ij}b^i c^j$, $i \in \{1,\dots,n\}$, $j \in \{2,\dots,m\}$.

2.1 Itô Formula

For a d-dimensional semimartingale $X = (X^1,\dots,X^d)$ and a function $f : [0,\infty)\times\mathbb{R}^d \to \mathbb{R}$ that is $\mathcal{C}^{1,2}([0,\infty)\times\mathbb{R}^d)$, the Itô Formula reads

$$\begin{aligned} f(t,X_t) = f(s,X_s) &+ \int_s^t \partial_r f(r,X_{r-})\,dr + \int_s^t \partial_j f(r,X_{r-})\,dX_r^j \\ &+ \frac{1}{2}\int_s^t \partial_i\partial_j f(r,X_{r-})\,d[X^i,X^j]_r^c \\ &+ \sum_{s\le r\le t}\Big(f(r,X_r) - f(r,X_{r-}) - \partial_j f(r,X_{r-})\,\Delta X_r^j\Big). \end{aligned}$$

Let $L = (L^1,\dots,L^m)$ be an m-dimensional Lévy process, built of independent

one dimensional square integrable Lévy component processes $\{L_t^p\}_{t\in[0,\infty)}$. The law of the component Lévy processes are determined by triples $(\gamma^p, \sigma(p), \nu^p)$, where $\gamma^p \in \mathbb{R}$ and $\sigma(p) \geq 0$ are constants and ν^p a Lévy measure.

Let the semimartingale X be a solution to the d-dimensional SDE

$$dX_t^j = a^j(t, X_{t-})\, dt + b_p^j(t, X_{t-}) dL_t^p, \quad j = 1, \ldots, m. \tag{1}$$

The continuous part $[X^i, X^j]^c$ of the covariation between X^i and X^j satisfies

$$\begin{aligned}
&[X^i, X^j]_r^c \\
&= \int_0^r a^i(u, X_{u-}) b_p^j(u, X_{u-})\, d[u, L^p]_u^c + \int_0^r a^j(u, X_{u-}) b_q^i(u, X_{u-})\, d[L^q, u]_u^c \\
&\qquad + \int_0^r b_p^i(u, X_{u-}) b_q^j(u, X_{u-})\, d[L^p, L^q]_u^c.
\end{aligned}$$

The covariation process $[u, L^p]_r$ is zero, as the following calculation shows: Taking $m = d = 1$, $f(t, x) = tx$ and $X_t = L_t^p$, the Itô Formula becomes

$$rL_r^p = \int_0^r L_u^p\, du + \int_0^r u\, dL_u^p + \sum_{0<u\leq r} (u(L_u^p - L_{u-}^p) - u(L_u^p - L_{u-}^p)).$$

And so we get $[u, L^p]_r = 0$, by the defining relation for the covariation

$$rL_r^p = \int_0^r L_u^p\, du + \int_0^r u\, dL_u^p + [u, L^p]_r.$$

As by Applebaum (2004), pp. 225 and 230, $[L^p, L^q]_r^c = \sigma(p)\sigma(q)\delta_{p,q} r$, we get

$$[X^i, X^j]_r^c = \sigma^2(p) \int_0^r b_p^i(u, X_{u-}) b_p^j(u, X_{u-})\, du.$$

This in turn shows that the Ito Formula for $f(t, X_t)$ can be expressed as

$$\begin{aligned}
f(t, X_t) = f(s, X_s) &+ \int_s^t \mathcal{L}_0 f(r, X_{r-})\, dr + \int_s^t \mathcal{L}_p f(r, X_{r-})\, dL_r^p \\
&+ \sum_{s\leq r\leq t} \Big(f(r, X_r) - f(r, X_{r-}) - \partial_j f(r, X_{r-}) \Delta X_r^j \Big),
\end{aligned}$$

where

$$\begin{aligned}\mathcal{L}_0 f(r, X_{r-}) &= \partial_r\, f(r, X_{r-}) + a^j(r, X_{r-})\partial_j\, f(r, X_{r-}) \\ &\qquad + \frac{\sigma^2(p)}{2} b_p^i(u, X_{u-}) b_p^j(u, X_{u-})\partial_i\partial_j\, f(r, X_{r-}) \\ \mathcal{L}_p f(r, X_{r-}) &= b_p^j(r, X_{r-})\partial_j\, f(r, X_{r-}).\end{aligned}$$

By the mean-value theorem, there exists a stochastic process $\theta(X) = \{\theta_t(X)\}_{t\geq 0}$ such that, if $\Delta X_t^j = 0$ then $\theta_t(X)^j = X_t^j$, while if $\Delta X_t^j \neq 0$ then

$$\theta(X)_t^j \in (X_{t-}^j \wedge X_t^j, X_{t-}^j \vee X_t^j) \quad \text{and} \quad f(t, X_t) - f(t, X_{t-}) = \partial_j f(t, \theta_t(X))\Delta X_t^j.$$

Assume that the following Lévy process J is well-defined:

$$J_t = \sum_{0<s\leq t} \Delta L_s \quad \text{for } t \geq 0. \tag{2}$$

By Bertoin (1996), p. 15, if $\int_{\mathbb{R}\setminus\{0\}} 1 \wedge |x|\, \nu(dx) < \infty$, then the sum (2) is absolutely convergent. However, the sum (2) is not absolutely convergent in general, see e.g., Applebaum (2004), p. 87: We do not know exact conditions for the convergence of the sum (2), but simply assume that it takes place.

If we define the operator $\mathcal{L}_{m+p}$ as

$$\mathcal{L}_{m+p} f(r, X_{r-}) = b_p^j(r, X_{r-})\Big(\partial_j f(r, \theta_r(X)) - \partial_j f(r, X_{r-})\Big),$$

then the Itô Formula for $f(t, X_t)$ can be written

$$\begin{aligned}&f(t, X_t) \\ &= f(s, X_s) + \int_s^t \mathcal{L}_0 f(r, X_{r-})\, dr + \int_s^t \mathcal{L}_p f(r, X_{r-})\, dL_r^p + \int_s^t \mathcal{L}_{m+p} f(r, X_{r-})\, dL_r^{m+p},\end{aligned}$$

where the differential dL_r^{m+p} is to be understood as $dJ_r^p = \Delta L_r^p$, so that

$$\int_s^t \mathcal{L}_{m+p} f(r, X_{r-})\, dL_r^{m+p} = \sum_{s\leq r\leq t} b_p^j(r, X_{r-})\Big(\partial_j f(r, \theta_r(X)) - \partial_j f(r, X_r)\Big)\, \Delta L_r^p.$$

The final form of the Itô Formula is

$$f(t, X_t) = f(s, X_s) + \int_s^t \mathcal{L}_k f(r, X_{r-})\, d\mathcal{L}_r^k, \tag{3}$$

where the differential $d\mathcal{L}_t^k$ is defined by

$$d\mathcal{L}_t^k = \begin{cases} dt & \text{for } k = 0, \\ L_t^k & \text{for } k \in \{1, \ldots, m\}k \\ dJ_t^{k-m} & \text{for } k \in \{m+1, \ldots, 2m\}. \end{cases}$$

2.2 Multi-index

Consider the set $\mathcal{M}$ of all *multi-indices* $(j_1, \ldots, j_k)$, where $j_1, \ldots, j_k \in \{0, 1, \ldots, m, m+1, \ldots, 2m\}$ and $k \in \mathbb{N}$, and where v denotes the empty multi-index.

Let $l(\alpha)$ denote the *length* of $\alpha \in \mathcal{M}$, and $n(\alpha)$ its number of 0-*components*. Introduce operators $-\alpha$, $\alpha-$ and $\alpha * \beta$ on multi-indices, by

$$\begin{aligned} -\alpha: &\quad -(j_1, \ldots, j_k) = (j_2, \ldots, j_k), \\ \alpha-: &\quad (j_1, \ldots, j_k)- = (j_1, \ldots, j_{k-1}), \\ \alpha * \beta: &\quad (i_1, \ldots, i_k) * (j_1, \ldots, j_l) = (i_1, \ldots, i_k, j_1, \ldots, j_l). \end{aligned}$$

A set of multi-indices $\mathcal{A} \subseteq \mathcal{M}$ is *hierarchical* if it has the properties that

$$\mathcal{A} \neq \emptyset, \quad \sup_{\alpha \in \mathcal{A}} l(\alpha) < \infty \quad \text{and} \quad \alpha \in \mathcal{A} \setminus \{v\} \Rightarrow -\alpha \in \mathcal{A}.$$

The *remainder* of a hierarchical set $\mathcal{A}$ is defined $\mathcal{B}(\mathcal{A}) = \{\alpha \in \mathcal{M} \setminus \mathcal{A} : -\alpha \in \mathcal{A}\}$.

