NUMERICAL METHODS FOR STOCHASTIC DIFFERENTIAL EQUATIONS

Part I: A brief introduction

Part II: Classical methods and results

Part III: Optimal L_2-approximation

Part IV: Optimal approximation at a single point

Part V: Outlook

Thomas Müller-Gronbach, Uni Magdeburg
METU, September 2004

I. A brief introduction

Problem:
Approximation of the solution X of a (scalar) sde

$$(1) \quad dX(t) = a(t, X(t))\,dt + \sigma(t, X(t))\,dW(t), \quad t \in [0,1],$$

with

initial condition	$X(0) = x_0 \in \mathbb{R}$
drift coefficient	$a : [0,1] \times \mathbb{R} \to \mathbb{R}$
diffusion coefficient	$\sigma : [0,1] \times \mathbb{R} \to \mathbb{R}$
(driving) Brownian motion	W (on $(\Omega, \mathcal{A}, P)$)

In this lecture:
Strong (pathwise) approximation of X
i.e.,
Construction of a stochastic process $\widehat{X}$ whose trajectories are close (in a certain sense) to the trajectories of the solution X.

How to interpret (1) ?

Example 1. (population dynamics, finance)

$$dX(t) = \mu \cdot X(t)\, dt + b \cdot X(t)\, dW(t), \quad X(0) = 1.$$

Here: $a(t,x) = \mu \cdot x, \quad \sigma(t,x) = b \cdot x.$

Discrete version:

$$0 = t_0 < t_1 < \ldots < t_n = 1$$

$$\frac{X(t_{\ell+1}) - X(t_\ell)}{X(t_\ell)} = \mu \cdot (t_{\ell+1} - t_\ell) + b \cdot (W(t_{\ell+1}) - W(t_\ell))$$

Average relative increase over time interval $[t_\ell, t_{\ell+1}]$ is proportional to length $t_{\ell+1} - t_\ell$:

$$E\left(\frac{X(t_{\ell+1}) - X(t_\ell)}{X(t_\ell)}\right) = \mu \cdot (t_{\ell+1} - t_\ell)$$

Random perturbation due to, e.g., random change of environment. Modelled by independent noise terms

$$b \cdot (W(t_{\ell+1}) - W(t_\ell)) \sim \mathrm{N}\big(0, b^2 \cdot (t_{\ell+1} - t_\ell)\big)$$

Idealization:

$$\max_\ell (t_{\ell+1} - t_\ell) \to 0$$

Pathwise interpretation of (1) as an ode?

X is a stochastic process with differentiable paths $X(\cdot, \omega),\ \omega \in \Omega$, that satisfy

$$\frac{d}{dt} X(t,\omega) = a(t, X(t,\omega)) + \sigma(t, X(t,\omega)) \cdot \underbrace{\frac{d}{dt} W(t,\omega)}_{?}$$

No! With probability 1, the paths of W are nowhere differentiable on $[0,1]$.

Reformulation of (1) as an integral equation:

$$(2) \quad X(t) = x_0 + \int_0^t a(s, X(s))\, ds + \underbrace{\int_0^t \sigma(s, X(s))\, dW(s)}_{(*)}$$

for $t \in [0,1]$.

Pathwise interpretation of $(*)$ as Riemann-Stieltjes integral?

Problem: With probability 1, the paths of W are of unbounded variation on $[0,1]$.

The way out:

For suitable processes Y define

$$\int_0^t Y(s)\, dW(s)$$

in the mean square sense, and use formulation (2).

The Itô-integral

Let $0 \le s \le t \le 1$ and consider a sequence of discretizations

(3) $$s = t_0^{(n)} < t_1^{(n)} < \ldots < t_n^{(n)} = t$$

with

(4) $$\lim_{n\to\infty} \max_{0\le \ell \le n-1} (t_{\ell+1}^{(n)} - t_\ell^{(n)}) = 0$$

For a NICE process $Y = (Y(u))_{0\le u\le 1}$ define

(5) $$\int_s^t Y(u)\, dW(u)$$

$$:= \operatorname*{l.i.m.}_{n\to\infty} \sum_{\ell=0}^{n-1} Y(t_\ell^{(n)}) \cdot \big(W(t_{\ell+1}^{(n)}) - W(t_\ell^{(n)})\big)$$

(l.i.m.: limit in $L_2(\Omega, \mathcal{A}, P)$)

Remark 1. Formally, NICE means:

(i) Y is adapted (to the standard filtration associated with W)

(ii) Y is measurable

(iii) $\int_0^1 E(Y^2(t))\, dt < \infty$

(iv) Y has continuous paths a.s.

In this lecture: All integrands are NICE

Example 2.

$$\int_s^t W(u)\, dW(u) = 1/2 \cdot \big(W^2(t) - W^2(s) - (t-s)\big)$$

Proof. Consider a sequence of discretizations (3) that satisfies (4). Then

$$S_n := \sum_\ell W(t_\ell^{(n)}) \cdot \big(W(t_{\ell+1}^{(n)}) - W(t_\ell^{(n)})\big)$$
$$= 1/2 \cdot \Big(W^2(t) - W^2(s) - \sum_\ell \big(W(t_{\ell+1}^{(n)}) - W(t_\ell^{(n)})\big)^2\Big).$$

Clearly,

$$E\Big(\sum_\ell \big(W(t_{\ell+1}^{(n)}) - W(t_\ell^{(n)})\big)^2\Big) = t - s.$$

Thus

$$E\Big(\Big(\sum_\ell \big(W(t_{\ell+1}^{(n)}) - W(t_\ell^{(n)})\big)^2 - (t-s)\Big)^2\Big)$$
$$= Var\Big(\sum_\ell \big(W(t_{\ell+1}^{(n)}) - W(t_\ell^{(n)})\big)^2\Big)$$
$$\le \sum_\ell E\Big(\big(W(t_{\ell+1}^{(n)}) - W(t_\ell^{(n)})\big)^4\Big)$$
$$= 3 \cdot \sum_\ell \big(t_{\ell+1}^{(n)} - t_\ell^{(n)}\big)^2$$
$$\le 3 \cdot (t-s) \cdot \max_{0\le \ell\le n-1} (t_{\ell+1}^{(n)} - t_\ell^{(n)}).$$

□

Remark 2. The common rules of Riemann-Stieltjes integration do not hold for the Itô-integral, see Example 2. This is a consequence of the choice of the evaluation points for the process Y in the definition (5).

Different (systematic) choices lead to different stochastic integrals, e.g., the *Stratonovich-integral*

$$\int_s^t Y(u) \circ dW(u)$$

$$:= \operatorname*{l.i.m.}_{n\to\infty} \sum_{\ell=0}^{n-1} Y\big((t_\ell^{(n)} + t_{\ell+1}^{(n)})/2\big) \cdot \big(W(t_{\ell+1}^{(n)}) - W(t_\ell^{(n)})\big),$$

which obeys the classical transformation rules but lacks of other nice properties of the Itô-integral.

Properties of the Itô-integral

(i) Linearity

(ii) For $s \le v \le t$:

$$\int_s^v Y(u)\, dW(u) + \int_v^t Y(u)\, dW(u) = \int_s^t Y(u)\, dW(u)$$

(iii) $E\left(\int_s^t Y(u)\, dW(u)\right) = 0$

(iv) For $s_1 \le s_2 \le t_1 \le t_2$:

$$E\left(\int_{s_1}^{s_2} Y(u)\, dW(u) \cdot \int_{t_1}^{t_2} Y(u)\, dW(u)\right) = 0$$

(v) Itô-isometry:

$$E\left(\left(\int_s^t Y(u)\, dW(u)\right)^2\right) = \int_s^t E(Y^2(u))\, du$$

Furthermore, the process

$$\mathcal{J} = \left(\int_0^t Y(u)\, dW(u)\right)_{0\le t\le 1}$$

satisfies:

(vi) $\mathcal{J}$ is adapted.

(vii) $\mathcal{J}$ is a martingale:

$$E\big(\mathcal{J}(t) \big| W(u), u \le s\big) = \mathcal{J}(s)$$

(viii) $\mathcal{J}$ has a continuous modification $\widetilde{\mathcal{J}}$, i.e.,

$\widetilde{\mathcal{J}}$ has continuous paths and $\forall t:\ \mathcal{J}(t) = \widetilde{\mathcal{J}}(t)$ a.s.

Itô's formula

Theorem 1. *For $f \in C^{1,2}([0,1] \times \mathbb{R})$ and $t \in [0,1]$:*

$$\begin{aligned} f(t, W(t) &= f(0,0) + \int_0^t f^{(0,1)}(s, W(s))\ dW(s) \\ &+ \int_0^t f^{(1,0)}(s, W(s))\ ds \\ &+ 1/2 \cdot \int_0^t f^{(0,2)}(s, W(s))\ ds \end{aligned}$$

Corollary 1. *For $g \in C^1([0,1])$ and $t \in [0,1]$:*

$$\int_0^t g(s)\ dW(s) = g(t) \cdot W(t) - \int_0^t g'(s) \cdot W(s)\ ds$$

Proof. Use Itô's formula with $f(t,x) = g(t) \cdot x$. □

Example 3.

Solving the sde from Example 1

$$dX(t) = \mu \cdot X(t)\ dt + b \cdot X(t)\ dW(t), \quad X(0) = 1.$$

Consider

$$f(t,x) = \exp\big((\mu - b^2/2) \cdot t + b \cdot x\big).$$

Geometric Brownian motion:

$$X(t) := f(t, W(t) = \exp\big((\mu - b^2/2) \cdot t + b \cdot W(t)\big)$$

Note that

$$f^{(1,0)} = (\mu - b^2/2) \cdot f, \quad f^{(0,1)} = b \cdot f, \quad f^{(0,2)} = b^2 \cdot f.$$

Thus, by Itô's formula,

$$\begin{aligned} X(t) &= \underbrace{f(0,0)}_{=1} + \int_0^t b \cdot X(s)\ dW(s) \\ &\quad + \int_0^t (\mu - b^2/2) \cdot X(s)\ ds + 1/2 \cdot \int_0^t b^2 \cdot X(s)\ ds \\ &= 1 + \int_0^t \mu \cdot X(s)\ ds + \int_0^t b \cdot X(s)\ dW(s). \end{aligned}$$

Existence and uniqueness of a solution of (1)

Regularity conditions for a function $f : [0,1] \times \mathbb{R} \to \mathbb{R}$:

(L_f) $\quad \exists K > 0\ \forall x, y \in \mathbb{R}\ \forall t \in [0,1]$:

$$|f(t,x) - f(t,y)| \le K \cdot |x - y|$$

(LG_f) $\quad \exists K > 0\ \forall x \in \mathbb{R}\ \forall t \in [0,1]$:

$$|f(t,x)| \le K \cdot (1 + |x|)$$

(LLG_f) $\exists K > 0\ \forall x \in \mathbb{R}\ \forall s,t \in [0,1]$:

$$|f(s,x) - f(t,x)| \le K \cdot (1 + |x|) \cdot |s - t|$$

Note: $(L_f) \wedge (LLG_f) \Rightarrow (LG_f) \wedge f$ is continuous

Theorem 2. *Assume that $f = a$ and $f = \sigma$ (are continuous and) satisfy (L_f) as well as (LG_f).*
Then there exists a continuous adapted process X that satisfies (2). Moreover, if the process $\widetilde{X}$ has the same properties then

$$P\big(\forall t: \ X(t) = \widetilde{X}(t)\big) = 1.$$

Idea of *proof of existence*:

Consider the Picard-Lindelöf-iteration

$$X^{(0)} = x_0,$$

$$X^{(n)}(t) = x_0 + \int_0^t a\big(s, X^{(n-1)}(s)\big)\, ds + \int_0^t \sigma\big(s, X^{(n-1)}(s)\big)\, dW(s).$$

With probability 1 the paths $X^{(n)}(\cdot,\omega)$ converge uniformly on $[0,1]$. The limiting process is adapted and satisfies (2).

Properties of the solution

Under the assumptions of Theorem 2:

(i) The trajectories of X up to time t are essentially determined by the corresponding trajectories of W up to time t, i.e.,

$$(X(s))_{0 \le s \le t} = g\big((W(s))_{0 \le s \le t}\big) \quad \text{a.s.}$$

for some measurable function $g : C([0,t]) \to C([0,t])$.

(ii) X is Hölder continuous of order $1/2$ in p-th mean sense, i.e.,

$\forall p \in [1, \infty[\ \exists c > 0\ \forall s,t \in [0,1]$:

$$\big(E|X(s) - X(t)|^p\big)^{1/p} \le c \cdot |s - t|^{1/2}. \tag{6}$$

(iii) X is a Markov-process. For $s \le t$:

$$P^{X(t)|(X(u))_{0 \le u \le s}} = P^{X(t)|X(s)}$$

Moreover, for $P^{X(s)}$ almost all $x \in \mathbb{R}$:

$$P^{X(t)|X(s)=x} = P^{X_{s,x}(t)},$$

where $X_{s,x}$ is the solution of

$$dX_{s,x}(t) = a(t, X_{s,x})\, dt + \sigma(t, X_{s,x})\, dW(t), \quad t \in [s,1],$$

with initial value $X_{s,x}(s) = x$.

(iv) X is a diffusion process. In particular,

$$E(X(t+\delta) - X(t)|X(t) = x) = a(t,x) \cdot \delta + o(\delta)$$

and

$$E\big((X(t+\delta) - X(t))^2 \big| X(t) = x\big) = \sigma^2(t,x) \cdot \delta + o(\delta). \tag{7}$$

Property (7) is crucial for the analysis of L_2-approximation, see Part III. The smoothness of X in mean square sense is locally in time and space determined by:

conditional Hölder constant $|\sigma(t, X(t))|$.

Example 4. For the trivial equation

$$dX(t) = dW(t), \quad X(0) = 0,$$

we have

$$\sigma(t, X(t) = 1.$$

For the linear equation

$$dX(t) = \mu \cdot X(t)\; dt + b \cdot X(t)\; dW(t), \quad X(0) = 1.$$

we have

$$\sigma(t, X(t) = b \cdot X(t) = b \cdot \exp\big((\mu - b^2/2) \cdot t + b \cdot W(t)\big),$$

see Example 3.

The problem of strong approximation

Given data about a, σ and W construct a stochastic process $\widehat{X}$ that is pathwise close to the solution X of (1).

Available data:

(i) evaluation of a, σ and/or partial derivatives of these functions at suitably chosen points in $[0,1] \times \mathbb{R}$, e.g.,

$$a(t,x),\ \sigma(t,x),\ a^{(0,1)}(t,x),\ \sigma^{(2,0)}(t,x).$$

In practice: by subroutines

(ii) evaluation of suitably chosen functionals of W; in particular

(a) Dirac-functionals: $W(t)$,

(b) Continuous linear functionals: $\int_s^t W(u)\; du$,

(c) Iterated Itô-integrals: $\int_s^t W(u)\; dW(u)$.

In practice: (a), (b) by random number generator;
(c) is a problem.