Example 1 The multi-index $\alpha = (j_1 \ldots, j_k)$ generates a hierarchical set $\mathcal{A}_\alpha = \{v, (j_k), \ldots, (j_1, \ldots, j_k)\}$, with remainder

$$\mathcal{B}(\mathcal{A}_\alpha) = \bigcup_{a=1}^{2m} \{(a, j_1, \ldots, j_k)\} \cup \bigcup_{i=1}^{k} \bigcup_{a \neq j_i} \{(a, j_{i+1}, \ldots, j_k)\}.$$

2.3 Multiple Itô integral

Define the *coefficient function* f_α for a sufficiently smooth function f and multi-index $\alpha \in \mathcal{M}$ recursively, by $f_\alpha = \mathcal{L}_{j_1} f_{-\alpha}$ for $\alpha \neq v$ and $f_v = f$.

Example 2 Consider the SDE (1) with $m = d = 1$ and a and b constants. For $\alpha = (0, 1, 2)$, we have, denoting $\tilde{\partial}_x = \partial_x|_{x=\theta_t(X)}$,

$$f_\alpha(t, X_{t-}) = b^2 \Big(\partial_t \partial_x \tilde{\partial}_x - \partial_t \partial_x^2 + a \partial_x^2 \tilde{\partial}_x - a \partial_x^3 + \frac{1}{2} \sigma^2 b^2 (\partial_x^3 - \partial_x^4) \Big) f(t, X_{t-}).$$

Define *multiple Itô integrals* for adapted cáglád processes f and multi-indices $\alpha = (j_1, \ldots, j_k) \in \mathcal{M}$ by

$$I_\alpha[f]_{s,t} = \int_s^t I_{\alpha-}[f]_{s,r} \, d\mathcal{L}_r^{j_{l(\alpha)}} \quad \text{for } \alpha \neq v, \quad I_v[f]_{s,t} = f(t).$$

3 The Stochastic Taylor Formula

To establish the Stochastic Taylor Formula, we need the following lemma, which is a version of the Itô Formula for multiple Itô integrals $I_\alpha[f_\beta(\cdot, X_{(\cdot-)})]_{s,t}$.

Lemma 3 *For multi-indices α and β such that $l(\beta) > 0$, we have*

$$I_\alpha[f_\beta(\cdot, X_{(\cdot-)})]_{s,t} = I_\alpha[f_\beta(s, X_s)]_{s,t} + I_{(j)*\alpha}[f_{(j)*\beta}(\cdot, X_{(\cdot-)})]_{s,t}. \tag{4}$$

Proof. For $l(\alpha) = 0$ we get the lemma from the Itô Formula, as

$$I_v[f_\beta(\cdot, X_{(\cdot-)})]_{s,t} = f_\beta(t, X_t) = I_v[f_\beta(s, X_s)]_{s,t} + I_{(j)*v}[f_{(j)*\beta}(\cdot, X_{(\cdot-)})]_{s,t}.$$

And so we get the lemma by induction, observing that if the lemma is true for $l(\alpha) = k \geq 0$, and taking $\alpha = (j_1, \ldots, j_{k+1})$, we have

$$
\begin{aligned}
I_\alpha[f_\beta(\cdot, X_{(\cdot-)})]_{s,t} &= I_{(j_{k+1})}[I_{\alpha-}[f_\beta(\cdot, X_{(\cdot-)})]_{s,\cdot}]_{s,t} \\
&= I_{(j_{k+1})}\left[I_{\alpha-}[f_\beta(s, X_s)]_{s,\cdot} + I_{(j)*\alpha-}[f_{(j)*\beta}(\cdot, X_{(\cdot-)})]_{s,\cdot}\right]_{s,t} \\
&= I_\alpha[f_\beta(s, X_s)]_{s,t} + I_{(j)*\alpha}[f_{(j)*\beta}(\cdot, X_{(\cdot-)})]_{s,t}. \quad \square
\end{aligned}
$$

Theorem 4 (Stochastic Taylor Formula) *If X satisfies the SDE* (1)*, then we have for hierarchical $\mathcal{A} \subset \mathcal{M}$,*

$$
f(t, X_t) = \sum_{\alpha \in \mathcal{A}} I_\alpha[f_\alpha(s, X_s)]_{s,t} + \sum_{\alpha \in \mathcal{B}(\mathcal{A})} I_\alpha[f_\alpha(\cdot, X_{(\cdot-)})]_{s,t}.
$$

Example 5 (The Itô Formula) Taking $\mathcal{A} = \{v\}$, so that $\mathcal{B}(\mathcal{A}) = \{(0), \ldots, (2m)\}$, the Stochastic Taylor Formula becomes the Itô Formula.

Example 6 (Taylor Formula) Picking $\mathcal{A} = \{\alpha \in \mathcal{M} : l(\alpha) = n(\alpha) \leq n\}$, $a = 1$, $b = 0$ and $d = 1$, so that $\mathcal{B}(\mathcal{A}) = \{\alpha \in \mathcal{M} : l(\alpha) = n(\alpha) = n + 1\}$, since $m = 0$, the Stochastic Taylor Formula becomes a Taylor expansion of f of order n with remainder.

Proof. [Proof of Theorem 4] Let $l_1(\mathcal{A}) = \sup_{\alpha \in \mathcal{A}} l(\alpha)$. For $l_1(\mathcal{A}) = 0$, so that $\mathcal{A} = \{v\}$, the theorem reduces to the Itô Formula (3), by Example 5.

Assume that the theorem holds for hierarchical sets $\mathcal{A}$ with $l_1(\mathcal{A}) = k - 1$, where $k \geq 1$. Take $l_1(\mathcal{A}) = k$. Consider the hierarchical set $\mathcal{E} = \{\alpha \in \mathcal{A} : l(\alpha) \leq k-1\}$ with $l_1(\mathcal{E}) = k-1$, so that (4) applies by the inductive assumption

$$
f(t, X_t) = \sum_{\alpha \in \mathcal{E}} I_\alpha[f_\alpha(s, X_s)]_{s,t} + \sum_{\alpha \in \mathcal{B}(\mathcal{E})} I_\alpha[f_\alpha(\cdot, X_{(\cdot-)})]_{s,t}. \tag{5}
$$

Pick $\alpha \in \mathcal{A} \setminus \mathcal{E}$, so that $l(\alpha) = k$. As $\mathcal{A} \setminus \mathcal{E} \subseteq \mathcal{B}(\mathcal{E})$, we have

$$
\sum_{\alpha \in \mathcal{B}(\mathcal{E})} I_\alpha[f_\alpha(\cdot, X_{(\cdot-)})]_{s,t} = \sum_{\alpha \in \mathcal{A} \setminus \mathcal{E}} I_\alpha[f_\alpha(\cdot, X_{(\cdot-)})]_{s,t} + \sum_{\alpha \in \mathcal{B}(\mathcal{E}) \setminus (\mathcal{A} \setminus \mathcal{E})} I_\alpha[f_\alpha(\cdot, X_{(\cdot-)})]_{s,t}. \tag{6}
$$

Applying Lemma 3 to $I_\alpha[f_\alpha(\cdot, X_{(\cdot-)})]_{s,t}$ for $\alpha \in \mathcal{A} \setminus \mathcal{E}$ in (6), we get

$$I_\alpha[f_\alpha(\cdot, X_{(\cdot-)})]_{s,t} = I_\alpha[f_\alpha(s, X_s)]_{s,t} + I_{(j)*\alpha}[f_{(j)*\alpha}(\cdot, X_{(\cdot-)})]_{s,t}.$$

Inserting this in (6), and using (5), we get

$$\begin{aligned} f(t, X_t) = \sum_{\alpha\in\mathcal{A}} I_\alpha[f_\alpha(s, X_s)]_{s,t} + \sum_{\alpha\in\mathcal{A}\setminus\mathcal{E}} I_{(j)*\alpha}[f_{(j)*\alpha}(\cdot, X_{(\cdot-)})]_{s,t} \\ + \sum_{\alpha\in\mathcal{B}(\mathcal{E})\setminus(\mathcal{A}\setminus\mathcal{E})} I_\alpha[f_\alpha(\cdot, X_{(\cdot-)})]_{s,t}. \end{aligned}$$

This equation can be expressed on the form (4), with

$$f(t, X_t) = \sum_{\alpha\in\mathcal{A}} I_\alpha[f_\alpha(s, X_s)]_{s,t} + \sum_{\alpha\in\mathcal{B}} I_\alpha[f_\alpha(\cdot, X_{(\cdot-)})]_{s,t}. \tag{7}$$