Discrete approximation:

$\widehat{X}$ pathwise close to X at finitely many points in $[0,1]$, e.g., at $t=1$ with

error pathwise: $|X(1) - \widehat{X}(1)|$

average: $\big(E|X(1) - \widehat{X}(1)|^q\big)^{1/q}, \quad q \in [1, \infty[.$

Global approximation:

$\widehat{X}$ pathwise close to X globally on $[0,1]$ with

error pathwise:

$$\|X-\widehat{X}\|_p = \begin{cases} \left(\int_0^1 |X(t)-\widehat{X}(t)|^p\,dt\right)^{1/p} & , p\in[1,\infty[\\ \sup_{t\in[0,1]} |X(t)-\widehat{X}(t)| & , p=\infty. \end{cases}$$

average: $\left(E\|X-\widehat{X}\|_p^q\right)^{1/q}, \quad q\in[1,\infty[.$

Remark 3. Closely related problems:

- Reconstruction of a stochastic process X
- Estimation of a weighted integral of X

based on data about X itself. See Ritter (2000), Remark 10 in Part III and Remark 14 in Part IV.

Note: For approximation of the solution X of (1) only data about the driving Brownian motion W is available.

II. Classical methods and results

Classical methods are based on a fixed discretization

$$0=t_0<t_1<\ldots<t_n=1.$$

Recursive one-step construction of a scheme

$$x_0=\widehat{X}(t_0), \widehat{X}(t_1), \ldots, \widehat{X}(t_n).$$

Discrete approximation at $t=1$: $\widehat{X}(1)$.

Global approximation by (piecewise linear) interpolation.

The Euler scheme (Maruyama 1955)

Linearization of integrals:

$$X(t_{\ell+1}) = X(t_\ell) + \underbrace{\int_{t_\ell}^{t_{\ell+1}} a(t,X(t))\,dt}_{(*)} + \underbrace{\int_{t_\ell}^{t_{\ell+1}} \sigma(t,X(t))\,dW(t)}_{(**)}$$

$$\begin{array}{lll} (*) & \text{approx} & a(t_\ell,X(t_\ell))\cdot(t_{\ell+1}-t_\ell), \\ (**) & \text{approx} & \sigma(t_\ell,X(t_\ell))\cdot(W(t_{\ell+1})-W(t_\ell)). \end{array}$$

Scheme:

$$\widehat{X}^E(t_0) = x_0,$$
$$\widehat{X}^E(t_{\ell+1}) = \widehat{X}^E(t_\ell) + a(t_\ell, \widehat{X}^E(t_\ell)) \cdot (t_{\ell+1} - t_\ell) + \sigma(t_\ell, \widehat{X}^E(t_\ell)) \cdot (W(t_{\ell+1}) - W(t_\ell)).$$

Example 5.

Trivial equation:

$$dX(t) = dW(t), \quad X(0) = 0$$
$$X = W$$
$$\widehat{X}^E(t_\ell) = W(t_\ell)$$

Geometric Brownian motion:

$$dX(t) = \mu \cdot X(t)\, dt + b \cdot X(t)\, dW(t), \quad X(0) = 1$$
$$X(t) = \exp\big((\mu - b^2/2) \cdot t + b \cdot W(t)\big)$$
$$a(t,x) = \mu \cdot x$$
$$\sigma(t,x) = b \cdot x$$

Here

$$\widehat{X}^E(t_{\ell+1}) = \widehat{X}^E(t_\ell) + \mu \cdot \widehat{X}^E(t_\ell) \cdot (t_{\ell+1} - t_\ell) + b \cdot \widehat{X}^E(t_\ell) \cdot (W(t_{\ell+1}) - W(t_\ell))\big)$$

and

$$\widehat{X}^E(t_\ell) = \prod_{j=0}^{\ell-1} \big(1 + \mu \cdot (t_{j+1} - t_j) + b \cdot (W(t_{j+1}) - W(t_j))\big).$$

Notation:

$$\Delta_{\max} = \max_{\ell=0,\dots,n-1} (t_{\ell+1} - t_\ell)$$

c: unspecified positive constants that only depend on error parameters p, q, on initial value x_0, and on constants from regularity conditions on a and σ

Theorem 3. *Assume* (L_f) *and* (LLG_f) *for* $f = a$ *and* $f = \sigma$.

(i) *(Maruyama 1955)*

$$\max_\ell \big(E|X(t_\ell) - \widehat{X}^E(t_\ell)|^q\big)^{1/q} \le c \cdot \Delta_{\max}^{1/2}$$

(ii) *For the piecewise linear interpolated Euler scheme*

$$\big(E\|X - \widehat{X}^E\|_p^q\big)^{1/q} \le c \cdot \begin{cases} \Delta_{\max}^{1/2} & \text{if } p < \infty \\ \Delta_{\max}^{1/2} \cdot \big(\ln \Delta_{\max}^{-1}\big)^{1/2} & \text{if } p = \infty \\ \text{(Fauré 1992)} & \end{cases}$$

Proof of the first inequality in Theorem 3(ii). Define

$$X^{lin}(t) = \frac{t_{\ell+1} - t}{t_{\ell+1} - t_\ell} \cdot X(t_\ell) + \frac{t - t_\ell}{t_{\ell+1} - t_\ell} \cdot X(t_{\ell+1}), \quad t \in [t_{\ell+1}, t_\ell].$$

By property (6) and part (i) of Theorem 3

$$\big(E\|X - X^{lin}\|_p^q\big)^{1/q} \le c \cdot \Delta_{\max}^{1/2},$$
$$\big(E\|X^{lin} - \widehat{X}^E\|_p^q\big)^{1/q} \le c \cdot \max_\ell \big(E|X(t_\ell) - \widehat{X}^E(t_\ell)|^q\big)^{1/q} \le c \cdot \Delta_{\max}^{1/2}.$$

□

The Milstein scheme (Milstein 1974)

Motivation: Assumptions in Theorem 3 on a and σ imply

$$E\left(\int_{t_\ell}^{t_{\ell+1}} a(s,X(s))\,ds - a(t_\ell,X(t_\ell))\cdot(t_{\ell+1}-t_\ell)\right)^2 \le c\cdot(t_{\ell+1}-t_\ell)^3$$

but only

$$E\left(\int_{t_\ell}^{t_{\ell+1}} \sigma(s,X(s))\,dW(s) - \sigma(t_\ell,X(t_\ell))\cdot\big(W(t_{\ell+1})-W(t_\ell)\big)\right)^2 \le c\cdot(t_{\ell+1}-t_\ell)^2$$

A better approximation of the Itô-integral is needed.

Idea: Use first order Taylor-expansion of $\sigma(t,\cdot)$ and ignore $O\big((t_{\ell+1}-t_\ell)^3\big)$-terms (in mean square sense).

Under sufficient regularity of a and σ for $t\in[t_\ell,t_{\ell+1}]$

$$\begin{aligned}
\sigma(t,X(t)) &\overset{\text{m.s.}}{=} \sigma(t_\ell,X(t_\ell)) + \sigma^{(0,1)}(t_\ell,X(t_\ell))\cdot(X(t)-X(t_\ell)) \\
&\quad + O\big((t_{\ell+1}-t_\ell)^2\big) \\
&\overset{\text{m.s.}}{=} \sigma(t_\ell,X(t_\ell)) + \sigma^{(0,1)}(t_\ell,X(t_\ell))\cdot\int_{t_\ell}^{t}\sigma(s,X(s))\,dW(s) \\
&\quad + O\big((t_{\ell+1}-t_\ell)^2\big) \\
&\overset{\text{m.s.}}{=} \sigma(t_\ell,X(t_\ell)) + \big(\sigma^{(0,1)}\cdot\sigma\big)(t_\ell,X(t_\ell))\cdot\big(W(t)-W(t_\ell)\big) \\
&\quad + O\big((t_{\ell+1}-t_\ell)^2\big),
\end{aligned}$$

which yields

$$\begin{aligned}
&\int_{t_\ell}^{t_{\ell+1}} \sigma(t,X(t))\,dW(t) \\
&\quad\overset{\text{m.s.}}{=} \sigma(t_\ell,X(t_\ell))\cdot\big(W(t_{\ell+1})-W(t_\ell)\big) \\
&\qquad + \big(\sigma^{(0,1)}\cdot\sigma\big)(t_\ell,X(t_\ell))\cdot\int_{t_\ell}^{t_{\ell+1}}\big(W(t)-W(t_\ell)\big)\,dW(t) \\
&\qquad + O\big((t_{\ell+1}-t_\ell)^3\big).
\end{aligned}$$

Recall from Example 2:

$$\int_{t_\ell}^{t_{\ell+1}}\big(W(t)-W(t_\ell)\big)\,dW(t) = 1/2\cdot\big((W(t_{\ell+1})-W(t_\ell))^2-(t_{\ell+1}-t_\ell)\big)$$

Scheme:

$$\begin{aligned}
\widehat{X}^M(t_0) &= x_0, \\
\widehat{X}^M(t_{\ell+1}) &= \widehat{X}^M(t_\ell) + a(t_\ell,\widehat{X}^M(t_\ell))\cdot(t_{\ell+1}-t_\ell) \\
&\quad + \sigma(t_\ell,\widehat{X}^M(t_\ell))\cdot(W(t_{\ell+1})-W(t_\ell)) \\
&\quad + 1/2\cdot\big(\sigma^{(0,1)}\cdot\sigma\big)(t_\ell,\widehat{X}^M(t_\ell)) \\
&\qquad \times\big((W(t_{\ell+1})-W(t_\ell))^2-(t_{\ell+1}-t_\ell)\big).
\end{aligned}$$

Example 6. Geometric Brownian motion

$$\begin{aligned} dX(t) &= \mu \cdot X(t)\, dt + b \cdot X(t)\, dW(t), \quad X(0) = 1 \\ X(t) &= \exp\big((\mu - b^2/2) \cdot t + b \cdot W(t)\big) \end{aligned}$$

Here: $\sigma^{(0,1)} = b$

Hence

$$\begin{aligned} \widehat{X}^M(t_{\ell+1}) = \widehat{X}^M(t_\ell) &+ \mu \cdot \widehat{X}^M(t_\ell) \cdot (t_{\ell+1} - t_\ell) \\ &+ b \cdot \widehat{X}^M(t_\ell) \cdot (W(t_{\ell+1}) - W(t_\ell))\big) \\ &+ b^2/2 \cdot \widehat{X}^M(t_\ell) \\ &\times \big((W(t_{\ell+1}) - W(t_\ell))^2 - (t_{\ell+1} - t_\ell)\big) \end{aligned}$$

and

$$\begin{aligned} \widehat{X}^M(t_\ell) = \prod_{j=0}^{\ell-1} \Big(1 &+ (\mu - b^2/2) \cdot (t_{j+1} - t_j) \\ &+ b \cdot (W(t_{j+1}) - W(t_j)) \\ &+ b^2/2 \cdot (W(t_{j+1}) - W(t_j))^2\Big). \end{aligned}$$

Theorem 4. *Assume* (L_f), (LLG_f) *and* $(L_{f^{(0,1)}})$ *for* $f = a$ *and* $f = \sigma$.

(i) *(Milstein 1974)*

$$\max_\ell \big(E|X(t_\ell) - \widehat{X}^M(t_\ell)|^q\big)^{1/q} \le c \cdot \Delta_{\max}$$

(ii) *For the piecewise linear interpolated Milstein scheme*

$$\begin{aligned} &\big(E\|X - \widehat{X}^M\|_p^q\big)^{1/q} \\ &\le c \cdot \begin{cases} \Delta_{\max}^{1/2} & \textit{if } p < \infty \\ \Delta_{\max}^{1/2} \cdot \big(\ln \Delta_{\max}^{-1}\big)^{1/2} & \textit{if } p = \infty \end{cases} \end{aligned}$$

Proof of Theorem 4(ii). The first inequality is straightforward from property (6) and part (i); c.f. the proof of the first inequality in Theorem 3(ii). The second inequality follows from Theorem 3 and Theorem 4(i) since

$$\begin{aligned} \big(E\|X - \widehat{X}^M\|_\infty^q\big)^{1/q} \le \big(E\|X - \widehat{X}^E\|_\infty^q\big)^{1/q} &+ \max_\ell \big(E|\widehat{X}^E(t_\ell) - X(t_\ell)|^q\big)^{1/q} \\ &+ \max_\ell \big(E|X(t_\ell) - \widehat{X}^M(t_\ell)|^q\big)^{1/q}. \quad \square \end{aligned}$$

See ?? for a proof of Theorem 4(i).

Can we do better?

Discrete:	NO	if only point evaluations of W are used see Part IV
	YES	if continuous linear functionals of W or iterated Itô-integrals are used, see next section
Global:	NO	see Part III and Part V

Itô-Taylor schemes (Kloeden and Platen 1995)

Iterated Itô-integrals:

Notation $\quad d_0 t := dt,\ d_1 t := dW(t)$

Let $\quad \mathcal{M} = \bigcup_{r\in\mathbb{N}} \{0,1\}^r$

For $\alpha = (\alpha_1, \ldots, \alpha_r) \in \mathcal{M}$ and $0 \le s \le t \le 1$ define

$$I_{\alpha,s,t} = \int_s^t \int_s^{t_r} \cdots \int_s^{t_2} 1\ d_{\alpha_1} t_1 \cdots d_{\alpha_{r-1}} t_{r-1}\, d_{\alpha_r} t_r.$$

Example 7.

$$I_{(0),s,t} = \int_s^t 1\ du = t - s$$

$$I_{(1),s,t} = \int_s^t 1\ dW(u) = W(t) - W(s)$$

$$I_{(1,0),s,t} = \int_s^t I_{(1),s,u}\ du = \int_s^t (W(u) - W(s))du$$

(continuous linear functional)

$$I_{(1,0,1),s,t} = \int_s^t \left(\int_s^v (W(u) - W(s))\ du \right)\ dW(v)$$

(nonlinear functional)

Note: In general, iterated Itô-integrals are not Gaussian.

Itô-Taylor scheme of order $\gamma \in \mathbb{N}/2$:

$$\widehat{X}^\gamma(t_0) = x_0$$

$$\widehat{X}^\gamma(t_{\ell+1}) = \widehat{X}^\gamma(t_\ell) + \sum_{\alpha\in A_\gamma} f_\alpha\big(t_\ell, \widehat{X}^\gamma(t_\ell)\big) \cdot I_{\alpha,t_\ell,t_{\ell+1}}$$

with certain hierarchical sets of multiindices $A_\gamma \subset \mathcal{M}$ and certain coefficient functions

$$f_\alpha : [0,1] \times \mathbb{R} \to \mathbb{R}$$

composed of a, σ (and partial derivatives).