Now the theorem follows from (7) by induction, because $\mathcal{B}$ is given by

$$(\mathcal{B}(\mathcal{E})\setminus(\mathcal{A}\setminus\mathcal{E})) \cup \bigcup_j \{(j)*\alpha : \alpha \in \mathcal{A}\setminus\mathcal{E}\} = (\mathcal{B}(\mathcal{E})\setminus(\mathcal{A}\setminus\mathcal{E})) \cup \mathcal{B}(\mathcal{A}\setminus\mathcal{E}) = \mathcal{B}(\mathcal{A}). \quad \square$$

The Stochastic Taylor Formula carries over to general semimartingales: For example, in one dimension, let X solve an SDE driven by a semimartingale Y

$$dX_t = a(t, X_{t-})\, dt + b(t, X_{t-})\, dY_t.$$

Consider the differential operators $\mathcal{L}_i$ and differentials $d\mathcal{L}_t^i$

$$\mathcal{L}_i = \begin{cases} \partial_t + a\partial_x + \frac{1}{2}b^2\partial_{xx}^2 & \text{for } i = 0 \\ b\partial_x & \text{for } i = 1 \\ \frac{1}{2}b^2\partial_{xx}^2 & \text{for } i = 2 \\ b(\tilde{\partial}_x - \partial_x) & \text{for } i = 3 \end{cases} \quad \text{and} \quad d\mathcal{L}_t^i = \begin{cases} dt & \text{for } i = 0 \\ dY_t & \text{for } i = 1 \\ d([Y,Y]_t^c - t) & \text{for } i = 2 \\ \Delta Y_t & \text{for } i = 3 \end{cases}.$$

Then we have

$$f(t, X_t) = \sum_{\alpha\in\mathcal{A}} f_\alpha(0, x_0) I_\alpha[1]_{0,t} + \sum_{\alpha\in\mathcal{B}(\mathcal{A})} I_\alpha[f_\alpha(\cdot, X_{(\cdot-)})]_{0,t}.$$

With $f(t,x) = x$ this result gives a representation of the solution of the SDE.

For $[Y,Y]^c_t$ linear, as for Lévy processes, we need only consider

$$\mathcal{L}_i = \begin{cases} \partial_t + a\partial_x + \frac{1}{2}b^2\partial^2_{xx} & \text{for } i=0 \\ b\partial_x & \text{for } i=1 \\ b(\tilde{\partial}_x - \partial_x) & \text{for } i=3 \end{cases} \quad \text{and} \quad d\mathcal{L}^i_t = \begin{cases} dt & \text{for } i=0 \\ dY_t & \text{for } i=1 \\ \Delta Y_t & \text{for } i=3 \end{cases}.$$

4 Uniform Mean-square Estimates

The Protter inequality, as we call it, is a version of Protter (2003), Theorem 66: It is a key to find the rates at which the remainder term, from the set $\mathcal{B}(\mathcal{A})$ in the Stochastic Taylor Formula goes to zero. As Protter presents the result with an unspecified constant $K_p > 0$, that we need to have a specified, we reiterate the proof, paying particular attention to the constant.

Theorem 7 (Protter inequality) *For L one dimensional, and a predictable one dimensional process H, we have*

$$\mathbb{E}\Big\{ \sup_{r\in[s,t]} \Big| \int_s^r H_u\, dL_u \Big|^2 \Big\} \leqslant 3\Big(4\mathrm{Var}\{L_1\} + (t-s)(\mathbb{E}\{L_1\})^2\Big) \int_s^t \mathbb{E}\{H_u^2\}\, du.$$

Proof. By the Lévy-Itô decomposition, we have $L_t = \beta t + \sigma W_t + M_t$, where $\beta \in \mathbb{R}$ is a constant, and the martingale M is independent of the Wiener process W. Hence the Hölder and Burkholder-Davis-Gundy inequalities give

$$\begin{aligned} &\frac{1}{3}\,\mathbb{E}\Big\{ \sup_{r\in[s,t]} \Big| \int_s^r H_u\, dL_u \Big|^2 \Big\} \\ &\leq \beta^2 \mathbb{E}\Big\{ \sup_{r\in[s,t]} \Big| \int_s^r H_u\, du \Big|^2 \Big\} + \sigma^2 \mathbb{E}\Big\{ \sup_{r\in[s,t]} \Big| \int_s^r H_u\, dW_u \Big|^2 \Big\} + \mathbb{E}\Big\{ \sup_{r\in[s,t]} \Big| \int_s^r H_u\, dM_u \Big|^2 \Big\} \\ &\leq \beta^2 (t-s) \int_s^t \mathbb{E}\{H_u^2\}\, du + 4\sigma^2 \int_s^t \mathbb{E}\{H_u^2\}\, du + 4\,\mathbb{E}\Big\{ \int_s^t H_u^2\, d[M,M]_u \Big\}. \end{aligned}$$

Here the quadratic variation $[M,M]_t$ is a Lévy process, with $\mathbb{E}\{[M,M]_t\} = t\,\mathbb{E}\{M_1^2\} \le t\,\mathbb{E}\{L_1^2\}$. Thus $X_t = [M,M]_t - \mathbb{E}\{[M,M]_t\}$ is a martingale, so that

$$\begin{aligned}&\mathbb{E}\Big\{\int_s^t H_u^2\, d[M,M]_u\Big\}\\ &= \mathbb{E}\Big\{\int_s^t H_u^2\, dX_u\Big\} + \mathbb{E}\Big\{\int_s^t H_u^2\, d\mathbb{E}\{[M,M]_u\}\Big\} = \mathbb{E}\{M_1^2\}\int_s^t \mathbb{E}\{|H_u|^2\}\, du.\end{aligned}$$

Upon assembling the results of the above computations, we obtain

$$\mathbb{E}\Big\{\sup_{r\in[s,t]}\Big|\int_s^r H_u\, dL_u\Big|^2\Big\} \le 3\Big(\beta^2(t-s) + 4\Big(\sigma^2 + \mathbb{E}\{M_1^2\}\Big)\Big)\int_s^t \mathbb{E}\{H_u^2\}\, du.$$

The theorem now follows since $\mathbb{E}\{L_1\} = \beta$ and $\mathrm{Var}\{L_1\} = \sigma^2 + \mathbb{E}\{M_1^2\}$. □

We use the Protter inequality to bound iterated integrals $I_\alpha[f]_{s,t}$: For $\alpha = (\alpha_1,\dots,\alpha_n) \in \mathcal{M}$, repeated useage of the Protter inequality gives

$$\begin{aligned}\mathbb{E}\Big\{\sup_{r\in[s,t]}\big|I_\alpha[f]_{s,r}\big|^2\Big\} &= \mathbb{E}\Big\{\sup_{r\in[s,t]}\Big|\int_s^r I_{\alpha-}[f]_{s,u}\, d\mathcal{L}_u^{i_n}\Big|^2\Big\}\\ &\le C_{i_n}\int_s^t \mathbb{E}\Big\{|I_{\alpha-}[f]_{s,r}|^2\Big\}\, dr\\ &\le C_{i_n}(t-s)\,\mathbb{E}\Big\{\sup_{r\in[s,t]}|I_{\alpha-}[f]_{s,r}|^2\Big\}\\ &\le \cdots \le C_{i_n}(t-s)\cdot\ldots\cdot C_{i_1}(t-s)\,\mathbb{E}\Big\{\sup_{r\in[s,t]}|f|^2\Big\}\\ &\le (\max_{i_k>0} C_{i_k})^{l(\alpha)-n(\alpha)}(t-s)^{l(\alpha)+n(\alpha)}\,\mathbb{E}\Big\{\sup_{r\in[s,t]}|f|^2\Big\},\end{aligned}$$

as $C_0 = t-s$. And so we arrive at

Theorem 8 (Uniform mean-square-estimate) *There exists a constant $K > 0$ such that, for $\alpha \in \mathcal{M}\setminus\{v\}$,*

$$\mathbb{E}\Big\{\sup_{r\in[s,t]}\big|I_\alpha[f(\cdot)]_{s,r}\big|^2\Big\} \le K^{l(\alpha)-n(\alpha)}(t-s)^{l(\alpha)+n(\alpha)}\,\mathbb{E}\Big\{\sup_{r\in[s,t]}|f(r)|^2\Big\}. \qquad (8)$$

5 Strong Convergence of the Stochastic Taylor Formula

From now on we consider the one dimensional SDE (1) with $d = 1$.