(Background: Truncated Itô-Taylor expansion of solution X)

Coefficient functions:

Consider the differential operators

$$L^0 = \frac{\partial}{\partial t} + a \cdot \frac{\partial}{\partial x} + \sigma^2/2 \cdot \frac{\partial^2}{\partial x^2}, \quad L^1 = \sigma \cdot \frac{\partial}{\partial x}$$

For $\alpha = (\alpha_1, \ldots, \alpha_r) \in \mathcal{M}$ define

$$f_\alpha = \begin{cases} a & \text{if } \alpha = (0) \\ \sigma & \text{if } \alpha = (1) \\ L^{\alpha_1} f_{(\alpha_2,\ldots,\alpha_r)} & \text{otherwise} \end{cases}$$

Example 8.

$$f_{(0,1)} = \sigma^{(1,0)} + a \cdot \sigma^{(0,1)} + \sigma^2/2 \cdot \sigma^{(0,2)}$$

$$f_{(1,0)} = \sigma \cdot a^{(0,1)}, \quad f_{(1,1)} = \sigma \cdot \sigma^{(0,1)}$$

Hierarchical sets:

For $\alpha = (\alpha_1, \ldots, \alpha_r) \in \mathcal{M}$ put

$$\zeta(\alpha) = \sharp\{j :\ \alpha_j = 0\}, \quad ||\alpha|| = r + \zeta(\alpha)$$

For $\gamma \in \mathbb{N}/2$ define

$$A_\gamma = \{\alpha \in \mathcal{M} :\ ||\alpha|| \le 2\gamma \text{ or } ||\alpha|| = 2\zeta(\alpha) = 2\gamma + 1\}$$

Note: $\gamma_1 < \gamma_2 \Rightarrow A_{\gamma_1} \subset A_{\gamma_2}$

Example 9.

$\gamma = 1/2$: $A_{1/2} = \{(0), (1)\}$
$f_{(0)} = a,\ f_{(1)} = \sigma$
Euler scheme

$\gamma = 1$: $A_1 = \{(0), (1), (1,1)\}$
$f_{(0)} = a,\ f_{(1)} = \sigma,\ f_{(1,1)} = \sigma \cdot \sigma^{(0,1)}$
Milstein scheme

$\gamma = 3/2$: $A_{3/2} = \{(0), (1), (0,0), (0,1), (1,0), (1,1), (1,1,1)\}$
yields the

Wagner-Platen scheme (Wagner and Platen 1978)

$$\widehat{X}^{3/2}(t_0) = x_0,$$

$$\begin{aligned}
\widehat{X}^{3/2}&(t_{\ell+1}) \\
&= \widehat{X}^{3/2}(t_\ell) + a(t_\ell, \widehat{X}^{3/2}(t_\ell)) \cdot (t_{\ell+1} - t_\ell) \\
&\qquad + \sigma(t_\ell, \widehat{X}^{3/2}(t_\ell)) \cdot (W(t_{\ell+1}) - W(t_\ell)) \\
&\quad + 1/2 \cdot \big(\sigma^{(0,1)} \cdot \sigma\big)(t_\ell, \widehat{X}^{3/2}(t_\ell)) \\
&\qquad \times \big((W(t_{\ell+1}) - W(t_\ell))^2 - (t_{\ell+1} - t_\ell)\big) \\
&\quad + \big(\sigma^{(1,0)} + a \cdot \sigma^{(0,1)} - \sigma/2 \cdot (\sigma^{(0,1)})^2\big)(t_\ell, \widehat{X}^{3/2}(t_\ell)) \\
&\qquad \times \big(W(t_{\ell+1}) - W(t_\ell)\big) \cdot (t_{\ell+1} - t_\ell) \\
&\quad + 1/6 \cdot \big(\sigma \cdot (\sigma^{(0,1)})^2 + \sigma^2 \cdot \sigma^{(0,2)}\big)(t_\ell, \widehat{X}^{3/2}(t_\ell)) \\
&\qquad \times \big(W(t_{\ell+1}) - W(t_\ell)\big)^3 \\
&\quad + 1/2 \cdot \big(a^{(1,0)} + a \cdot a^{(0,1)} + \sigma^2/2 \cdot a^{(0,2)}\big)(t_\ell, \widehat{X}^{3/2}(t_\ell)) \\
&\qquad \times (t_{\ell+1} - t_\ell)^2 \\
&\quad + \big(\sigma \cdot a^{(0,1)} - \sigma^{(1,0)} \\
&\qquad\qquad - a \cdot \sigma^{(0,1)} - \sigma^2/2 \cdot \sigma^{(0,2)}\big)(t_\ell, \widehat{X}^{3/2}(t_\ell)) \\
&\qquad \times \int_{t_\ell}^{t_{\ell+1}} \big(W(s) - W(t_\ell)\big)\, ds
\end{aligned}$$

Note: The Wagner-Platen scheme uses point evaluations <u>and</u> continuous linear functionals of W.

For $\gamma \in \mathbb{N}/2$ put

$$B_\gamma = \{(\alpha_1, \dots, \alpha_r) \in \mathcal{M} \setminus A_\gamma : \ (\alpha_2, \dots, \alpha_r) \in A_\gamma\}$$

Theorem 5. *Assume* (L_{f_α}) *and* (LLG_{f_α}) *for all* $\alpha \in A_\gamma \cup B_\gamma$.

(i) *(Kloeden and Platen 1995)*

$$\max_\ell \left(E|X(t_\ell) - \widehat{X}^\gamma(t_\ell)|^q\right)^{1/q} \le c \cdot \Delta_{\max}^\gamma$$

(ii) *For the piecewise linear interpolated scheme* $\widehat{X}^\gamma$

$$\left(E\|X - \widehat{X}^\gamma\|_p^q\right)^{1/q} \le c \cdot \begin{cases} \Delta_{\max}^{1/2} & \text{if } p < \infty \\ \Delta_{\max}^{1/2} \cdot \left(\ln \Delta_{\max}^{-1}\right)^{1/2} & \text{if } p = \infty \end{cases}$$

Remark 4. Part (ii) of Theorem 5 is shown in the same manner as Theorem 4(ii).

Note again: In general, the simulation of iterated Itô-integrals is an unsolved problem. See, e.g., Gaines and Lyons (1994), and Rydén and Wiktorsson (2001) for that problem.

Remark 5. Meanwhile there exists a multitude of schemes including, e.g., Runge-Kutta methods and multistep methods. See Kloeden and Platen (1995).

II. Optimal L_2-approximation based on point evaluations of W

Motivation:

Consider a classical strong approximation $\widehat{X}$ based on a discretization $0 = t_0 < \dots < t_n = 1$. Typical result:

$$\left(E\|X - \widehat{X}\|_2^2\right)^{1/2} \le c \cdot \Delta_{\max}^{1/2}$$

(e.g., Theorems 3, 4 and 5 for $p = q = 2$ in Part II)

Shortcomings of this kind of result are

(i) Computational cost of $\widehat{X}$ is not taken into account, e.g.,

(a) number of calls to random number generator

(b) number of evaluations of a, σ, $a^{(0,1)}$, etc.

(c) number of arithmetic operations to compute $\widehat{X}$

A more useful result would be of the type

$$\left(E\|X - \widehat{X}\|_2^2\right)^{1/2} \le c \cdot \left(\mathrm{cost}(\widehat{X})\right)^{-1/2}$$

with an appropriate definition of $\mathrm{cost}(\widehat{X})$, e.g.,

$$\mathrm{cost}(\widehat{X}) = (a) + (b) + (c)$$

$$\text{or} \quad \mathrm{cost}(\widehat{X}) = (a).$$

Thus: Relate the error to the cost

(ii) The result gives only an upper bound with an unspecified constant. Different methods or different discretizations can not be compared.

Thus: Lower bounds and explicit constants

Questions:

- Optimal method?
- Adaptive discretization?
- Complexity?

  Given $\varepsilon > 0$ what is the minimal cost to spend in order to achieve an error of at most ε?

  or, equivalently,

  Given a (cost)-budget N what is the minimal error that can be achieved using methods with cost at most N?

For a systematic treatment of these questions we need to specify

- class of methods
- error criterion
- cost measure

(Approach in the framework of *information based complexity*, see Traub, Wasilkowski and Woźniakowski 1988)

Class of methods

$\mathbb{X}$: All (up to measurability) approximations $\widehat{X}$ that are are based on a finite number of evaluations of W at points

$$\tau_1, \ldots, \tau_\nu \in [0,1]$$

where the choice of $\tau_{\ell+1}$ may depend in any (measurable) way on the previously computed values $W(\tau_1), \ldots, W(\tau_\ell)$ (in particular $\tau_\ell < \tau_{\ell+1}$ is not required).

and the total number ν of evaluations may be determined by a (measurable) stopping rule.

(See below for a formal definition of $\mathbb{X}$)

Error

$$e(\widehat{X}) = \big(E\|X - \widehat{X}\|_2^2\big)^{1/2} = \left(E \int_0^1 \big(X(t) - \widehat{X}(t)\big)^2 \, dt\right)$$

Cost

$$c(\widehat{X}) = E(\nu) = \text{average number of calls to random number generator}$$

Minimal errors

$$e_N = \inf\{e(\widehat{X}) : \ \widehat{X} \in \mathbb{X},\ c(\widehat{X}) \le N\}, \quad N > 0$$

The quantity e_N is the minimal L_2-error that can be achieved with methods in the class $\mathbb{X}$ that use at most N point evaluations of W on the average.

Mathematical goals:

(i) Exact rate of convergence of e_N, i.e., order and asymptotic constant (answers complexity question in asymptotic sense)

(ii) Easy to implement method $\widehat{X}_n$ that achieves this rate, i.e.,

$$\lim_{n\to\infty} \frac{e(\widehat{X}_n)}{e_{c(\widehat{X}_n)}} = 1$$

((strong) asymptotic optimality)

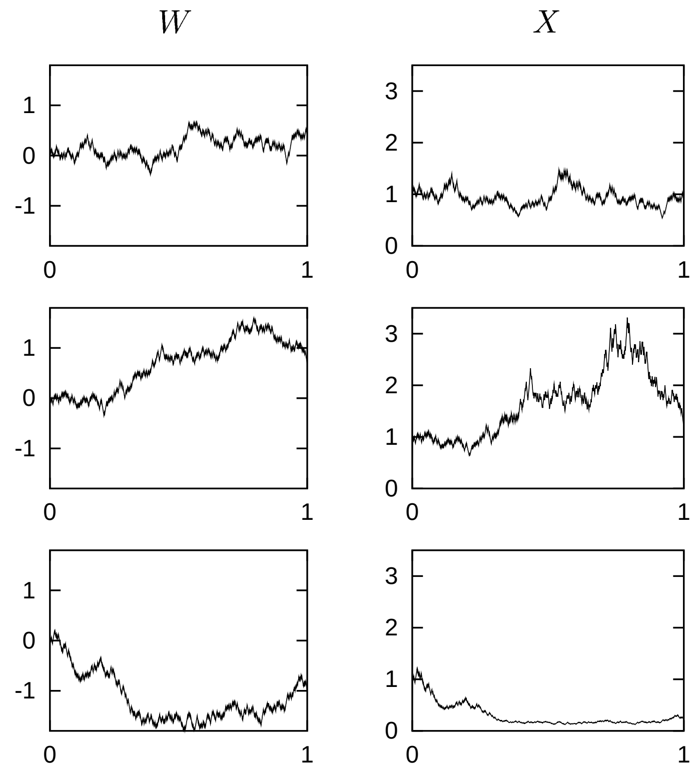
Remark 6. The class $\mathbb{X}$ is specified by the type of information about W. Here only Dirac functionals are admitted. In particular, $\mathbb{X}$ contains (piecewise) interpolated Euler and Milstein schemes but not interpolated Itô-Taylor schemes of order $\gamma > 1$. See Part V for results concerning continuous linear functionals or iterated Itô-integrals.

Why should the points $\tau_1, \ldots, \tau_\nu$ be chosen sequentially?

Example 10. (Geometric Brownian motion)

$$dX(t) = X(t)\, dW(t), \quad X(0) = 1$$
$$X(t) = \exp(-t/2 + W(t))$$

Three trajectories of W and X:



Discretization should be adapted to trajectories of solution

Recall from Part I:

Under appropriate conditions on a and σ

$$E\big((X(t+\delta)-X(t))^2 \big| X(t)=x\big) = \sigma^2(t,x)\cdot\delta + o(\delta)$$

Oszillation of trajectories of X is (in mean square sense) locally in time and space determined by:

conditional Hölder constant $|\sigma(t, X(t))|$

In the preceding example: $|\sigma(t, X(t))| = X(t)$
(c.f. Example 4)

Basic idea: The *larger* the value of $|\sigma(t, X(t))|$
the *finer* the discretization around t.

Problems:

(i) Quantification of *larger* and *finer*

(ii) We can not observe the value of $|\sigma(t, X(t))|$ in general

The fully adaptive method $\widehat{X}_n^*$

Start: $\tau_0 = 0, \quad \widehat{X}_n^*(\tau_0) = x_0$

ℓ-th step:

- known data
$$\tau_j,\ x_j = \widehat{X}_n^*(\tau_j), \quad j = 0,\ldots,\ell-1$$
- estimate the conditional Hölder constant at $\tau_{\ell-1}$ by
$$|\sigma(\tau_{\ell-1}, x_{\ell-1})|$$
and compute candidate for next discretization point
$$\tau_\ell = \tau_{\ell-1} + \frac{1}{n\cdot|\sigma(\tau_{\ell-1}, x_{\ell-1})|}$$
- If $\tau_\ell < 1$ then do Milstein step
$$\begin{aligned}\widehat{X}_n^*(\tau_\ell) = x_{\ell-1} &+ a(\tau_{\ell-1}, x_{\ell-1}))\cdot(\tau_\ell - \tau_{\ell-1}) \\ &+ \sigma(\tau_{\ell-1}, x_{\ell-1}))\cdot(W(\tau_\ell) - W(\tau_{\ell-1})) \\ &+ 1/2\cdot\big(\sigma\cdot\sigma^{0,1)}\big)(\tau_{\ell-1}, x_{\ell-1})) \\ &\quad\times\big((W(\tau_\ell) - W(\tau_{\ell-1}))^2 - (\tau_\ell - \tau_{\ell-1})\big)\end{aligned}$$
and proceed with $(\ell+1)$-th step.
Else set $\tau_\ell = 1$, calculate $\widehat{X}_n^*(1)$ as above and stop

End: Piecewise linear interpolation

Number of evaluations of W pathwise

$$\nu_n = \min\{\ell \in \mathbb{N} : \ \tau_\ell = 1\}$$
$$\text{approx } n \cdot \int_0^1 |\sigma(t, X(t))| \, dt$$

The *harder* the trajectory the more points are used

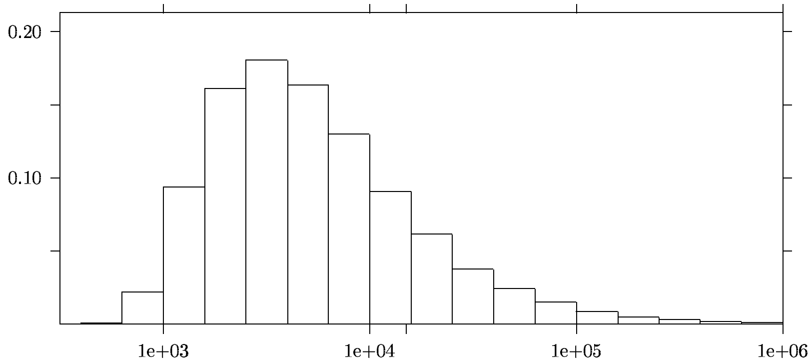
Simulation experiment:

$$dX(t) = 3X(t)\, dW(t)$$
$$n = 5000$$

10000 trajectories $w_1, \dots, w_{10000}$ of W
Average number of evaluations

$$\frac{1}{10^4} \sum_{i=1}^{10^4} \nu_{5000}(w_i) = 14981$$

Relative frequencies of ν_{5000}:



Unfortunately, it's hard to analyze $\widehat{X}_n^*$

The simplified adaptive method $\widehat{X}_n$

Idea: Do not update the step-size in every step

Algorithm:

1) Dependent on n determine *coarse* equidistant prediscretization
$$t_\ell = \ell / k_n, \quad \ell = 0, 1, \dots, k_n$$
Long Milstein-steps
$$\widehat{X}_n(t_\ell) = \widehat{X}^M(t_\ell)$$

2) Determine number of additional equidistant points in subintervals $]t_\ell, t_{\ell+1}[$ by
$$\mu_\ell = \lceil n/k_n \cdot |\sigma(t_\ell, \widehat{X}^M(t_\ell))| \rceil$$
Resulting points
$$\tau_{\ell,j} = t_\ell + j \cdot (k_n \cdot (\mu_\ell + 1))^{-1}, \quad j = 0, 1, \dots, \mu_\ell$$
Short Euler-steps with "frozen" values of drift and diffusion coefficient
$$\widehat{X}_n(\tau_{\ell,j}) = \widehat{X}_n(\tau_{\ell,j-1}) + a(t_\ell, \widehat{X}^M(t_\ell)) \cdot (\tau_{\ell,j} - \tau_{\ell,j-1}) + \sigma(t_\ell, \widehat{X}^M(t_\ell)) \cdot (W(\tau_{\ell,j}) - W(\tau_{\ell,j-1}))$$

3) Piecewise linear interpolation

Adjusting the number k_n of coarse grid points

k_n should be small compared to the total number $\sum_\ell \mu_\ell$ of adaptive points in order to keep track of the local smoothness of the solution:

$$\lim_{n\to\infty} k_n/n = 0$$

k_n should be large enough to guarantee sufficiently good estimates of the values of the drift and diffusion coefficients:

$$\lim_{n\to\infty} k_n^2/n = \infty$$

Assumptions throughout the rest of Part III:

$$(L_f),\ (LLG_f),\ (L_{f(0,1)}) \quad \text{for } f = a \text{ and } f = \sigma$$

Put

$$C = \frac{1}{\sqrt{6}} \cdot E \int_0^1 |\sigma(t, X(t))|\, dt$$

Theorem 6. *(Hofmann, M-G, Ritter 2001)*

(i) $$\lim_{n\to\infty} \big(c(\widehat{X}_n)\big)^{1/2} \cdot e(\widehat{X}_n) = C$$

(ii) $$\lim_{n\to\infty} n^{-1} \cdot c(\widehat{X}_n) = \sqrt{6} \cdot C.$$

A first comparison:

$\widehat{X}_n^M$: Piecewise linear interpolated Milstein scheme with equidistant discretization $0, 1/n, 2/n, \ldots, 1$

Note: $c(\widehat{X}_n^M) = n$

Put

$$C^{\text{equi}} = \frac{1}{\sqrt{6}} \cdot \left(E \int_0^1 \sigma^2(t, X(t))\, dt\right)^{1/2}$$

Theorem 7. *(Hofmann, M-G, Ritter 2001)*

$$\lim_{n\to\infty} n^{1/2} \cdot e(\widehat{X}_n^M) = C^{equi}$$

Example 11. Geometric Brownian motion

$$dX(t) = b \cdot X(t)\, dW(t), \quad X(0) = 1$$

Here

$$C^{\text{equi}} = \frac{1}{\sqrt{6}} \cdot \big(\exp(b^2) - 1\big)^{1/2}, \quad C = \frac{1}{\sqrt{6}} \cdot |b|$$

<u>The class $\mathbb{X}$</u>

An approximation $\widehat{X} \in \mathbb{X}$ is determined by 3 sequences of measurable mappings

$$\psi = (\psi_k)_{k\in\mathbb{N}},\ \chi = (\chi_k)_{k\in\mathbb{N}},\ \varphi = (\varphi_k)_{k\in\mathbb{N}},$$

with

$$\begin{aligned} &\psi_\ell : \mathbb{R}^\ell \to [0,1] && \text{(determines } \tau_\ell\text{)}, \\ &\chi_\ell : \mathbb{R}^{\ell+1} \to \{\text{STOP}, \text{GO}\} && \text{(determines stopping)}, \\ &\varphi_\ell : \mathbb{R}^{\ell+1} \to L_2([0,1]) && \text{(determines approximation)}. \end{aligned}$$

First evaluation point: $\tau_1 = \psi_1(x_0)$

ℓ-th step: known data

$$D_\ell = (x_0, W(\tau_1), \ldots, W(\tau_\ell))$$

if $\chi_\ell(D_\ell) = \text{GO}$ then evaluate W at

$$\tau_{\ell+1} = \psi_{\ell+1}(D_\ell)$$

if $\chi_\ell(D_\ell) = \text{STOP}$ then use approximation

$$\varphi_\ell(D_\ell)$$

Total number of evaluations:

$$\nu_{(\psi,\chi,\varphi)} = \min\{\ell \in \mathbb{N} : \chi_\ell(D_\ell) = \text{STOP}\}$$

Put

$$S = \{(\psi,\chi,\varphi) : P(\nu_{(\psi,\chi,\varphi)} < \infty) = 1\}$$

For $(\psi,\chi,\varphi) \in S$ define

$$\widehat{X}_{(\psi,\chi,\varphi)} = \varphi_{\nu_{(\psi,\chi,\varphi)}}\big(D_{\nu_{(\psi,\chi,\varphi)}}\big)$$

andlet

$$\mathbb{X} = \big\{\widehat{X}_{(\psi,\chi,\varphi)} : (\psi,\chi,\varphi) \in S\big\}$$

Note:

- $\mathbb{X}$ contains every implementable method;
  in particular all methods based on evaluation of W at a fixed discretization
  $$0 = t_0 < \ldots < t_n = 1.$$
  In this case
  $$\psi_\ell \equiv t_\ell, \quad \ell = 1, \ldots, n,$$
  $$\chi_\ell \equiv \begin{cases} \text{GO} & \text{if } \ell \le n-1 \\ \text{STOP} & \text{if } \ell = n \end{cases}$$
- $\mathbb{X}$ contains the adaptive methods $\widehat{X}_n$ and $\widehat{X}_n^*$
- $\mathbb{X}$ even contains methods, which may not be implementable, e.g., conditional expectations
  $$\widehat{X} = E(X \mid W(\tau_1), \ldots, W(\tau_\nu))$$
- No restrictions on information about a and σ.

Recall minimal errors $\quad e_N = \inf\big\{e(\widehat{X}) : \widehat{X} \in \mathbb{X},\ c(\widehat{X}) \le N\big\}$

Theorem 8. *(Hofmann, M-G, Ritter 2001)*

$$\lim_{N\to\infty} N^{1/2} \cdot e_N = C$$

Corollary 2. *If $C > 0$ then $\widehat{X}_n$ is asymptotically optimal.*

Remark 7.

$$\begin{aligned} C = 0 &\Leftrightarrow P(\forall t \in [0,1] : \ \sigma(t, X(t) = 0) = 1 \\ &\Leftrightarrow \text{equation (1) is deterministic} \end{aligned}$$

Can we be as good with

- a fixed number of discretization points?
- a fixed discretization?
- an equidistant discretization?

Interesting subclasses of $\mathbb{X}$

Corresponding to the above questions we consider

$$\begin{aligned}
\mathbb{X}^{\text{fixed } \nu} &= \{\widehat{X}_{(\psi,\chi,\varphi)} \in \mathbb{X} : \nu_{(\psi,\chi,\varphi)} \text{ constant}\} \\
\mathbb{X}^{\text{fixed}} &= \{\widehat{X}_{(\psi,\chi,\varphi)} \in \mathbb{X}^{\text{fixed } \nu} : \psi_\ell \text{ constant for } \ell \le \nu_{(\psi,\chi,\varphi)}\} \\
\mathbb{X}^{\text{equi}} &= \{\widehat{X}_{(\psi,\chi,\varphi)} \in \mathbb{X}^{\text{fixed}} : \psi_\ell = \ell/\nu_{(\psi,\chi,\varphi)} \text{ for } \ell \le \nu_{(\psi,\chi,\varphi)}\}
\end{aligned}$$

with respective minimal errors

$$\begin{aligned}
e_N^{\text{fixed } \nu} &= \inf\{e(\widehat{X}) : \widehat{X} \in \mathbb{X}^{\text{fixed } \nu},\ c(\widehat{X}) \le N\} \\
e_N^{\text{fixed}} &= \inf\{e(\widehat{X}) : \widehat{X} \in \mathbb{X}^{\text{fixed}},\ c(\widehat{X}) \le N\} \\
e_N^{\text{equi}} &= \inf\{e(\widehat{X}) : \widehat{X} \in \mathbb{X}^{\text{equi}},\ c(\widehat{X}) \le N\}
\end{aligned}$$

Note: (Piecewise linear) interpolated Euler schemes and Milstein schemes belong to the class $\mathbb{X}^{\text{fixed}}$.
The (piecewise linear) interpolated equidistant Milstein schemes $\widehat{X}_n^M$ belong to the class $\mathbb{X}^{\text{equi}}$.

Put

$$\begin{aligned}
C^{\text{fixed } \nu} &= \frac{1}{\sqrt{6}} \cdot \left(E\left(\int_0^1 |\sigma(t, X(t))|\, dt \right)^2 \right)^{1/2} \\
C^{\text{fixed}} &= \frac{1}{\sqrt{6}} \cdot \int_0^1 \left(E\big(\sigma^2(t, X(t))\big) \right)^{1/2} dt
\end{aligned}$$

and recall the constant

$$C^{\text{equi}} = \frac{1}{\sqrt{6}} \cdot \left(E \int_0^1 \sigma^2(t, X(t))\, dt \right)^{1/2}$$

Theorem 9. *(Hofmann, M-G, Ritter 2001)*

$$\begin{aligned}
\lim_{N\to\infty} N^{1/2} \cdot e_N^{\textit{fixed } \nu} &= C^{\textit{fixed } \nu} \\
\lim_{N\to\infty} N^{1/2} \cdot e_N^{\textit{fixed}} &= C^{\textit{fixed}} \\
\lim_{N\to\infty} N^{1/2} \cdot e_N^{\textit{equi}} &= C^{\textit{equi}}
\end{aligned}$$

Recall the asymptotic constant for e_N:

$$C = \frac{1}{\sqrt{6}} \cdot E \int_0^1 |\sigma(t, X(t))|\, dt$$

Note that

$$C \le C^{\text{fixed } \nu} \le C^{\text{fixed}} \le C^{\text{equi}}$$

with strict inequality in most cases, see Remark 9.

Example 12. Geometric Brownian motion

$$dX(t) = b \cdot X(t)\, dW(t), \quad X(0) = 1$$

Here

$$C^{\text{equi}} = \frac{1}{\sqrt{6}} \cdot \big(\exp(b^2) - 1\big)^{1/2}$$

$$C^{\text{fixed}} = \frac{1}{\sqrt{6}} \cdot \frac{2}{|b|} \cdot \big(\exp(b^2/2) - 1\big)$$

$$C^{\text{fixed } \nu} = \frac{1}{\sqrt{6}} \cdot \frac{1}{|b|} \cdot \big(2 \cdot \exp(b^2) - b^2 - 1\big)^{1/2}$$

$$C = \frac{1}{\sqrt{6}} \cdot |b|$$

Exponential dependence on b except for the constant C.

Remark 8. (Interpretation of asymptotic constants)

Recall: conditional Hölder constant $|\sigma(t, X(t))|$ describes smoothness of X locally in time and space.

Constants C and $C^{\text{fixed } \nu}$ are based on average

$$H = \int_0^1 |\sigma(t, X(t))|\, dt$$

of the conditional Hölder constant along a trajectory:

$$C = 1/\sqrt{6} \cdot E(H), \quad C^{\text{fixed } \nu} = 1/\sqrt{6} \cdot (E(H^2))^{1/2}$$

Due to (7)

$$(8) \qquad E(X(t+\delta) - X(t))^2 = E|\sigma(t, X(t))|^2 \cdot \delta + o(\delta)$$

Thus

$$\alpha(t) = \big(E|\sigma(t, X(t))|^2\big)^{1/2}$$

describes smoothness of X only locally in time.

Constants C^{fixed} and C^{equi} are based on α:

$$C^{\text{fixed}} = \frac{1}{\sqrt{6}} \cdot \int_0^1 \alpha(t)\, dt, \quad C^{\text{equi}} = \frac{1}{\sqrt{6}} \cdot \left(\int_0^1 \alpha^2(t)\, dt\right)^{1/2}$$

Remark 9. (Equality of asymptotic constants)

In general

$C^{\text{equi}} = C^{\text{fixed}}$ iff α is constant

$C^{\text{fixed}} = C^{\text{fixed } \nu}$

iff $\exists\, t_0 \in [0,1],\ \gamma \in C([0,1])$:

$P\big(\forall t \in [0,1] : |\sigma(t, X(t))| = \gamma(t) \cdot |\sigma(t_0, X(t_0))|\big) = 1$

$C^{\text{fixed } \nu} = C$

iff $P\big(\forall t \in [0,1] : |\sigma(t, X(t))| = \alpha(t)\big) = 1$

For equations with additive noise, i.e., $\sigma^{0,1} = 0$: $C = C^{\text{fixed}}$

Furthermore

$C = C^{\text{fixed } \nu} = C^{\text{fixed}} = C^{\text{equi}} = 0$ iff $C = 0$

iff eq (1) is deterministic

Remark 10. (L_2-approximation and L_2-reconstruction)

Consider weighted Brownian motion

$$Y(t) = \rho(t) \cdot W(t), \quad t \in [0,1],$$

with continuous $\rho : [0,1] \to [0,\infty[$.

Best L_2-approximation of Y based on $Y(t_1), \dots, Y(t_N)$:

$$\widehat{Y}_{(t_1,\dots,t_N)}(t) = E\big(Y(t) \,\big|\, Y(t_1), \dots, Y(t_N)\big)$$

Resulting minimal mean square L_2-error:

$$\widetilde{e}_N = \inf\big\{(E\|Y - \widehat{Y}_{(t_1,\dots,t_N)}\|_2^2)^{1/2} : \, 0 < t_1 < \dots < t_N \le 1\big\}$$

Note: Adaption does not help here since Y is Gaussian

From general results on L_2-reconstruction (see Ritter 2000):

$$\lim_{N\to\infty} N^{1/2} \cdot \widetilde{e}_N = C_\rho := \frac{1}{\sqrt{6}} \cdot \int_0^1 \rho(t)\,dt$$

Taking the weight $\rho = \alpha$ yields

$$C^{\text{fixed}} = C_\alpha$$

Taking the *random weight* $\rho = |\sigma(\cdot, X(\cdot))|$ we get

$$C = E\big(C_{|\sigma(\cdot,X(\cdot))|}\big)$$

Heuristics:

- Locally in time, the solution X behaves like a Brownian motion weighted with α, see (8).

- Locally in time and space, the solution X behaves like a Brownian motion randomly weighted with $|\sigma(\cdot, X(\cdot))|$, see (7).
- Approximation of X at discrete points is *cheap* compared to L_2-approximation of X, see Theorems 4 and 7.