Theorem 9 (Strong convergence) *Consider the hierarchical sets*

$$\mathcal{A}_k = \{\alpha \in \mathcal{M} : l(\alpha) + n(\alpha) \leq k\} \cup \{\alpha \in \mathcal{M} : l(\alpha) = n(\alpha) = (k+1)/2\} \quad (9)$$

for $k \in \mathbb{N}$. Take $f(t,x) = x$ and consider the stochastic Taylor scheme

$$X_r^k = \sum_{\alpha \in \mathcal{A}_k} I_\alpha[f_\alpha(s, X_s)]_{s,r}.$$

Let the SDE (1) with $d = 1$ be such that

$$\mathbb{E}\left\{ \sup_{r\in[s,t]} |f_\alpha(r, X_r)|^2 \right\} \leq C_1 C_2^{l(\alpha)+n(\alpha)} \quad \text{for } \alpha \in \mathcal{M},$$

for some constants $C_1, C_2 > 0$. We have

$$\mathbb{E}\left\{ \sup_{r\in[s,t]} |X_r - X_r^k|^2 \right\} \leq C_3 (t-s)^k C_4^k \quad \text{for } t - s \leq 1, \quad (10)$$

for some constants $C_3, C_4 > 0$. In particular,

$$\mathbb{P}\left\{ \lim_{k\to\infty} \sup_{r\in[s,t]} \left| X_r^k - X_r \right| = 0 \right\} = 1. \quad (11)$$

Proof. By the Stochastic Taylor Formula, and writing N_k for the number of elements in $\mathcal{B}(\mathcal{A}_k)$, we have

$$\begin{aligned}
\mathbb{E}\left\{ \sup_{r\in[s,t]} |X_r - X_r^k|^2 \right\} &= \mathbb{E}\left\{ \sup_{r\in[s,t]} \left| \sum_{\alpha\in\mathcal{B}(\mathcal{A}_k)} I_\alpha[f_\alpha(\cdot, X_{(\cdot-)})]_{s,r} \right|^2 \right\} \\
&\leq N_k \sum_{\alpha\in\mathcal{B}(\mathcal{A}_k)} \mathbb{E}\left\{ \sup_{r\in[s,t]} |I_\alpha[f_\alpha(\cdot, X_{(\cdot-)})]_{s,r}|^2 \right\} \\
&\leq N_k \sum_{\alpha\in\mathcal{B}(\mathcal{A}_k)} K^{l(\alpha)-n(\alpha)} (t-s)^{l(\alpha)+n(\alpha)} C_1 C_2^{l(\alpha)+n(\alpha)}.
\end{aligned} \quad (12)$$

Now we relate $l(\alpha)$ and $n(\alpha)$ to k for $\alpha \in \mathcal{B}(\mathcal{A}_k)$: Let $l(\alpha^+)$ denote the number of non-zero components of α, so that $l(\alpha) + n(\alpha) = l(\alpha^+) + 2n(\alpha)$. Pick an $\alpha \in \mathcal{A}_k$ and add a component j, thus forming the multi-index $(j) * \alpha$.

If $j = 0$, then $n((j) * \alpha) = 1 + n(\alpha)$ and $l(((j) * \alpha)^+) = l(\alpha^+)$, so that $l((j)*\alpha)+n((j)*\alpha) = l(\alpha^+)+2n(\alpha)+2$. If $j \geq 1$, then $n((j)*\alpha) = n(\alpha)$ and $l(((j) * \alpha)^+) = 1 + l(\alpha^+)$, so that $l((j) * \alpha) + n((j) * \alpha) = l(\alpha^+) + 2n(\alpha) + 1$.

By the previous paragraph, $l(\alpha) + n(\alpha) \in \{k + 1, k + 2\}$ for $\alpha \in \mathcal{B}(\mathcal{A}_k)$. Further, the number of elements of $\mathcal{B}(\{\alpha \in \mathcal{M} : l(\alpha)+n(\alpha) \leq k\})$ is at most $(2m + 1)^k$, see Kloeden and Platen (1992), p. 207, which readily gives N_k $\leq (2m + 1)^k + (2m + 1) \leq 2(2m + 1)^{k+1}$. Inserting this in (12), we get (10).

Clearly, it is sufficient to prove (11) for $t-s$ sufficiently small. However, picking an $\varepsilon > 0$ and using the Chebychev inequality together with (10), we get

$$\sum_{k\in\mathbb{N}} \mathbb{P}\left\{ \sup_{r\in[s,t]} |X_r - X_r^k| > \varepsilon \right\} \leq \frac{1}{\varepsilon^2} \sum_{k\in\mathbb{N}} \mathbb{E}\left\{ \sup_{r\in[s,t]} |X_r - X_r^k|^2 \right\} \leq \sum_{k\in\mathbb{N}} \frac{C_3(t-s)^k C_4^k}{\varepsilon^2} < \infty$$

for $t - s$ small enough. And so (11) follows from the Borel-Cantelli lemma. □

The approximations of the solution of an SDE take the form

$$X_k(t) = \sum_{\alpha\in\mathcal{A}_k} I_\alpha[f_\alpha(s, X_s)]_{s,t}. \tag{13}$$

Since $I_\alpha[g]$ is linear in g, we get

$$I_\alpha[f_\alpha(s, X_s)]_{s,t} = f_\alpha(s, X_s) \int_s^t I_{\alpha-}[1]_{s,s_{l(\alpha)}} \, d\mathcal{L}_{s_{l(\alpha)}}^{j_{l(\alpha)}} = \ldots = f_\alpha(s, X_s) I_\alpha^L(s, t),$$

where

$$I_\alpha^L(s, t) = I_\alpha[1]_{s,t} = \int_{s_1=s}^t \ldots \int_{s_{l(\alpha)}=s}^{s_{l(\alpha)-1}} d\mathcal{L}_{s_1}^{j_1} \ldots d\mathcal{L}_{s_{l(\alpha)}}^{j_{l(\alpha)}}.$$

And so we may write the approximation (13)

$$X_k(t) = X_s + \sum_{\alpha \in \mathcal{A}_k \setminus \{v\}} f_\alpha(s, X_s) I_\alpha^L(s,t). \tag{14}$$

6 Approximations by the Stochastic Taylor Formula

Take a uniform partition $s = \tau_0 < \cdots < \tau_{n_t} \leq t$ of $[s,t]$, with $\tau_{j+1} - \tau_j = h \leq 1$ and $\max(t - \tau_{n_t}, \sup_j\{\tau_j - \tau_{j-1}\}) \leq h$, where $n_t = \sup\{k \in \mathbb{N} : \tau_k \leqslant t\}$. Write

$$X_t = X_s + (X_{\tau_1} - X_s) + \cdots + (X_{\tau_{n_t}} - X_{\tau_{n_t-1}}) + (X_t - X_{\tau_{n_t}}). \tag{15}$$

The Stochastic Taylor Formula can be applied to (15) [cf. (14)], to give

$$\begin{aligned} X_t = X_s &+ \sum_{\alpha \in \mathcal{A}_k \setminus \{v\}} \left(f_\alpha(\tau_{n_t}, X_{\tau_{n_t}}) I_\alpha[1]_{\tau_{n_t},t} + \sum_{j=0}^{n_t-1} f_\alpha(\tau_j, X_{\tau_j}) I_\alpha[1]_{\tau_j,\tau_{j+1}} \right) \\ &+ \sum_{\alpha \in \mathcal{B}(\mathcal{A}_k)} \left(I_\alpha[f_\alpha(\cdot, X_{(\cdot_)})]_{\tau_{n_t},t} + \sum_{j=0}^{n_t-1} I_\alpha[f_\alpha(\cdot, X_{(\cdot_)})]_{\tau_j,\tau_{j+1}} \right). \end{aligned}$$

Let the functions $f_\alpha(t,x)$ based on $f(t,x) = x$ be globally Lipschitz in x, with a Lipschitz constant ℓ_1 that is bounded for $\alpha \in \mathcal{A}_k$ and t, and linear growth in x, with a linear growth constant ℓ_2 that is bounded for $\alpha \in \mathcal{A}_k$ and t.

The *stochastic Taylor approximation* $Y^k = \{Y_t^k\}_{t \geq 0}$ interpolates the values $X_{\tau_j}^k$ of the discrete approximations $\{X_{\tau_j}^k\}_{j=0}^{n_t}$, and are given by

$$X_{\tau_{j+1}}^k = X_{\tau_j}^k + \sum_{\alpha \in \mathcal{A}_k \setminus \{v\}} f_\alpha(\tau_j, X_{\tau_j}^k) I_\alpha[1]_{\tau_j,\tau_{j+1}}, \tag{16}$$

$$Y_r^k = X_{\tau_{n_r}}^k + \sum_{\alpha \in \mathcal{A}_k \setminus \{v\}} f_\alpha(\tau_{n_r}, X_{\tau_{n_r}}^k) I_\alpha[1]_{\tau_{n_r},r}.$$

Since $n_{\tau_j} = j$, we have $Y_{\tau_j}^k = X_{\tau_j}^k$. Now represent Y_r^k as

$$Y_r^k = Y_{\tau_0}^k + \sum_{j=0}^{n_r-1} (Y_{\tau_{j+1}}^k - Y_{\tau_j}^k) + (Y_r^k - Y_{\tau_{n_r}}^k) = X_{\tau_0}^k + \sum_{j=0}^{n_r-1} (X_{\tau_{j+1}}^k - X_{\tau_j}^k) + (Y_r^k - X_{\tau_{n_r}}^k). \tag{17}$$

Next we study the difference $X_r - Y_r^k$ for $r \in [s,t]$, using (17):

$$\begin{aligned} X_r - Y_r^k = X_s - X_s^k &+ \sum_{\alpha \in \mathcal{A}_k \setminus \{v\}} \Big(f_\alpha(\tau_{n_r}, X_{\tau_{n_r}}) - f_\alpha(\tau_{n_r}, X^k_{\tau_{n_r}}) \Big) I_\alpha[1]_{\tau_{n_r}, r} \\ &+ \sum_{\alpha \in \mathcal{A}_k \setminus \{v\}} \left(\sum_{j=0}^{n_r - 1} \Big(f_\alpha(\tau_j, X_{\tau_j}) - f_\alpha(\tau_j, X^k_{\tau_j}) \Big) I_\alpha[1]_{\tau_j, \tau_{j+1}} \right) \\ &+ \sum_{\alpha \in \mathcal{B}(\mathcal{A}_k)} \left(I_\alpha[f_\alpha(\cdot, X_{(\cdot -)})]_{\tau_{n_r}, r} + \sum_{j=0}^{n_r - 1} I_\alpha[f_\alpha(\cdot, X_{(\cdot -)})]_{\tau_j, \tau_{j+1}} \right). \end{aligned}$$

Let $\mathcal{A}_k \setminus \{v\}$ and $\mathcal{B}(\mathcal{A}_k)$ have M_k and N_k elements, respectively. We have

$$\mathbb{E}\Big\{ \sup_{r \in [s,t]} \big| X_r - Y_r^k \big|^2 \Big\} \le 3\, \mathbb{E}\Big\{ g |X_s - Y_s^k|^2 \Big\} + 3 M_k \sum_{\alpha \in \mathcal{A}_k \setminus \{v\}} F_\alpha + 3 N_k \sum_{\alpha \in \mathcal{B}(\mathcal{A}_k)} G_\alpha, \tag{18}$$

where

$$\begin{aligned} F_\alpha &= \mathbb{E}\Big\{ \sup_{r \in [s,t]} \Big| \Big(f_\alpha(\tau_{n_r}, X_{\tau_{n_r}}) - f_\alpha(\tau_{n_r}, X^k_{\tau_{n_r}}) \Big) I_\alpha[1]_{\tau_{n_r}, r} \\ &\qquad + \sum_{j=0}^{n_r - 1} \Big(f_\alpha(\tau_j, X_{\tau_j}) - f_\alpha(\tau_j, X^k_{\tau_j}) \Big) I_\alpha[1]_{\tau_j, \tau_{j+1}} \Big|^2 \Big\}, \\ G_\alpha &= \mathbb{E}\Big\{ \sup_{r \in [s,t]} \Big| I_\alpha[f_\alpha(\cdot, X_{(\cdot -)})]_{\tau_{n_r}, r} + \sum_{j=0}^{n_r - 1} I_\alpha[f_\alpha(\cdot, X_{(\cdot -)})]_{\tau_j, \tau_{j+1}} \Big|^2 \Big\}. \end{aligned}$$

In order to estimate F_α and G_α we use the following lemma:

Lemma 10 *For an adapted cáglád process g with $\mathbb{E}\Big\{ \sup_{r \in [s,t]} |g(r)|^2 \Big\} < \infty$, writing $\varphi(\alpha) = l(\alpha -) + n(\alpha -)$, there exist $K_\alpha > 0$ such that*

$$\mathbb{E}\Big\{ \sup_{r \in [s,t]} \Big| I_\alpha[g(\cdot)]_{\tau_{n_r}, r} + \sum_{j=0}^{n_r - 1} I_\alpha[g(\cdot)]_{\tau_j, \tau_{j+1}} \Big|^2 \Big\} \le K_\alpha h^{\varphi(\alpha)} \quad \text{for } \alpha \in \mathcal{M} \setminus \{v\}.$$

Proof. The proof has three parts: To take off, let $\alpha = (i_1, \ldots, i_n)$ and denote

$$S_r^\alpha = \sum_{j=0}^{n_r - 1} I_\alpha[g(\cdot)]_{\tau_j, \tau_{j+1}} + I_\alpha[g(\cdot)]_{\tau_{n_r}, r}.$$

Part 1. Assume that $n(\alpha) = l(\alpha)$. As $\tau_{n_u} = \tau_j$ for $u \in [\tau_j, \tau_{j+1})$, we have

$$\begin{aligned}\int_s^r I_{\alpha-}[g(\cdot)]_{\tau_{n_u},u}\,du &= \sum_{j=1}^{n_r}\int_{\tau_{j-1}}^{\tau_j} I_{\alpha-}[g(\cdot)]_{\tau_{n_u},u}\,du + \int_{\tau_{n_r}}^{r} I_{\alpha-}[g(\cdot)]_{\tau_{n_u},u}\,du \\ &= \sum_{j=1}^{n_r}\int_{\tau_{j-1}}^{\tau_j} I_{\alpha-}[g(\cdot)]_{\tau_{j-1},u}\,du + \int_{\tau_{n_r}}^{r} I_{\alpha-}[g(\cdot)]_{\tau_{n_r},u}\,du = S_r^\alpha.\end{aligned} \tag{19}$$

We use the Cauchy-Schwartz inequality and the estimate (8), to obtain

$$\begin{aligned}\mathbb{E}\Big\{\sup_{r\in[s,t]}\left|S_r^\alpha\right|^2\Big\} &= \mathbb{E}\Big\{\sup_{r\in[s,t]}\left|\int_s^r I_{\alpha-}[g(\cdot)]_{\tau_{n_u},u}\,du\right|^2\Big\} \\ &\le \mathbb{E}\Big\{\sup_{r\in[s,t]}(r-s)\int_s^r\left|I_{\alpha-}[g(\cdot)]_{\tau_{n_u},u}\right|^2 du\Big\} \\ &\le (t-s)\int_s^t \mathbb{E}\Big\{\Big|\sup_{v\in[\tau_u,u]} I_{\alpha-}[g(\cdot)]_{\tau_{n_u},v}\Big|^2\Big\}\,du \\ &\le (t-s)^2 h^{l(\alpha-)+n(\alpha-)}K^{l(\alpha-)-n(\alpha-)}\,\mathbb{E}\Big\{\sup_{r\in[s,t]}|g(r)|^2\Big\}.\end{aligned}$$

Part 2. Assume that $n(\alpha) < l(\alpha)$ and $i_n = 0$. We still have (19). By the Cauchy-Schwartz inequality and the estimate (8), we therefore get

$$\begin{aligned}\mathbb{E}\Big\{\sup_{r\in[s,t]}|S_r^\alpha|^2\Big\} &\le (t-s)\,\mathbb{E}\Big\{\sup_{r\in[s,t]}|I_{\alpha-}[g(\cdot)]_{\tau_{n_r},r}|^2\Big\} \\ &\le (t-s)^2 h^{l(\alpha-)+n(\alpha-)}K^{l(\alpha-)-n(\alpha-)}\,\mathbb{E}\Big\{\sup_{r\in[s,t]}|g(r)|^2\Big\}.\end{aligned}$$

Part 3. Assume that $n(\alpha) < l(\alpha)$ and $i_n \neq 0$. By the same reasoning as was used for (19), we may represent $S^\alpha = \int_s^r I_{\alpha-}[g(\cdot)]_{\tau_{n_u},u}\,d\mathcal{L}_u^{i_n}$, where $\mathcal{L}^{i_n}$ is a Lévy process given by

$$\mathcal{L}_u^{i_n} = \begin{cases} L_u^{i_n} & \text{for } i_n \in \{1,\dots,m\}, \\ J_u^{i_n-m} & \text{for } i_n \in \{m+1,\dots,2m\}. \end{cases}$$