Remark 1. (Asymptotically optimal methods)

For the class $\mathbb{X}$: The adaptive method $\widehat{X}_n$ (Corollary 2)

For the class $\mathbb{X}^{\text{fixed }\nu}$: Same definition as $\widehat{X}_n$ with numbers μ_ℓ (essentially) replaced by

$$\mu_\ell = \left\lceil n \cdot |\sigma(t_\ell, \widehat{X}^M(t_\ell))| \cdot \Big(\sum_{j=0}^{k_n-1} |\sigma(t_j, \widehat{X}^M(t_j))|\Big)^{-1} \right\rceil$$

For the class $\mathbb{X}^{\text{fixed}}$: Piecewise linear interpolation of Milstein scheme with *regular* discretization

$$\int_{t_\ell^{(n)}}^{t_{\ell+1}^{(n)}} \alpha(t)\,dt = \frac{1}{n} \cdot \int_0^1 \alpha(t)\,dt, \quad \ell = 0, 1, \dots, n-1$$

For the class $\mathbb{X}^{\text{equi}}$: Piecewise linear interpolation of the equidistant Milstein scheme $\widehat{X}_n^M$ (Theorems 6 and 8)

Simulation experiments

Equation

$$dX(t) = 3 \cdot X(t)\, dW(t), \quad X(0) = 1.$$

Comparison of fully adaptive method $\widehat{X}_n^*$ with simplified adaptive method $\widehat{X}_n$:

- Error and cost of $\widehat{X}_n^*$ by simulation
- Error and cost of $\widehat{X}_n$ by asymptotic formulas from Theorem 6:

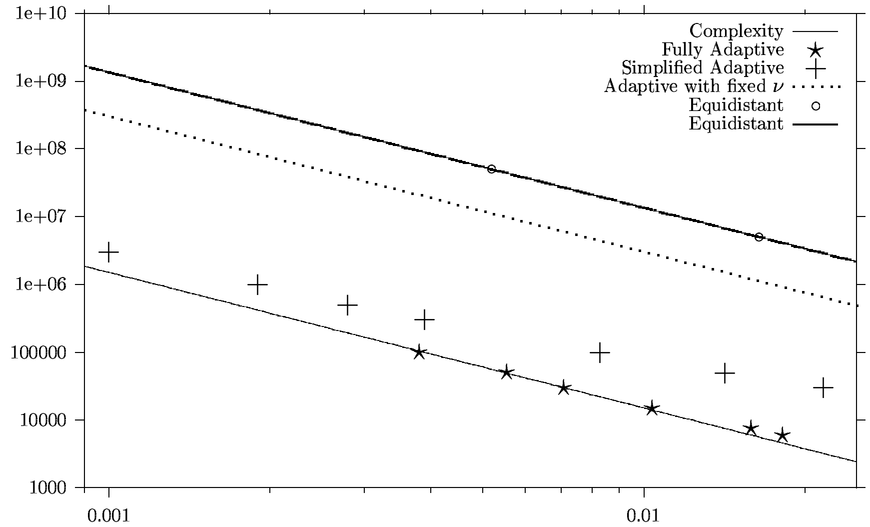
$$c(\widehat{X}_n) \approx 3n, \quad e(\widehat{X}_n) \approx 1/\sqrt{2n}$$

n	**Error**		**Cost**	
	Simul.	Asymp.	Simul.	Asymp.
$2 \cdot 10^3$	$1.82 \cdot 10^{-2}$	$1.58 \cdot 10^{-2}$	6009	6000
$2.5 \cdot 10^3$	$1.59 \cdot 10^{-2}$	$1.41 \cdot 10^{-2}$	7624	7500
$5 \cdot 10^3$	$1.04 \cdot 10^{-2}$	$1.00 \cdot 10^{-2}$	14981	15000
$1 \cdot 10^4$	$7.09 \cdot 10^{-3}$	$7.07 \cdot 10^{-3}$	29843	30000
$1/6 \cdot 10^5$	$5.55 \cdot 10^{-3}$	$5.48 \cdot 10^{-3}$	51297	50000
$1/3 \cdot 10^5$	$3.81 \cdot 10^{-3}$	$3.87 \cdot 10^{-3}$	100122	100000

Conjecture: The fully adaptive method $\widehat{X}_n^*$ is also asymptotically optimal in the class $\mathbb{X}$

Dependence of cost on error within range $[10^{-3}, 2 \cdot 10^{-2}]$:

- Solid or dotted lines by asymptotic formulas from Theorems 6, 7 and 9.
- $\bigstar$ from simulated values in previous table
- $+$ and $\bigcirc$ from explicit formulas for $c(\widehat{X}_n)$ and $e(\widehat{X}_n)$ as well as $e(\widehat{X}_n^M)$



Using the fully or simplified adaptive method the computation time (in terms of the average number of calls to random number generator) decreases by a factor close to

$$\left(C/C^{\text{equi}}\right)^2 = 9/(\exp(9) - 1) = 0.0011108...$$

IV. Optimal approximation at a single point based on point evaluations of W

Classes of methods as in Part III, i.e.,

$\mathbb{X}$: All approximations $\widehat{X}$ that are based on a finite (random) number ν of point evaluations of W

$\mathbb{X}^{\text{fixed }\nu}$: All approximations $\widehat{X} \in \mathbb{X}$ with deterministic ν

$\mathbb{X}^{\text{fixed}}$: All approximations $\widehat{X} \in \mathbb{X}^{\text{fixed }\nu}$ with fixed discretization

$\mathbb{X}^{\text{equi}}$: All approximations $\widehat{X} \in \mathbb{X}^{\text{fixed}}$ with equidistant discretization

Cost measure as in Part III, i.e.,

$$c(\widehat{X}) = E(\nu) = \text{ average number of evaluations of } W$$

Error

$$e(\widehat{X}) = \left(E|X(1) - \widehat{X}(1)|^2\right)^{1/2}$$

Minimal errors

$$e_N = \inf\{e(\widehat{X}) : \widehat{X} \in \mathbb{X},\ c(\widehat{X}) \le N\}$$

Canonical definition of minimal errors $e_N^{\text{fixed }\nu}$, e_N^{fixed}, e_N^{equi}

Assumption throughout the rest of Part IV:

$$a^{(i,j)},\ \sigma^{(i,j)}, \quad i = 0,1,2,\ j = 0,1,2,3,$$

exist and are continuous and bounded.

Example 13. (Geometric Brownian motion)

$$dX(t) = \mu \cdot X(t)\,dt + b \cdot X(t)\,dW(t), \quad X(0) = 1$$

Here

$$X(1) = \exp\bigl((\mu - b^2/2) + b \cdot W(1)\bigr), \quad e_N = e_1 = 0$$

Example 14. (Additive noise)

$$dX(t) = a(t)\,dt + \sigma(t)\,dW(t), \quad X(0) = 0,$$

Here

$$X(1) = \int_0^1 a(t)\,dt + \sigma(1) \cdot W(1) - \int_0^1 \sigma'(t) \cdot W(t)\,dt$$

Essentially, approximation of

$$\int_0^1 \sigma'(t) \cdot W(t)\,dt$$

Example 15. (Langevin equation)

$$dX(t) = \alpha \cdot X(t)\,dt + dW(t), \quad X(0) = 1$$

Here

$$X(1) = W(1) + \int_0^1 \alpha \cdot e^{\alpha\cdot(1-t)} \cdot W(t)\,dt$$

Essentially, approximation of

$$\int_0^1 \alpha \cdot e^{\alpha\cdot(1-t)} \cdot W(t)\,dt$$

Thus: Mean square approximation of X at a single point is strongly connected to approximation of the integral of a weighted Brownian motion

The weighting process

Define

$$\Theta(t) = \eta(t, X(t)) \cdot \Phi(t), \quad t \in [0,1],$$

where

$$\eta = \sigma a^{(0,1)} - \sigma^{(1,0)} - a\sigma^{(0,1)} - 1/2 \cdot \sigma^2 \sigma^{(0,2)}$$

and

$$\Phi(t) = \exp\Big(\int_t^1 \big(a^{(0,1)} - 1/2 \cdot (\sigma^{(0,1)})^2\big)(u, X(u))\,du + \int_t^1 \sigma^{(0,1)}(u, X(u))\,dW(u) \Big)$$

Motivation for η:

General results on approximation of the integral of a weighted Brownian motion suggest (for nonzero weight)

$$(*) \qquad e_N \asymp 1/N,$$

see Ritter 2000. (Upper bound is a consequence of Theorem 4)

For the Wagner-Platen scheme $\widehat{X}_N^{3/2}$ with N equidistant discretization points

$$\big(E|X(1) - \widehat{X}_N^{3/2}(1)|^2\big)^{1/2} \le c/N^{3/2},$$

see Theorem 5.

Note that $\widehat{X}_N^{3/2} \notin \mathbb{X}$ since this method uses integrals of the Brownian motion, see Part II.

If $(*)$ is true then approximation of $X(1)$ is equivalent to approximation of $\widehat{X}_N^{3/2}(1)$.

The function η is composed from the coefficient function associated with the Brownian integrals used by the Wagner-Platen scheme

$$\eta = f_{(1,0)} - f_{(0,1)}$$

$$I_{(1,0),t_\ell,t_{\ell+1}} = \int_{t_\ell}^{t_{\ell+1}} (W(t) - W(t_\ell))\,dt$$

$$I_{(0,1),t_\ell,t_{\ell+1}} = (t_{\ell+1} - t_\ell) \cdot ((W(t_{\ell+1}) - W(t_\ell)) - \int_{t_\ell}^{t_{\ell+1}} (W(t) - W(t_\ell))\,dt$$

All other Itô-integrals are functions of point evaluations of W

Interpretation of the process Φ:

Roughly, $\Phi(t)$ is the mean square derivative of the solution at the final time point $X(1)$ with respect to its state at time t.

For $x \in \mathbb{R}$ and $t \in [0,1]$ consider the equation

$$dX_{t,x}(s) = a(s, X_{t,x}(s))\, ds + \sigma(s, X_{t,x}(s))\, dW(s), \quad t \le s \le 1,$$

with initial value $X_{t,x}(t) = x$.

Note that

$$P^{(X_{t,x}(s))_{t\le s\le 1}} = P^{(X(s))_{t\le s\le 1}\,|\,X(t)=x}$$

For every $s \in [t,1]$ there exists the mean square derivative $X'_{t,x}(s)$ of $X_{t,x}(s)$ with respect to the initial value x, i.e.,

$$\lim_{h\to 0} E\big(1/h \cdot (X_{t,x+h}(s) - X_{t,x}(s)) - X'_{t,x}(s)\big)^2 = 0.$$

The derivative process is explicitly given by

$$X'_{t,x}(s) = \exp\bigg(\int_t^s \big(a^{(0,1)} - 1/2 \cdot (\sigma^{(0,1)})^2\big)(u, X_{t,x}(u))\, du + \int_t^s \sigma^{(0,1)}(u, X_{t,x}(u))\, dW(u)\bigg).$$

Taking $s = 1$ and replacing $X_{t,x}$ by the solution X in the right hand side above yields the definition of $\Phi(t)$.

Analysis of minimal errors

Define the constants

$$C = \frac{1}{\sqrt{12}} \cdot \left(E \int_0^1 |\Theta(t)|^{2/3}\, dt\right)^{3/2}$$

$$C^{\text{fixed }\nu} = \frac{1}{\sqrt{12}} \cdot \left(E\left(\int_0^1 |\Theta(t)|^{2/3}\, dt\right)^3\right)^{1/2}$$

$$C^{\text{fixed}} = \frac{1}{\sqrt{12}} \cdot \left(\int_0^1 \big(E|\Theta(t)|^2\big)^{1/3}\, dt\right)^{3/2}$$

$$C^{\text{equi}} = \frac{1}{\sqrt{12}} \cdot \left(\int_0^1 E|\Theta(t)|^2\, dt\right)^{1/2}$$

Theorem 10. *(M-G 2004)*

$$\lim_{N\to\infty} N \cdot e_N = C$$

$$\lim_{N\to\infty} N \cdot e_N^{\textit{fixed }\nu} = C^{\textit{fixed }\nu}$$

$$\lim_{N\to\infty} N \cdot e_N^{\textit{fixed}} = C^{\textit{fixed}}$$

$$\lim_{N\to\infty} N \cdot e_N^{\textit{equi}} = C^{\textit{equi}}$$

Note: Constants are alltogether either zero or positve
In the latter case the order of convergence is $1/N$ for all classes, however

$$C \le C^{\text{fixed }\nu} \le C^{\text{fixed}} \le C^{\text{equi}}$$

with strict inequality in most cases, see Remark 13

Example 16. (Additive noise)

$$dX(t) = a(t)\,dt + \sigma(t)\,dW(t), \quad X(0) = 0,$$

Here

$$\eta = -\sigma', \quad \Phi(t) \equiv 1$$

Thus deterministic weighting process

$$\Theta = \eta = -\sigma'$$

yields

$$C = C^{\text{fixed }\nu} = C^{\text{fixed}} = \frac{1}{\sqrt{12}} \cdot \Big(\int_0^1 |\sigma'(t)|^{2/3}\,dt\Big)^{3/2}$$

$$C^{\text{equi}} = \frac{1}{\sqrt{12}} \cdot \Big(\int_0^1 |\sigma'(t)|^2\,dt\Big)^{1/2}$$

similar to mean square L_2-approximation

Example 17. (Linear equation)

$$dX(t) = \alpha(t) \cdot X(t)\,dt + \beta(t) \cdot X(t)\,dW(t), \quad X(0) = 1,$$

Solution

$$X(t) = \exp\Big(\int_0^t (\alpha - 1/2 \cdot \beta^2)(u)\ du + \int_0^t \beta(u)\ dW(u)\Big)$$

Here

$$\eta(t, x) = -\beta'(t) \cdot x$$

and

$$\Phi(t) = X(1)/X(t)$$

Weighting process

$$\Theta(t) = -\beta'(t) \cdot X(1)$$

yields constants

$$C = \frac{1}{\sqrt{12}} \cdot \|\beta'\|_{2/3} \cdot \exp\big(\|\alpha\|_1 - 1/6 \cdot \|\beta\|_2^2\big)$$

$$C^{\text{fixed }\nu} = C^{\text{fixed}} = \frac{1}{\sqrt{12}} \cdot \|\beta'\|_{2/3} \cdot \exp\big(\|\alpha\|_1 + 1/2 \cdot \|\beta\|_2^2\big)$$

$$C^{\text{equi}} = \frac{1}{\sqrt{12}} \cdot \|\beta'\|_2 \cdot \exp\big(\|\alpha\|_1 + 1/2 \cdot \|\beta\|_2^2\big)$$

For instance, if $\alpha = 0$ and $\beta(t) = b \cdot t$ then

$$C = |b|/\sqrt{12} \cdot \exp(-b^2/18)$$

$$C^{\text{fixed }\nu} = C^{\text{fixed}} = C^{\text{equi}} = |b|/\sqrt{12} \cdot \exp(b^2/6)$$

C is exponentially decreasing in b while the other constants are exponentially increasing in b.

Adapting the number ν of evaluations of W to the trajectory of the solution X is essential here.

Remark 12.

Theorem 10 determines rate of convergence of minimal errors only for nonzero constants.