And so the Protter inequality and the estimate (8) of an iterated integral give

$$\begin{aligned}\mathbb{E}\Big\{\sup_{r\in[s,t]}\left|S_r^\alpha\right|^2\Big\} &\le C_{i_n}\int_s^t \mathbb{E}\Big\{|I_{\alpha-}[g(\cdot)]_{\tau_{n_u},u}|^2\Big\}\,du \\ &\le C_{i_n}\int_s^t \mathbb{E}\Big\{\sup_{v\in[\tau_{n_u},u]}|I_{\alpha-}[g(\cdot)]_{\tau_{n_u},v}|^2\Big\}\,du \\ &\le h^{l(\alpha-)+n(\alpha-)}C_{i_n}K^{l(\alpha-)-n(\alpha-)}\int_s^t \mathbb{E}\Big\{\sup_{r\in[\tau_{n_u},u]}|g(r)|^2\Big\}\,du. \quad \square\end{aligned}$$

If we apply Lemma 10 to the inequality (18), we get

$$\begin{aligned}
&\mathbb{E}\Big\{ \sup_{r\in[s,t]} |X_r - Y_r^k|^2 \Big\} - 3\,\mathbb{E}\Big\{ |X_s - Y_s^k|^2 \Big\} \\
&\quad \le 3M_k \sum_{\alpha\in\mathcal{A}_k\setminus\{v\}} K_\alpha h^{\varphi(\alpha)} \int_s^t \mathbb{E}\Big\{ \sup_{r\in[s,u]} \Big| f_\alpha(\tau_{n_r}, X_{\tau_{n_r}}) - f_\alpha(\tau_{n_r}, X^k_{\tau_{n_r}}) \Big|^2 \Big\} du \\
&\qquad + 3N_k \sum_{\alpha\in\mathcal{B}(\mathcal{A}_k)} K_\alpha h^{\varphi(\alpha)} \int_s^t \mathbb{E}\Big\{ \sup_{r\in[s,u]} |f_\alpha(r, X_r)|^2 \Big\} du \\
&\quad \le 3M_k \sum_{\alpha\in\mathcal{A}_k\setminus\{v\}} K_\alpha h^{\varphi(\alpha)} \ell_1^2 \int_s^t \mathbb{E}\Big\{ \sup_{r\in[s,u]} |X_r - X_r^k|^2 \Big\} du \\
&\qquad + 3N_k \sum_{\alpha\in\mathcal{B}(\mathcal{A}_k)} K_\alpha h^{\varphi(\alpha)} \ell_2^2 \int_s^t \mathbb{E}\Big\{ \sup_{r\in[s,u]} |1 + |X_r||^2 \Big\} du.
\end{aligned}$$

As $\sup_r |X_r - X_r^k| \le \sup_r |X_r - Y_r^k|$, we arrive at the following inequality

$$\mathbb{E}\Big\{ \sup_{r\in[s,t]} |X_r - Y_r^k|^2 \Big\} \le 3\mathbb{E}\{ |X_s - Y_s^k|^2 \} + Ah^{\varphi(\alpha)} + Bh^{\varphi(\alpha)} \int_s^t \mathbb{E}\Big\{ \sup_{r\in[s,u]} |X_r - Y_r^k|^2 \Big\} du,$$

where $A, B > 0$ are constants. And so the Grönwall inequality (see e.g., Kloeden and Platen, 1992, Lemma 4.5.1), gives

$$\mathbb{E}\Big\{ \sup_{r\in[s,t]} |X_r - Y_r^k|^2 \Big\} \le \Big(3\mathbb{E}\{ |X_s - Y_s^k|^2 \} + Ah^{\varphi(\alpha)} \Big) e^{Bh^{\varphi(\alpha)}(t-s)}.$$

We have to investigate the function φ on $\mathcal{B}(\mathcal{A}_k)$: For $\alpha \in \mathcal{B}(\mathcal{A}_k)$, we have

$$l(\alpha)+n(\alpha) \ge k+1 \quad \text{and} \quad \begin{cases} 2l(\alpha) \ne k+1 \\ \text{or} \\ 2n(\alpha) \ne k+1 \end{cases} \quad \text{and} \quad \begin{cases} l(-\alpha)+n(-\alpha) \le k \\ \text{or} \\ 2l(-\alpha) = 2n(-\alpha) = k+1 \end{cases}.$$

From this we may deduce that $\varphi(\alpha) \in \{k, k+1\}$ for $i_1 = 0$ and $\varphi(\alpha) = k+1$ $i_1 \ne 0$. Since $h \le 1$ we get $h^{\varphi(\alpha)} \le h^k$ for $\alpha \in \mathcal{B}(\mathcal{A}_k)$. And so we arrive at

Theorem 11 (Order of Convergence) *With the notation of this section, there exists a constant $C > 0$ such that*

$$\mathbb{E}\Big\{ \sup_{r\in[s,t]} |X_r - Y_r^k|^2 \Big\} \le 3\mathbb{E}\Big\{ |X_s - Y_s^k|^2 \Big\} + Ch^k.$$

7 Stochastic Taylor Schemes, $k \leq 2$

Consider the SDE (1) with $d = m = 1$. We have $\mathcal{A}_1 = \{v, (0), (1), (2)\}$ and $\mathcal{A}_2 = \{v, (0), (1), (2), (1,1), (1,2), (2,1), (2,2)\}$. With $f(t,x) = x$ we have $f_{(0)} = a$, $f_{(1)} = b$, $f_{(2)} = f_{(1,2)} = f_{(2,2)} = 0$, $f_{(1,1)} = b\partial_x b$ and

$$f_{(2,1)}(t, X_{t-}) = b(t, X_{t-})\Big(\frac{\partial b}{\partial x}(t,x)\Big|_{x=\Theta_t(X)} - \frac{\partial b}{\partial x}(t, X_{t^-})\Big).$$

Moreover, $I_v[1]_{\tau_j,\tau_{j+1}} = 1$, $I_{(0)}[1]_{\tau_j,\tau_{j+1}} = h$, $I_{(1)}[1]_{\tau_j,\tau_{j+1}} = L_{\tau_{j+1}} - L_{\tau_j}$, and

$$\begin{aligned} I_{(1,1)}[1]_{\tau_j,\tau_{j+1}} &= \frac{1}{2}\Big((L_{\tau_{j+1}} - L_{\tau_j})^2 - ([L,L]_{\tau_{j+1}} - [L,L]_{\tau_j})\Big), \\ I_{(2,1)}[1]_{\tau_j,\tau_{j+1}} &= \int_{\tau_j}^{\tau_{j+1}} \sum_{\tau_j < s \leqslant u} \Delta L_s \, dL_u. \end{aligned}$$

7.1 The Euler-Maruyama scheme, $k = 1$

For the hierarchical set $\mathcal{A}_1$, the stochastic Taylor scheme (16) becomes

$$X^1_{\tau_{j+1}} = X^1_{\tau_j} + a(\tau_j, X^1_{\tau_j})(\tau_{j+1} - \tau_j) + b(\tau_j, X^1_{\tau_j})(L_{\tau_{j+1}} - L_{\tau_j}).$$

This scheme has order of convergence 1/2, i.e., $\mathbb{E}\{|X_T - X^1_T|^2\} \leq Ch^{1/2}$.

7.2 The Milstein scheme, $k = 2$

For the hierarchical set $\mathcal{A}_2$ the stochastic Taylor scheme (16) becomes

$$\begin{aligned} X^2_{\tau_{j+1}} = X^2_{\tau_j} &+ a(\tau_j, X^2_{\tau_j})(\tau_{j+1} - \tau_j) + b(\tau_j, X^2_{\tau_j})(L_{\tau_{j+1}} - L_{\tau_j}) \\ &+ \frac{1}{2} b(\tau_j, X^2_{\tau_j}) \frac{\partial b}{\partial x}(\tau_j, X^2_{\tau_j})\Big((L_{\tau_{j+1}} - L_{\tau_j})^2 - ([L,L]_{\tau_{j+1}} - [L,L]_{\tau_j})\Big) \\ &+ b(\tau_j, X^2_{\tau_j})\Big(\frac{\partial b}{\partial x}(\tau_j, x)\Big|_{x=\Theta_{\tau_j}(X^2)} - \frac{\partial b}{\partial x}(\tau_j, X^2_{\tau_j})\Big) \int_{\tau_j}^{\tau_{j+1}} \sum_{\tau_j < s \leq u} \Delta L_s dL_u. \end{aligned} \tag{20}$$

This scheme has order of convergence 1, i.e., $\mathbb{E}\{|X_T - X_T^2|^2\} \leq Ch$.