Clearly: $\quad C = C^{\text{fixed }\nu} = C^{\text{fixed}} = C^{\text{equi}} = 0 \quad$ iff

(9) $$P(\forall t \in [0,1] :\ \eta(t, X(t)) = 0) = 1$$

Conjecture:

$$(9) \quad \text{iff} \quad \exists \text{ measurable } f : [0,1] \times \mathbb{R} \to \mathbb{R} : \quad P(\forall t \in [0,1] : X(t) = f(t, W(t)) = 1$$

If the conjecture is true then

$$e_N \asymp 1/N \quad \text{or} \quad e_N = e_1 = 0.$$

The conjecture is known to be true in many cases:

- $\inf_{t,x} |\sigma(t,x)| > 0$
  (Clark and Cameron 1980)
- a and σ have parial derivatives of any order
  (Yamoto 1979)
- For the linear equation from Example 17
  ((9) implies $\beta' = 0$)

Remark 13. (Equality of asymptotic constants)

$$C^{\text{equi}} = C^{\text{fixed}}$$
$$\text{iff} \quad \exists\, \theta \in \mathbb{R}\ \forall t \in [0,1] : E(\Theta^2(t) = \theta$$

$$C^{\text{fixed}} = C^{\text{fixed}\ \nu}$$
$$\text{iff} \quad \exists\, t_0 \in [0,1],\ \theta \in C([0,1]) : \quad P\big(\forall t \in [0,1] : \Theta(t) = \theta(t) \cdot \Theta(t_0)\big) = 1$$

$$C^{\text{fixed}\ \nu} = C$$
$$\text{iff} \quad \exists\, \theta \in C([0,1]) : \ P\big(\forall t \in [0,1] : \Theta(t) = \theta(t)\big) = 1$$

See Examples 16 and 17

Remark 14. (mean square approximation at a single point and weighted integration)

Consider integral of weighted Brownian motion

$$Y = \int_0^1 \rho(t) \cdot W(t)$$

with continuous $\rho : [0,1] \to [0,\infty[$.

Best mean square approximation of Y based on $W(t_1), \ldots, W(t_N)$:

$$\widehat{Y}_{(t_1,\ldots,t_N)} = E\big(Y \,\big|\, W(t_1), \ldots, W(t_N)\big)$$

Resulting minimal mean square error:

$$\widetilde{e}_N = \inf\Big\{ \big(E\big|Y - \widehat{Y}_{(t_1,\ldots,t_N)}\big|^2\big)^{1/2} : \ 0 < t_1 < \ldots < t_N \le 1 \Big\}$$

Note: Adaption does not help here since Y is Gaussian

From general results on weighted integration (see Ritter 2000):

$$\lim_{N\to\infty} N \cdot \widetilde{e}_N = C_\rho := \frac{1}{\sqrt{12}} \cdot \left(\int_0^1 |\rho(t)|^{2/3}\, dt \right)^{3/2}$$

We have

$$C^{\text{fixed}} = C_{(E|\Theta(\cdot)|^2)^{1/2}}$$

Taking the *random weight* $\rho = \Theta$ yields

$$C = \Big(E\big(C_\Theta^{2/3}\big)\Big)^{3/2}$$

The adaptive approximation $\widehat{X}_n(1)$

Basic idea:

- Replace integrals
$$J_\ell = \int_{t_\ell}^{t_{\ell+1}} \big(W(t) - W(t_\ell)\big)\, dt$$
in the definition of the Wagner-Platen scheme by suitable approximations based on point evaluations of W.
- Adapt the discretization to the random weight Θ:

  The *larger* the value of $\Theta(t)$
  the *finer* the discretization around t

Algorithm (essential structure):

1) Dependent on n determine *coarse* equidistant prediscretization
$$t_\ell = \ell/k_n, \quad \ell = 0, 1, \ldots, k_n$$
2) Compute corresponding *truncated* Wagner-Platen scheme
$$\widehat{X}^{\text{trunc}}(t_\ell), \quad \ell = 0, 1, \ldots, k_n$$
3) Use *truncated* Wagner-Platen scheme to compute a discrete approximation of the random weight
$$\widehat{\Theta}(t_\ell), \quad \ell = 0, 1, \ldots, k_n - 1$$

4) Determine number of additional equidistant points in subintervals $]t_\ell, t_{\ell+1}[$ by
$$\mu_\ell = \left\lceil n/k_n \cdot |\widehat{\Theta}(t_\ell)|^{2/3} \right\rceil$$
Resulting points in $[t_\ell, t_{\ell+1}]$
$$\tau_{\ell,j} = t_\ell + j \cdot (k_n \cdot (\mu_\ell + 1))^{-1}, \quad j = 0, 1, \ldots, \mu_\ell + 1$$
5) Estimate integrals J_ℓ by (adaptive) trapezoidal rule
$$\widehat{J}_\ell = \frac{1}{2k_n \cdot (\mu_\ell + 1)} \cdot \sum_{j=0}^{\mu_\ell} \big(W(\tau_{\ell,j}) + W(\tau_{\ell,j+1}) - 2W(t_\ell)\big)$$
6) Compute approximation at $t = 1$
$$\widehat{X}_n(1) = \widehat{X}^{\text{trunc}}(1) + \sum_{\ell=0}^{k_n-1} \widehat{\Theta}(t_\ell) \cdot \widehat{J}_\ell$$

Note that $\widehat{X}_n(1) \in \mathbb{X}$.

Missing details:

ad 1) Choose k_n such that
$$\lim_{n\to\infty} k_n/n = 0$$
(keeping track of size of random weight)

and
$$\lim_{n\to\infty} k_n^{3/2}/n = \infty$$
(estimating size of random weight sufficiently well)

ad 2) *Truncated* Wagner-Platen scheme

$$
\begin{aligned}
&\widehat{X}^{\text{trunc}}(t_0) = x_0,\\
&\widehat{X}^{\text{trunc}}(t_{\ell+1})\\
&\quad = \widehat{X}^{\text{trunc}}(t_\ell) + a(t_\ell, \widehat{X}^{\text{trunc}}(t_\ell)) \cdot (t_{\ell+1} - t_\ell)\\
&\qquad\qquad + \sigma(t_\ell, \widehat{X}^{\text{trunc}}(t_\ell)) \cdot (W(t_{\ell+1}) - W(t_\ell))\\
&\qquad + 1/2 \cdot \big(\sigma^{(0,1)} \cdot \sigma\big)(t_\ell, \widehat{X}^{\text{trunc}}(t_\ell))\\
&\qquad\qquad \times \big((W(t_{\ell+1}) - W(t_\ell))^2 - (t_{\ell+1} - t_\ell)\big)\\
&\qquad + \big(\sigma^{(1,0)} + a \cdot \sigma^{(0,1)} - \sigma/2 \cdot (\sigma^{(0,1)})^2\big)(t_\ell, \widehat{X}^{\text{trunc}}(t_\ell))\\
&\qquad\qquad \times \big(W(t_{\ell+1}) - W(t_\ell)\big) \cdot (t_{\ell+1} - t_\ell)\\
&\qquad + 1/6 \cdot \big(\sigma \cdot (\sigma^{(0,1)})^2 + \sigma^2 \cdot \sigma^{(0,2)}\big)(t_\ell, \widehat{X}^{\text{trunc}}(t_\ell))\\
&\qquad\qquad \times \big(W(t_{\ell+1}) - W(t_\ell)\big)^3\\
&\qquad + 1/2 \cdot \big(a^{(1,0)} + a \cdot a^{(0,1)} + \sigma^2/2 \cdot a^{(0,2)}\big)(t_\ell, \widehat{X}^{\text{trunc}}(t_\ell))\\
&\qquad\qquad \times (t_{\ell+1} - t_\ell)^2\\
&\quad = \widehat{X}^{\text{trunc}}(t_\ell) + \sum_{\alpha \in A_{3/2} \setminus \{(1,0),(0,1)\}} f_\alpha\big(t_\ell, \widehat{X}^{\text{trunc}}(t_\ell)\big) \cdot I_{\alpha, t_\ell, t_{\ell+1}}\\
&\qquad + f_{(0,1)}\big(t_\ell, \widehat{X}^{\text{trunc}}(t_\ell)\big) \cdot (t_{\ell+1} - t_\ell) \cdot (W(t_{\ell+1}) - W(t_\ell))
\end{aligned}
$$

Note: This scheme uses only point evaluations of W

ad 3) Estimation of random weight

Recall

$$\Theta(t) = \eta(t, X(t)) \cdot \Phi(t)$$

Consider random field

$$\Psi(t, s) = \Phi(t)/\Phi(s), \quad t \le s \le 1$$

Note: $\Phi(t) = \Psi(t, 1)$ and $\Psi(t, \cdot)$ satisfies

$$
\begin{aligned}
d\Psi(t, s) &= a^{(0,1)}(s, X(s)) \cdot \Psi(t, s)\, ds\\
&\quad + \sigma^{(0,1)}(s, X(s)) \cdot \Psi(t, s)\, dW(s), \quad t \le s \le 1,
\end{aligned}
$$

Euler-type scheme for approximation of $\Psi(t_\ell, \cdot)$

$$
\begin{aligned}
&\widehat{\Psi}(t_\ell, t_\ell) = 1,\\
&\widehat{\Psi}(t_\ell, t_{j+1}) = \widehat{\Psi}(t_\ell, t_j)\\
&\qquad + a^{(0,1)}(t_j, \widehat{X}^{\text{trunc}}(t_j)) \cdot \widehat{\Psi}(t_\ell, t_j) \cdot (t_{j+1} - t_j)\\
&\qquad + \sigma^{(0,1)}(t_j, \widehat{X}^{\text{trunc}}(t_j)) \cdot \widehat{\Psi}(t_\ell, t_j) \cdot (W(t_{j+1}) - W(t_j))
\end{aligned}
$$

for $j = \ell, \ell + 1, \ldots, k_n - 1$ yields approximation of $\Phi(t_\ell)$

$$\widehat{\Phi}(t_\ell) = \widehat{\Psi}(t_\ell, 1)$$

Use

$$\widehat{\Theta}(t_\ell) = \eta(t_\ell, \widehat{X}^{\text{trunc}}(t_\ell)) \cdot \widehat{\Phi}(t_\ell), \quad \ell = 0, 1, \ldots, k_n - 1.$$

Recall from Theorem 10

$$\lim_{n\to\infty} N \cdot e_N = C$$

with

$$C = \frac{1}{\sqrt{12}} \cdot \left(E \int_0^1 |\Theta(t)|^{2/3} \, dt \right)^{3/2}$$

Theorem 1. *(M-G 2004)*

(i) $$\lim_{n\to\infty} c\big(\widehat{X}_n\big) \cdot e\big(\widehat{X}_n\big) = C,$$

(ii) $$\lim_{n\to\infty} n^{-1} \cdot c\big(\widehat{X}_n\big) = \sqrt{12} \cdot C.$$

Corollary 3. *If $C > 0$ then $\widehat{X}_n(1)$ is asymptotically optimal in the class $\mathbb{X}$ for mean square approximation of $X(1)$.*

Remark 15. (Asymptotically optimal methods for subclasses)

Same definition as $\widehat{X}_n(1)$ with numbers μ_ℓ (essentially) replaced by

$$\mu_\ell = \left\lceil n \cdot |\widehat{\Theta}(t_\ell)|^{2/3} \cdot \Big(\sum_{j=0}^{k_n-1} |\widehat{\Theta}(t_j)|^{2/3} \Big)^{-1} \right\rceil \qquad \text{for } \mathbb{X}^{\text{fixed } \nu},$$

$$\mu_\ell = \left\lceil n \cdot \big(E|\Theta(t_\ell)|^2\big)^{1/3} \cdot \Big(\sum_{j=0}^{k_n-1} \big(E|\Theta(t_j)|^2\big)^{1/3} \Big)^{-1} \right\rceil \qquad \text{for } \mathbb{X}^{\text{fixed}},$$

and $\mu_\ell = 0$ for the class $\mathbb{X}^{\text{equi}}$.

Remark 16. (Performance of Milstein scheme)

The Milstein scheme is, in general, not asymptotically optimal for approximation of the solution at a single point.

Example:

$$dX(t) = \sigma(t)\, dW(t), \quad X(0) = 0$$

Milstein scheme based on fixed discretization

$$0 = t_0 < t_1 < \ldots < t_n = 1$$

is given by

$$\widehat{X}^M_{(t_1,\ldots,t_n)}(t_\ell) = \sum_{j=0}^{\ell-1} \sigma(t_j) \cdot (W(t_{j+1}) - W(t_j))$$

and satisfies

$$e\big(\widehat{X}^M_{(t_1,\ldots,t_n)}\big) = \frac{1}{3} \sum_{\ell=0}^{n-1} \big(\sigma'(t_\ell)\big)^2 \cdot (t_{\ell+1} - t_\ell)^3.$$

It follows

$$\lim_{n\to\infty} n \cdot \inf_{0<t_1<\cdots<t_n=1} e\big(\widehat{X}^M_{(t_1,\ldots,t_n)}\big) = 2 \cdot C^{\text{fixed}}$$

and

$$\lim_{n\to\infty} n \cdot e\big(\widehat{X}^M_{(1/n,2/n,\ldots,1)}\big) = 2 \cdot C^{\text{equi}}$$

V. Outlook

	Treated in lecture	What about?
global error	$\left(E\|X-\widehat{X}\|_2^2\right)^{1/2}$	$\left(E\|X-\widehat{X}\|_p^q\right)^{1/q}$
error at $t=1$	$\left(E\lvert X(1)-\widehat{X}(1)\rvert^2\right)^{1/2}$	$\left(E\lvert X(1)-\widehat{X}(1)\rvert^q\right)^{1/q}$
information about W	point evaluations	cont. lin. functionals iterated Itô-integrals
dimension	scalar equations	systems

Optimal global approximation based on

- point evaluations of W:

  Hofmann, M-G, Ritter 2001: $p=q=2$, scalar equations

  M-G 2002a: $p<\infty$, $q\geq 1$, systems

  M-G 2002b: $p=\infty$, $q\geq 1$, systems
- continuous linear functionals of W:

  Hofmann, M-G, Ritter 2002: $p=q=2$, scalar equations
- iterated Itô-integrals:

  Hofmann, M-G 2004: $p=q=2$, scalar equations

L_p-approximation of scalar equations:

Define

$$C_{p,q}(a,\sigma,x_0)=\left(E\left\|\sigma(\cdot,X(\cdot))\right\|_{2p/(p+2)}^{2q/(q+2)}\right)^{(q+2)/2q},$$

for $q\in[1,\infty[$, $p\in[1,\infty]$ $(2\infty/(\infty+2):=2)$, and put

$$m_p=\left(\int_{-\infty}^{\infty}|z|^p\cdot(2\pi)^{-1/2}\cdot\exp(-z^2/2)\,dz\right)^{1/p}$$

for $p\in[1,\infty[$.

info about W	min. errors order	asymp. const.	asymp. opt. meth. essentially
point evaluation $1\leq p=q<\infty$	$N^{-1/2}$	$\frac{m_p}{\sqrt{6}}\cdot C_{p,q}(a,\sigma,x_0)$	Milstein with stepsize $\sim\lvert\sigma(t,X(t))\rvert^{-2p/(p+2)}$
$p=\infty$ $1\leq q<\infty$	$(N/\ln N)^{-1/2}$	$\frac{1}{\sqrt{2}}\cdot C_{\infty,q}(a,\sigma,x_0)$	Euler with stepsize $\sim\lvert\sigma(t,X(t))\rvert^{-2}$
cont. lin. func. $p=q=2$	$N^{-1/2}$	$\frac{1}{\pi}\cdot C_{2,2}(a,\sigma,x_0)$	Uses adaptive Karhunen-Loeve expansion of W
iter. Itô-integrals $p=q=2$	$N^{-1/2}$ (non-adaptive)	?	?