The Milstein scheme is complicated by the last term on the right in (20). However, as $\Delta X^2_{\tau_j} = 0$, because X^2 does not jump in fixed points, being stochastically continuous, the last term is zero, and the Milstein scheme takes a form formally identical to that of an SDE driven by a Wiener process:

$$X^2_{\tau_{j+1}} = X^2_{\tau_j} + a(\tau_j, X^2_{\tau_j})(\tau_{j+1} - \tau_j) + b(\tau_j, X^2_{\tau_j})(L_{\tau_{j+1}} - L_{\tau_j}) + \frac{1}{2} b(\tau_j, X^2_{\tau_j}) \frac{\partial b}{\partial x}(\tau_j, X^2_{\tau_j}) \Big((L_{\tau_{j+1}} - L_{\tau_j})^2 - ([L,L]_{\tau_{j+1}} - [L,L]_{\tau_j}) \Big).$$

7.3 The Kailath-Segall identity

Let $\alpha_k = (j_1, \ldots, j_k)$, where $j_1 = \ldots = j_k \in \{1, 2\}$. By Segall and Kailath (1976), the iterated integrals $I^L_{\alpha_k}(s,t)$ can be computed recursively, as

$$kI^L_{\alpha_k}(s,t) = I^L_{\alpha_{k-1}}(s,t)(L_t - L_s) - I^L_{\alpha_{k-2}}(s,t)([L,L]_t - [L,L]_s) + \\ + \sum_{3 \leq n \leq k} (-1)^{n-1} I^L_{\alpha_{k-n}}(s,t) \sum_{s < r \leq t} (\Delta L_r)^n.$$

Example 12 By the Kailath-Segall identity, we can compute $I^L_{\alpha_3}(s,t)$ as

$$3I^L_{\alpha_3}(s,t) = I^L_{\alpha_2}(s,t)(L_t - L_s) - I^L_{\alpha_1}(s,t)([L,L]_t - [L,L]_s) + I^L_{\alpha_0}(s,t) \sum_{s<r\leq t} \Delta L_r^3 \\ = \frac{1}{2}(L_t - L_s)^3 - \frac{3}{2}(L_t - L_s)([L,L]_t - [L,L]_s) + \sum_{s<r\leq t} (\Delta L_r)^3.$$

And so we have

$$\sum_{\alpha \in \mathcal{A}_3 \setminus \mathcal{A}_2} f_\alpha(s, X_s) I^L_\alpha(s,t) \\ = b^2 \gamma X_s (L_t - L_s)(t - s) \\ + b^3 X_s \left(\frac{1}{6}(L_t - L_s)^3 - \frac{1}{2}(L_t - L_s)([L,L]_t - [L,L]_s) + \frac{1}{3} \sum_{s<r\leq t} (\Delta L_s)^3 \right).$$

References

[1] D. Applebaum, *Lévy Processes and Stochastic Calculus*, Cambridge University Press, Cambridge, 2004.

[2] J. Bertoin, *Lévy Processes*, Cambridge University Press, Cambridge, 1996.

[3] P.E. Kloeden, E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer, Berlin, 1992.

[4] P.E. Protter, *Stochastic Integration and Differential Equations*, 2nd Ed, Springer, Berlin, 2003.

[5] A. Segall, T. Kailath, Orthogonal functions of independent-increment processes, IEEE Trans. Information Theory IT-22 (1976), pp. 287–298.

PAPER IV

A NOTE ON THE LIPSCHITZ CONDITION FOR EULER APPROXIMATION OF STOCHASTIC DIFFERENTIAL EQUATIONS*

ANDERS MUSZTA†, MARTIN RICHTER‡, J. M. P. ALBIN§, AND BJARNE ASTRUP JENSEN¶

Abstract. We show how the common global Lipschitz condition can be relaxed to local Lipschitz, near finite boundaries, for the Euler approximation of homogeneous stochastic differential equations with linear growth.

Key words. Euler method, Lipschitz condition, local Lipschitz condition, non-linear stochastic differential equation, stochastic differential equation

AMS subject classifications. 60H10, 60H35, 60J60, 65C30, 68U20

1. Introduction. For an open interval $I = (l, r)$, and measurable functions $a : I \to \mathbb{R}$ and $b : I \to [0, \infty)$, consider a one dimensional time homogeneous stochastic differential equation (`sde`)

$$dX_t = a(X_t)dt + b(X_t)dW_t \quad \text{for } t \in [0, T], \quad X_0 = x, \tag{1.1}$$

driven by a standard Brownian motion $\{W_t\}_{t\geq 0}$, where $x \in I$ is a constant.

Assume that X does not explode, and thus stays within the interval I for finite times, necessary and sufficients conditions for which are given by Feller's test for explosion (see e.g., Karatzas and Shreve, 1991).

Consider a uniform partition $0 = t_0 < \ldots < t_n = T$ of $[0, T]$ with $t_k - t_{k-1} = \Delta > 0$ for $k = 1, \ldots, n$. Writing $\Delta W_k = W(t_k) - W(t_{k-1})$, the Euler scheme for numerical approximation of the `sde` (1.1) is given by

$$Y_k = Y_{k-1} + a(Y_{k-1})\Delta + b(Y_{k-1})\Delta W_k \quad \text{for } k = 1, \ldots, n. \tag{1.2}$$

The common condition in the literature for uniform mean-square convergence

$$\lim_{\Delta \downarrow 0} \mathbf{E}\Big\{ \sup_{0\leq k\leq n} |X_{t_k} - Y_k|^2 \Big\} = 0, \tag{1.3}$$

is that the coefficients a and b satisfy global Lipschitz conditions on I.

Very little is done in the literature without the global Lipschitz condition. However, Gyöngy (1998) shows almost sure uniform convergence

$$\lim_{\Delta \downarrow 0} \mathbf{P}\Big\{ \sup_{0\leq k\leq n} |X_{t_k} - Y_k| = 0 \Big\} = 1,$$

instead of in mean-square, assuming only a one-sided Lipschitz condition for the drift a and a locally Lipschitz volatility b. Despite this, in practice, it is possible to have `sde` with a very ill-behaved Euler scheme, for practical sizes of Δ, under these conditions.

*Recived March 2005.

†Department of Mathematics, Chalmers University of Technology, 412 96 Göteborg, Sweden. (muszta@math.chalmers.se).

‡Department of Finance, Copenhagen Business School, 2000 Frederiksberg, Denmark. (mr.fi@cbs.dk).

§Department of Mathematics, Chalmers University of Technology, 412 96 Göteborg, Sweden. (palbin@math.chalmers.se), supported by The Swedish Research Council grant no. 621-2003-5214.

¶Department of Finance, Copenhagen Business School, 2000 Frederiksberg, Denmark. (ba.fi@cbs.dk).

Higham, Mao and Stuart (2002), Theorem 2.2, makes the assuption that

$$\sup_{\Delta>0} \mathbf{E}\Big\{ \sup_{0\le k\le n} |Y_k|^p \Big\} < \infty \quad \text{and} \quad \mathbf{E}\Big\{ \sup_{0\le t\le T} |X_t|^p \Big\} < \infty \tag{1.4}$$

for some $p > 2$, with which they establish (1.3) only requiring that a and b are locally Lipschitz. However, to show (1.4) in turn, they have to require that the drift a is one-sided Lipschitz, and the volatility b globally Lipschitz.

It should be noted that for so called stochastic differential delay equations, local Lipschitz conditions, together with linear growth, is enough for convergence of Euler schemes, see Mao and Sabanis (2003).

In this note we will show that local Lipschitz conditions and linear growth are sufficient for the convergence (1.3) for the `sde` (1.1), when I is finite. If I is half-finite, then the global Lipschitz condition is required at the infinite boundary, while local Lipschitz is enough at the finite boundary.

2. Relaxing the Lipschitz condition. The next theorem is our main result:

THEOREM 2.1. *Assume that the coefficients a and b of the* `sde` (1.1) *are globally Lipshitz on any closed sub-interval of I if I is finite, and globally Lipshitz on any closed half-finite sub-interval of I if I is half-finite. Further, assume that a and b have linear growth on I, and that the* `sde` *does not explode. The Euler scheme* (1.2) *is uniformly mean-square convergent* (1.3).

Proof. We only prove the theorm for $I = (0, \infty)$, as it will be obvious from that proof how to deal with any other case of the hypothesis.