General picture seems to be:

- order of convergence of minimal errors

$$= \begin{cases} N^{-1/2} & \text{if } p < \infty \\ (N/\ln N)^{-1/2} & \text{if } p = \infty \end{cases}$$

- asymptotic constant

$$= \kappa_{p,q}(\text{info}) \cdot C_{p,q}(a, \sigma, x_0)$$

  where $\kappa_{p,q}(\text{info})$ only depends on error parameters p, q, and the type of information about W.

- key quantity for asymptotic constants and for construction of asymptotically optimal methods

$$= \text{ conditional Hölder constant } |\sigma(t, X(t))|$$

L_p-approximation of systems based on point evaluations of W:

- order of convergence of minimal errors as above
- key quantities for asymptotic constants and for construction of asymptotically optimal methods:

  Conditional Hölder constants of components of solution
  and
  Degree of non-commutativity of diffusion coefficient

Optimal approximation at $t = 1$ based on

- point evaluations of W:

  M-G 2004: $q \geq 1$, scalar equations
  M-G 2002a: $q = 2$, non-commutative systems

- continuous linear functionals of W or iterated Itô-integrals:

  Only upper bounds from Theorem 5, see Part II.

Scalar equations and point evaluations of W:

- order of convergence of minimal errors

$$= N^{-1} \quad \text{(up to specific cases)}$$

- key quantity for asymptotic constants and for construction of asymptotically optimal methods:

  random weight composed of Itô-Taylor coefficient functions and means square derivative of solution

Non-commutative systems and point evaluations of W:

- order of convergence of minimal errors

$$= N^{-1/2}$$

- key quantity for asymptotic constants and for construction of asymptotically optimal methods:

  random weight composed of degree of non-commutativity and means square derivative of solution

Appendix A. Proof of Theorem 4(*i*)

Define a continuous version of the Milstein scheme by

$$\begin{aligned}\widetilde{X}^M(t_0) &= x_0,\\ \widetilde{X}^M(t) &= \widehat{X}^M(t_\ell) + a(t_\ell, \widehat{X}^M(t_\ell)) \cdot (t - t_\ell)\\ &\quad + \sigma(t_\ell, \widehat{X}^M(t_\ell)) \cdot (W(t) - W(t_\ell))\\ &\quad + 1/2 \cdot \big(\sigma^{(0,1)} \cdot \sigma\big)(t_\ell, \widehat{X}^M(t_\ell))\\ &\qquad \times \big((W(t) - W(t_\ell))^2 - (t - t_\ell)\big).\end{aligned}$$

for $t \in [t_\ell, t_{\ell+1}]$, $\ell = 0, 1, \ldots, n-1$.

Note: $\quad \widetilde{X}^M(t_\ell) = \widehat{X}^M(t_\ell)$

Remark 17. The process $\widetilde{X}^M$ is not a numerical method for global approximation of X since the complete trajectories of W are needed for its construction.

Notation $\quad |Y|_q := (E|Y|^q)^{1/q}$ for a random variable Y

Instead of Theorem 4(i) we prove the stronger

Theorem 12. *Under the assumptions of Theorem 4*

$$\sup_{t\in[0,1]} |X(t) - \widetilde{X}^M(t)|_q \le c \cdot \Delta_{\max}.$$

The main tool for the proof is

Lemma 1. *(Gronwall's inequality)*
Let $\alpha, \beta \in \mathbb{R}$ with $\beta \ge 0$, and $f : [0,1] \to [0,\infty[$ be bounded and Borel-measurable. If

$$\forall t \in [0,1]: \qquad f(t) \le \alpha + \beta \cdot \int_0^t f(s)\, ds$$

then

$$\sup_{t\in[0,1]} f(t) \le \alpha \cdot \exp(\beta).$$

For a proof see, e.g., Revuz and Yor 1991.

We start with moment inequalities for $\widetilde{X}^M$.

Lemma 2. *Under the assumptions of Theorem 4*

(i) $$\sup_{t\in[0,1]} |\widetilde{X}^M(t)|_q \le c,$$

(ii) $$\sup_{t\in[0,1]} |\widetilde{X}^M(s) - \widetilde{X}^M(t)|_q \le c \cdot |s-t|^{1/2},$$

Proof. Define $f : [0,1] \to [0,\infty]$ by

$$f(t) = \sup_{s\in[0,t]} |\widetilde{X}^M(s)|_q .$$

Clearly, f is Borel-measurable. We show that f is bounded. Let $t \in [t_\ell, t_{\ell+1}]$ and observe the regularity conditions on a and σ to obtain

$$
\begin{aligned}
(10)\quad & |\widetilde{X}^M(t) - \widetilde{X}^M(t_\ell)|_q \\
&\le |a(t_\ell, \widetilde{X}^M(t_\ell)) \cdot (t - t_\ell)|_q \\
&\qquad + |\sigma(t_\ell, \widetilde{X}^M(t_\ell)) \cdot (W(t) - W(t_\ell))|_q \\
&\qquad + \big|\big(\sigma \cdot \sigma^{(0,1)}\big)(t_\ell, \widetilde{X}^M(t_\ell)) \\
&\qquad\qquad \times \big((W(t) - W(t_\ell))^2 - (t - t_\ell)\big)\big|_q \\
&\le (t - t_\ell) \cdot |a(t_\ell, \widetilde{X}^M(t_\ell))|_q \\
&\qquad + c \cdot (t - t_\ell)^{1/2} \cdot |\sigma(t_\ell, \widetilde{X}^M(t_\ell))|_q \\
&\le c \cdot (t - t_\ell)^{1/2} \cdot \big(1 + |\widetilde{X}^M(t_\ell)|_q\big),
\end{aligned}
$$

which yields $f(1) < \infty$ since $|\widetilde{X}^M(t_0)|_q = |x_0|$.

Next, we use Gronwall's inequality to prove an upper bound for f, which does not depend on the discretization. Put

$$
Z(s) = \sum_j \sigma(t_j, \widetilde{X}^M(t_j)) \cdot 1_{]t_j, t_{j+1}]}(s)
$$

and note that

$$
|Z(s)|_q \le c \cdot \big(1 + \sup_{u \in [0,s]} |\widetilde{X}^M(u)|_q\big) \le c \cdot (1 + f(s)).
$$

Assume $q \ge 2$. By the Burkholder martingale moment inequality,

$$
\begin{aligned}
|\widetilde{X}^M(t)|_q &\le |x_0| + c \cdot \sum_{j=0}^{\ell-1} \big(1 + |\widetilde{X}^M(t_j)|_q\big) \cdot (t_{j+1} - t_j) \\
&\qquad + c \cdot \big(1 + |\widetilde{X}^M(t_\ell)|_q\big) \cdot (t - t_\ell) + \Big| \int_0^t Z(s)\, dW(s) \Big|_q
\end{aligned}
$$

$$
\begin{aligned}
&\le c \cdot \Big(1 + \int_0^t f(s)\, ds\Big) + c \cdot \Big(\int_0^t |Z(s)|_q^q\, ds\Big)^{1/q} \\
&\le c \cdot \Big(1 + \int_0^t f(s)\, ds + \Big(\int_0^t (f(s))^q\, ds\Big)^{1/q}\Big) \\
&\le c \cdot \Big(1 + \int_0^t (f(s))^q\, ds\Big)^{1/q}.
\end{aligned}
$$

We conclude that

$$
f^q(t) \le c + c \cdot \int_0^t f^q(s)\, ds
$$

for all $t \in [0,1]$, which implies inequality (i) by Gronwall's Lemma. Clearly, inequality (ii) is a consequence of (10) and inequality (i). $\square$

Next, let

$$
V(t) = \sum_\ell \big(a^{(0,1)} \cdot \sigma\big)(t_\ell, \widetilde{X}^M(t_\ell)) \cdot (W(t) - W(t_\ell)) \cdot 1_{]t_j, t_{j+1}]}(t)
$$

and define

$$
U(t) = \int_0^t V(s)\, ds.
$$

Lemma 3. *Under the assumptions of Theorem 4*

$$
\sup_{t \in [0,1]} |U(t)|_q \le c \cdot \Delta_{\max}.
$$

Proof. Assume $q \in N$ and let $(\mathcal{F}_t)_{0\le t\le 1}$ denote the filtration that is generated by W. Let $t \in [t_\ell, t_{\ell+1}]$. By the regularity properties of a and σ,

$$\begin{aligned}
E\big(|U(t)|^{2q}\big|\mathcal{F}_{t_\ell}\big)
&= E\bigg(\bigg(U(t_\ell) + \int_{t_\ell}^t V(s)\,ds\bigg)^{2q}\bigg|\mathcal{F}_{t_\ell}\bigg) \\
&= \sum_{r=0}^{2q}\binom{2q}{r}\cdot U^{2q-r}(t_\ell)\cdot E\bigg(\bigg(\int_{t_\ell}^t V(s)\,ds\bigg)^r\bigg|\mathcal{F}_{t_\ell}\bigg) \\
&= \sum_{r=0}^{2q}\binom{2q}{r}\cdot U^{2q-r}(t_\ell)\cdot \big(a^{(0,1)}\cdot\sigma\big)^r(t_\ell,\widetilde{X}^M(t_\ell)) \\
&\qquad \times E\bigg(\int_{t_\ell}^t (W(s)-W(t_\ell))\,ds\bigg)^r \\
&\le U^{2q}(t_\ell) + c\cdot\sum_{r=1}^{q}\binom{2q}{2r}\cdot U^{2q-2r}(t_\ell) \\
&\qquad \times \big(a^{(0,1)}\cdot\sigma\big)^{2r}(t_\ell,\widetilde{X}^M(t_\ell))\cdot(t-t_\ell)^{3r} \\
&\le U^{2q}(t_\ell) + c\cdot(t-t_\ell)\cdot\sum_{r=1}^{q}\binom{2q}{2r}\cdot U^{2q-2r}(t_\ell) \\
&\qquad \times \big((t-t_\ell)\cdot(1+|\widetilde{X}^M(t_\ell)|)\big)^{2r}
\end{aligned}$$

Lemma 2 yields

$$\begin{aligned}
E\big(U^{2q-2r}(t_\ell)\cdot(1+|\widetilde{X}^M(t_\ell)|)^{2r}\big)
&\le |U(t_\ell)|_{2q}^{2q-2r}\cdot\big|1+|\widetilde{X}^M(t_\ell)|\big|_{2q}^{2r} \\
&\le c\cdot|U(t_\ell)|_{2q}^{2q-2r}.
\end{aligned}$$

Hence

$$\begin{aligned}
&|U(t)|_{2q}^{2q} \\
&\quad\le |U(t_\ell)|_{2q}^{2q} + c\cdot(t-t_\ell)\cdot\sum_{r=1}^{q}\binom{2q}{2r}\cdot|U(t_\ell)|_{2q}^{2q-2r}\cdot(t-t_\ell)^{2r} \\
&\quad\le |U(t_\ell)|_{2q}^{2q} + c\cdot(t-t_\ell)\cdot\big(|U(t_\ell)|_{2q}+(t-t_\ell)\big)^{2q} \\
&\quad\le |U(t_\ell)|_{2q}^{2q}\cdot(1+c\cdot(t-t_\ell)) + c\cdot(t-t_\ell)^{2q+1},
\end{aligned}$$

which implies

$$\sup_{s\in[0,t]}|U(s)|_{2q}^{2q} \le c\cdot\Delta_{\max}^{2q} + c\cdot\int_0^t \sup_{s\in[0,u]}|U(s)|_{2q}^{2q}\,du$$

for every $t\in[0,1]$. Use Gronwall's Lemma to complete the proof. □

Proof of Theorem 12.

By definition, $X(t) = x_0 + A(t) + \Sigma(t)$, where

$$A(t) = \int_0^t a(s,X(s))\,ds, \quad \Sigma(t) = \int_0^t \sigma(s,X(s))\,dW(s).$$

Similarly, $\widetilde{X}^M(t) = x_0 + \widetilde{A}(t) + \widetilde{\Sigma}(t)$, where

$$\widetilde{A}(t) = \int_0^t \sum_{\ell=0}^{n-1} a(t_\ell,\widetilde{X}^M(t_\ell))\cdot 1_{]t_\ell,t_{\ell+1}]}(s)\,ds$$

$$\widetilde{\Sigma}(t) = \int_0^t \sum_{\ell=0}^{n-1} \sigma(t_\ell, \widetilde{X}^M(t_\ell)) \cdot 1_{]t_\ell, t_{\ell+1}]}(s)\, dW(s)$$
$$+ \int_0^t \sum_{\ell=0}^{n-1} \big(\sigma \cdot \sigma^{(0,1)}\big)(t_\ell, \widetilde{X}^M(t_\ell)) \cdot (W(s) - W(t_\ell)) \times \cdot 1_{]t_\ell, t_{\ell+1}]}(s)\, dW(s).$$

Fix $s \in]t_\ell, t_{\ell+1}]$ and recall the above definition of $V(s)$. By the regularity properties of a and σ,

$$\begin{aligned}
&|a(s, X(s)) - a(t_\ell, \widetilde{X}^M(t_\ell)) - V(s)| \\
&\quad \le |a(s, X(s)) - a(t_\ell, X(s))| + |a(t_\ell, X(s)) - a(t_\ell, \widetilde{X}^M(s))| \\
&\qquad + |a(t_\ell, \widetilde{X}^M(s)) - a(t_\ell, \widetilde{X}^M(t_\ell)) \\
&\qquad\qquad - a^{(0,1)}(t_\ell, \widetilde{X}^M(t_\ell)) \cdot (\widetilde{X}^M(s) - \widetilde{X}^M(t_\ell))| \\
&\qquad + \big|a^{(0,1)}(t_\ell, \widetilde{X}^M(t_\ell)) \cdot \big(\widetilde{X}^M(s) - \widetilde{X}^M(t_\ell) \\
&\qquad\qquad - \sigma(t_\ell, \widetilde{X}^M(t_\ell)) \cdot (W(s) - W(t_\ell))\big)\big| \\
&\quad \le c \cdot (1 + |X(s)|) \cdot (s - t_\ell) + c \cdot |X(s) - \widetilde{X}^M(s)| \\
&\qquad + c \cdot |\widetilde{X}^M(s) - \widetilde{X}^M(t_\ell)|^2 \\
&\qquad + c \cdot \big(1 + |\widetilde{X}^M(t_\ell)|\big) \cdot \big((s - t_\ell) + (W(s) - W(t_\ell))^2\big).
\end{aligned}$$