By Higham, Mao and Stuart (2002), Theorem 2.2, it is enough to prove (1.4). But replacing (1.1) with a non-exploding SDE on I with global Lipschitz drift $\tilde{a}$ and volatility $\tilde{b}$, that coincide with a and b on (ε, ∞), where $\varepsilon \in (0, x)$, and denoting the solution to the new SDE $\tilde{X}$ and $\tau_\varepsilon = \inf\{t > 0 : X_t = \varepsilon\}$, we have

$$\begin{aligned}
&\mathbf{E}^x\Big\{ \sup_{t\in[0,T]} X(t)^p \Big\} \\
&\le \mathbf{E}^x\Big\{ \sup_{t\in[0,\tau_\varepsilon]} \tilde{X}(t)^p \mathbf{1}_{\{\tau_\varepsilon<T\}} \Big\} + \mathbf{E}^x\Big\{ \sup_{t\in[0,T]} X(t)^p \Big\} \mathbf{P}^x\{\tau_\varepsilon<T\} + \mathbf{E}^x\Big\{ \sup_{t\in[0,T]} \tilde{X}(t)^p \mathbf{1}_{\{\tau_\varepsilon\ge T\}} \Big\} \\
&\le \mathbf{E}^x\Big\{ \sup_{t\in[0,T]} \tilde{X}(t)^p \Big\} + \mathbf{E}^x\Big\{ \sup_{t\in[0,T]} X(t)^p \Big\} \mathbf{P}^x\{\tau_\varepsilon<T\}.
\end{aligned} \tag{2.1}$$

Since $\mathbf{P}\{\tau_\varepsilon < T\} < 1$, we get the moment bound (1.4) on X from such a bound on $\tilde{X}$. But such a bound, for any $p \ge 2$, is given e.g., in Kloeden and Platen (1992, Exercise 4.5.5).

As for the Euler scheme, we continue with the above modified diffusion $\tilde{X}$, with Euler scheme $\tilde{Y}_k$, which is then further extended to a diffusion $\hat{X}$ on $\mathbb{R}$, with Euler scheme $\hat{Y}_k$, by extending $\tilde{a}$ and $\tilde{b}$ to drifts and volatilities $\hat{a}$ and $\hat{b}$ on $\mathbb{R}$, by making them symmetric $\hat{a}(x) = \tilde{a}(|x|)$ and $\hat{b}(x) = \tilde{b}(|x|)$ for $x \in \mathbb{R}$. (Albeit $\tilde{X}$ will not take on negative values, $\tilde{Y}_k$ might do that!) By a trivial modification of the argument (2.1), together with the symmetry, we have

$$\mathbf{E}^x\Big\{ \sup_{0\le k\le n} Y_k^p \Big\} \le \mathbf{E}^x\Big\{ \sup_{0\le k\le n} |\hat{Y}_k|^p \Big\} + \mathbf{E}^x\Big\{ \sup_{0\le k\le n} Y_k^p \Big\} \mathbf{P}^x\{\tau_\varepsilon < T\}.$$

And so we get (1.4) for Y_k from such a bound on $\hat{Y}_k$.

Taking $\Delta \le 1$ in the sequel, by linear growth, we readily obtain

$$\hat{Y}_k^2 \le \hat{Y}_{k-1}^2 + 2\hat{b}(\hat{Y}_{k-1})\hat{Y}_{k-1}\Delta W_k + C_1\big(1+\hat{Y}_{k-1}^2\big)\big(\Delta + \Delta W_k^2\big). \tag{2.2}$$

For the sequence

$$\hat{Z}_k^2 = \hat{Z}_{k-1}^2 + 2\hat{b}(\hat{Y}_{k-1})\hat{Y}_{k-1}\Delta W_k + C_1\big(1+\hat{Z}_{k-1}^2\big)\big(\Delta + \Delta W_k^2\big), \quad \hat{Z}_0^2 = x^2, \tag{2.3}$$

we therefore have $\hat{Z}_k^2 \ge \hat{Y}_k^2$, by induction. Because this inequality holds trivially for $k=1$, and if it holds for $k=\ell \ge 1$, then we have, by (2.2),

$$\begin{aligned}\hat{Z}_{\ell+1}^2 &= \hat{Z}_\ell^2 + 2\hat{b}(\hat{Y}_\ell)\hat{Y}_\ell\Delta W_{\ell+1} + C_1\big(1+\hat{Z}_\ell^2\big)\big(\Delta+\Delta W_{\ell+1}^2\big)\\ &\ge \hat{Y}_\ell^2 + 2\hat{b}(Y_\ell)\hat{Y}_\ell\Delta W_{\ell+1} + C_1\big(1+\hat{Y}_\ell^2\big)\big(\Delta+\Delta W_{\ell+1}^2\big) \ge \hat{Y}_{\ell+1}^2.\end{aligned}$$

Notice that, for some constant $C_2 > 0$,

$$\hat{Z}_k^4 \le \hat{Z}_{k-1}^4 + 4\hat{b}(\hat{Y}_{k-1})\hat{Y}_{k-1}\hat{Z}_{k-1}^2\Delta W_k + C_2\big(1+\hat{Z}_{k-1}^4\big)\big(\Delta + \Delta W_k^2 + \Delta W_k^4\big), \tag{2.4}$$

and, more generally, for some constant $C_n > 0$,

$$\hat{Z}_k^{2^n} \le \hat{Z}_{k-1}^{2^n} + 2^n\hat{b}(\hat{Y}_{k-1})\hat{Y}_{k-1}\hat{Z}_{k-1}^{2^n-2}\Delta W_k + C_n\big(1+\hat{Z}_{k-1}^{2^n}\big)\big(\Delta+\Delta W_k^2+\Delta W_k^{2^n}\big). \tag{2.5}$$

From (2.5), using that $\mathbf{E}\{\hat{Z}_0^{2^n}\} = \mathbf{E}\{\hat{Y}_0^{2^n}\} = x^{2^n}$ is finite, we get that $\mathbf{E}\{\hat{Z}_1^{2^{n-1}}\} \ge \mathbf{E}\{\hat{Y}_1^{2^{n-1}}\}$ is finite, which gives that $\mathbf{E}\{\hat{Z}_2^{2^{n-2}}\} \ge \mathbf{E}\{\hat{Y}_1^{2^{n-2}}\}$ is finite, and so one. And so the sequences $\hat{Z}_k^2$ and $\hat{Y}_k^2$ have all moments.

As $\hat{Z}_k^2$ have all moments, it is a submartingale, by inspection of (2.3). Further, from (2.4) we get, for some constant $K > 0$,

$$\mathbf{E}\{\hat{Z}_k^4\} \le \mathbf{E}\{\hat{Z}_{k-1}^4\} + K\big(1+\mathbf{E}\{\hat{Z}_{k-1}^4\}\big)\Delta. \tag{2.6}$$

Hence, by induction, we have

$$\mathbf{E}\{\hat{Z}_k^4\} \le (1+x^4)(1+K\Delta)^k - 1, \tag{2.7}$$

as (2.7) holds trivially for $k=1$, and if (2.7) holds for $k=\ell\ge 1$, then (2.6) gives

$$\mathbf{E}\{\hat{Z}_{\ell+1}^4\} \le (1+x^4)(1+K\Delta)^\ell - 1 + K\Delta(1+x^4)(1+K\Delta)^\ell = (1+x^4)(1+K\Delta)^{\ell+1} - 1.$$

By (2.7) and Doob's maximal inequality (see e.g., Karatzas and Shreve, 1991), we have

$$\sup_{\Delta>0}\mathbf{E}\Big\{\sup_{0\le k\le n}\hat{Y}_k^4\Big\} \le \sup_{\Delta>0}\mathbf{E}\Big\{\sup_{0\le k\le n}\hat{Z}_k^4\Big\} \le 4\sup_{\Delta>0}\mathbf{E}\{\hat{Z}_n^4\} \le 4\sup_{\Delta>0}(1+x^4)(1+K\Delta)^{1/\Delta},$$

where the right hand side is finite: As we have shown (1.4), the proof is complete. □

3. An example. As an example for which Theorem 2.1 is useful, we may consider the `sde` on the interval $I=(0,\infty)$

$$dX_t = \alpha X_t^a\,dt + \beta X_t^b dW_t \quad \text{for } t\in[0,T], \quad X_0 = x, \tag{3.1}$$

where $\alpha,\beta>0$ and $a,b\in(0,1)$ are constants, and $x\in I$.

Clearly, the coefficients of the `sde` (3.1) are globally Lipschitz on any closed half-finite sub-interval of I, and have linear growth, but are not globally Lipschitz.

To ensure that the `sde` (3.1) does not explode, and has a unique strong solution, it is necessary and sufficient that $2b > a+1$.

REFERENCES

[1] I. Gyöngy, *A note on Euler approximations*, Potential Anal., 8 (1998), pp. 205–216.
[2] D. J. Higham, X. Mao and A. Stuart, *Strong convergence of numerical methods for nonlinear stochastic differential equations*, SIAM J. Numer. Anal., 40 (2002) pp. 1041–1063.
[3] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, Springer, Berlin, 1991.
[4] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer, Berlin, 1992.
[5] X. Mao and S. Sabanis, *Numerical solution of stochastic differential delay equations under local Lipshitz condition*, J. Comput. Appl. Math., 151 (2003), pp. 215–227.