Hence, by property (6) and Lemma 2,

$$\begin{aligned}
&|a(s, X(s)) - a(t_\ell, \widetilde{X}^M(t_\ell)) - V(s)|_q^q \\
&\quad \le c \cdot \big((1 + |X(s)|_q^q + |\widetilde{X}^M(t_\ell)|_q^q) \cdot \Delta_{\max}^q + |X(s) - \widetilde{X}^M(s)|_q^q\big) \\
&\quad \le c \cdot \big(\Delta_{\max}^q + |X(s) - \widetilde{X}^M(s)|_q^q\big)
\end{aligned}$$

Use Lemma 3 and the Hölder inequality to conclude

$$\begin{aligned}
&|A(t) - \widetilde{A}(t)|_q^q \\
&\quad \le c \cdot \int_0^t \sum_{\ell=0}^{n-1} |a(s, X(s)) - a(t_\ell, \widetilde{X}^M(t_\ell)) - V(s)|_q^q \times 1_{]t_\ell, t_{\ell+1}]}(s)\, ds \\
&\qquad + c \cdot |U(t)|_q^q \\
&\quad \le c \cdot \Delta_{\max}^q + c \cdot \int_0^t |X(s) - \widetilde{X}^M(s)|_q^q\, ds. \qquad (11)
\end{aligned}$$

Similarly, for $s \in]t_\ell, t_{\ell+1}]$,

$$\begin{aligned}
&\big|\sigma(s, X(s)) - \sigma(t_\ell, \widetilde{X}^M(t_\ell)) \\
&\qquad - \big(\sigma \cdot \sigma^{(0,1)}\big)(t_\ell, \widetilde{X}^M(t_\ell)) \cdot (W(s) - W(t_\ell))\big|_q^q \\
&\quad \le c \cdot \Delta_{\max}^q + c \cdot |X(s) - \widetilde{X}^M(s)|_q^q.
\end{aligned}$$

Thus, by the Burkholder inequality (assume $q \ge 2$),

$$|\Sigma(t) - \widetilde{\Sigma}(t)|_q^q \le c \cdot \Delta_{\max}^q + c \cdot \int_0^t |X(s) - \widetilde{X}^M(s)|_q^q\, ds. \qquad (12)$$

Combine (11) with (12) to obtain

$$|X(t) - \widetilde{X}^M(t)|_q^q \le c \cdot \Delta_{\max}^q + c \cdot |X(s) - \widetilde{X}^M(s)|_q^q$$

for every $t \in [0,1]$. It remains to apply Gronwall's Lemma to finish the proof. □

Appendix B. Proof of Theorems 6 and 8

We start with a result on approximation of Brownian bridges:

Lemma 4. *Consider a Brownian bridge B on an interval $[S,T]$ and let $\widetilde{B}$ denote the piecewise linear interpolation of B at the points*

$$s_\ell = \ell \cdot (T-S)/(m+1), \quad \ell = 1, \dots, m.$$

Then

$$\int_S^T E\big(B(t) - \widetilde{B}(t)\big)^2 \, dt = \frac{(T-S)^2}{6 \cdot (m+1)}.$$

Proof. Straightforward, using the fact that

$$E\big(B(t) - \widetilde{B}(t)\big)^2 = \frac{(s_{\ell+1} - t) \cdot (t - s_\ell)}{s_{\ell+1} - s_\ell}$$

for $t \in [s_\ell, s_{\ell+1}]$. □

Proof of the lower bound in Theorem 8

Consider an arbitrary sequence of approximations

$$\overline{X}_N = \widehat{X}_{(\psi^{(N)}, \chi^{(N)}, \varphi^{(N)})} \in \mathbb{X}$$

with

$$c(\overline{X}_N) \le N.$$

Let

$$D^{(N)} = D_{\nu_{(\psi^{(N)}, \chi^{(N)}, \varphi^{(N)})}}$$

denote the corresponding underlying Brownian data and define

$$Z_N(t) = W(t) - E(W(t)|D^{(N)}), \quad t \in [0,1].$$

Take a sequence of positive integers k_N such that

$$\lim_{N\to\infty} N^{1/2}/k_N = \lim_{N\to\infty} k_N/N = 0. \qquad (13)$$

Since $k_N = o(N)$ we may assume that $\overline{X}_N$ uses in particular the knots

$$t_\ell^{(N)} = \ell/k_N, \quad \ell = 0, 1, \dots, k_N.$$

Let $\widetilde{X}_N^M$ denote the corresponding time-continuous Milstein scheme, see Appendix A, and put

$$U_\ell^{(N)} = \big(t_\ell^{(N)}, \widetilde{X}_N^M(t_\ell^{(N)})\big).$$

Lemma 5.

$$\liminf_{N\to\infty} N \cdot e^2(\overline{X}_N) \ge \liminf_{N\to\infty} N \cdot \sum_{\ell=0}^{k_N-1} \int_{t_\ell^{(N)}}^{t_{\ell+1}^{(N)}} E\big|\sigma(U_\ell^{(N)}) \cdot Z_N(t)\big|^2 \, dt$$

Proof. Theorem 12 implies

$$E\big\|X - \widetilde{X}_N^M\big\|_2^2 \le c/k_N^2.$$

Since $c/k_N^2 = o(N^{-1})$ by (13), it therefore suffices to analyze the difference $\widetilde{X}_N^M - \overline{X}_N$.

Clearly,

$$E\big|\widetilde{X}_N^M(t) - \overline{X}_N(t)\big|^2 \geq E\big|\widetilde{X}_N^M(t) - E(\widetilde{X}_N^M(t)|D^{(N)})\big|^2$$

Let $t \in [t_\ell^{(N)}, t_{\ell+1}^{(N)}]$. By definition of $\widetilde{X}_N^M$,

$$\begin{aligned}
&\widetilde{X}_N^M(t) - E(\widetilde{X}_N^M(t)\big|D^{(N)}) \\
&\quad = \sigma(U_\ell^{(N)}) \cdot Z_N(t) + 1/2 \cdot (\sigma \cdot \sigma^{(0,1)})(U_\ell^{(N)}) \\
&\qquad \times \big((W(t) - W(t_\ell^{(N)}))^2 - E((W(t) - W(t_\ell^{(N)}))^2|D^{(N)})\big)
\end{aligned}$$

By boundedness of $\sigma^{(0,1)}$ and linear growth of σ,

$$\begin{aligned}
&E\big|(\sigma \cdot \sigma^{(0,1)})(U_\ell^{(N)}) \\
&\qquad \times \big((W(t) - W(t_\ell^{(N)}))^2 - E((W(t) - W(t_\ell^{(N)}))^2|D^{(N)})\big)\big|^2 \\
&\quad \leq c \cdot E\big(\sigma^2(U_\ell^{(N)}) \cdot (W(t) - W(t_\ell^{(N)}))^4\big) \\
&\quad \leq c \cdot \big(1 + E|\widetilde{X}_N^M(t_\ell^{(N)})|^2\big) \\
&\quad \leq c/k_N^2,
\end{aligned}$$

where the last estimate follows from Lemma 2(i).
We conclude that

$$\begin{aligned}
&\big(E\|\widetilde{X}_N^M - \overline{X}_N\|_2^2\big)^{1/2} \\
&\qquad \geq \left(\sum_{\ell=0}^{k_N-1} \int_{t_\ell^{(N)}}^{t_{\ell+1}^{(N)}} E\big|\sigma(U_\ell^{(N)}) \cdot Z_N(t)\big|^2 \, dt\right)^{1/2} - c/k_N,
\end{aligned}$$

and it remains to observe $c/k_N = o(N^{-1/2})$ due to (13). □

The random set of discretization points used by $\overline{X}_N$ is given by

$$\begin{aligned}
\mathcal{M}^{(N)} = \Big\{&\psi_1^{(N)}(x_0), \psi_2^{(N)}\big(x_0, W(\psi^{(N)}(x_0)\big), \\
&\ldots, \psi^{(N)}_{\nu_{(\psi^{(N)},\chi^{(N)},\varphi^{(N)})}}\big(D_{\nu_{(\psi^{(N)},\chi^{(N)},\varphi^{(N)})}-1}\big)\Big\}
\end{aligned}$$

Let

$$m_\ell^{(N)} = \#\big(\mathcal{M}^{(N)} \cap]t_\ell^{(N)}, t_{\ell+1}^{(N)}[\big)$$

denote the random number of points in $]t_\ell^{(N)}, t_{\ell+1}^{(N)}[$ and put

$$A_N = \sum_{\ell=0}^{k_N-1} \frac{\sigma^2(t_\ell^{(N)}, X(t_\ell^{(N)}))}{m_\ell^{(N)} + 1}.$$

Lemma 6.

$$\begin{aligned}
\liminf_{N\to\infty} N \cdot \sum_{\ell=0}^{k_N-1} \int_{t_\ell^{(N)}}^{t_{\ell+1}^{(N)}} E\big|\sigma(U_\ell^{(N)}) \cdot Z_N(t)\big|^2 \, dt \\
\geq \liminf_{N\to\infty} N/(6k_N^2) \cdot E(A_N).
\end{aligned}$$

Proof. Clearly,

$$E\big(|\sigma(U_\ell^{(N)}) \cdot Z_N(t)|^2 \,\big|\, D^{(N)}\big) = \sigma^2(U_\ell^{(N)}) \cdot E\big(Z_N^2(t) \,\big|\, D^{(N)}\big)$$

Conditioned on the data $D^{(N)}$, the discretization $\mathcal{M}^{(N)}$ is fixed and the process Z_N is a Brownian bridge on each of the corresponding subintervals.

Thus

$$\int_{t_\ell^{(N)}}^{t_{\ell+1}^{(N)}} E\big(Z_N^2(t) \mid D^{(N)}\big)\, dt \geq \frac{1}{6k_N^2 \cdot (m_\ell^{(N)}+1)}$$

due to Lemma 4, and consequently

$$\int_{t_\ell^{(N)}}^{t_{\ell+1}^{(N)}} E\big|\sigma(U_\ell^{(N)}) \cdot Z_N(t)\big|^2\, dt \geq 1/(6k_N^2) \cdot E\left(\frac{\sigma^2(U_\ell^{(N)})}{m_\ell^{(N)}+1}\right)$$

By the regularity assumptions on σ,

$$\begin{aligned}&\big|\sigma^2(U_\ell^{(N)}) - \sigma^2(t_\ell^{(N)}, X(t_\ell^{(N)})\big| \\ &\quad \leq c \cdot |\widetilde{X}_N^M(t_\ell^{(N)}) - X(t_\ell^{(N)})| \cdot \big(1 + |\widetilde{X}_N^M(t_\ell^{(N)})| + |X(t_\ell^{(N)})|\big).\end{aligned}$$

Hence, by Theorem 12, Lemma 4 and (6),

$$E\big|\sigma^2(U_\ell^{(N)}) - \sigma^2(t_\ell^{(N)}, X(t_\ell^{(N)})\big|^2 \leq c/k_N,$$

which yields

$$\sum_{\ell=0}^{k_N-1} \int_{t_\ell^{(N)}}^{t_{\ell+1}^{(N)}} E\big|\sigma(U_\ell^{(N)}) \cdot Z_N(t)\big|^2\, dt \geq 1/(6k_N^2) \cdot E(A_N) - c/k_N^2.$$

Employing (13) completes the proof. □

Lemma 7.

$$\liminf_{N\to\infty} N/(6k_N^2) \cdot E(A_N) \geq C^2$$

Proof. By assumption on $\overline{X}_N$ we have

$$\sum_{\ell=0}^{k_N-1} E\big(m_\ell^{(N)}+1\big) = c(\overline{X}_N) \leq N.$$

The Hölder inequality yields

$$\begin{aligned} &N \cdot E(A_N) \\ &\quad \geq \sum_{\ell=0}^{k_N-1} E\left(\frac{\sigma^2(t_\ell^{(N)}, X(t_\ell^{(N)}))}{m_\ell^{(N)}+1}\right) \cdot \sum_{\ell=0}^{k_N-1} E\big(m_\ell^{(N)}+1\big) \\ &\quad \geq \left(\sum_{\ell=0}^{k_N-1}\left(E\left(\frac{\sigma^2(t_\ell^{(N)}, X(t_\ell^{(N)}))}{m_\ell^{(N)}+1}\right)\right)^{1/2} \cdot \Big(E\big(m_\ell^{(N)}+1\big)\Big)^{1/2}\right)^2 \\ &\quad \geq \left(\sum_{\ell=0}^{k_N-1} E\big|\sigma(t_\ell^{(N)}, X(t_\ell^{(N)}))\big|\right)^2 \end{aligned}$$

Thus

$$\begin{aligned} &\liminf_{N\to\infty} N/(6k_N^2) \cdot E(A_N) \\ &\qquad \geq \frac{1}{6} \cdot \left(\lim_{N\to\infty} \frac{1}{k_N} \cdot \sum_{\ell=0}^{k_N-1} E\big|\sigma(t_\ell^{(N)}, X(t_\ell^{(N)}))\big|\right)^2 \\ &\qquad = C^2, \end{aligned}$$

as claimed. □

Combine Lemmas 5-7 to obtain the lower bound in Theorem 8. □

We turn to the analysis of the approximation $\widehat{X}_n$.

Recall the *coarse* discretization

$$t_\ell^{(n)} = \ell/k_n, \quad \ell = 0, \dots, k_n,$$

with

$$\lim_{n\to\infty} n^{1/2}/k_n = \lim_{n\to\infty} k_n/n = 0,$$

that appears in the definition of $\widehat{X}_n$, and let $\widetilde{X}_n^M$ denote the corresponding time-continuous version of the Milstein scheme.

Note that $\widetilde{X}_n^M$ and $\widehat{X}_n$ coincide at the points $t_\ell^{(n)}$.

Proof of Theorem 6(ii). By definition of $\widehat{X}_n$,

$$\begin{aligned} k_n + n/k_n \cdot \sum_{\ell=0}^{k_n-1} E\big|\sigma(t_\ell^{(n)}, \widetilde{X}_n^M(t_\ell^{(n)}))\big| &\le c(\widehat{X}_n) \\ &\le 2k_n + n/k_n \cdot \sum_{\ell=0}^{k_n-1} E\big|\sigma(t_\ell^{(n)}, \widetilde{X}_n^M(t_\ell^{(n)}))\big|. \end{aligned}$$

By the Lipschitz property of σ and Theorem 12,

$$\begin{aligned} E\big|\sigma(t_\ell^{(n)}, \widetilde{X}_n^M(t_\ell^{(n)})) - \sigma(t_\ell^{(n)}, X(t_\ell^{(n)}))\big| &\le c \cdot E\big|\widetilde{X}_n^M(t_\ell^{(n)}) - X(t_\ell^{(n)})\big| \\ &\le c/k_n. \end{aligned}$$

Hence

$$\begin{aligned} &\big|1/n \cdot c(\widehat{X}_n) - \sqrt{6}\cdot C\big| \\ &\quad \le \Big|1/n \cdot c(\widehat{X}_n) - 1/k_n \cdot \sum_{\ell=0}^{k_n-1} E\big|\sigma(t_\ell^{(n)}, \widetilde{X}_n^M(t_\ell^{(n)}))\big|\Big| \\ &\qquad + 1/k_n \cdot \sum_{\ell=0}^{k_n-1} E\big|\sigma(t_\ell^{(n)}, \widetilde{X}_n^M(t_\ell^{(n)})) - \sigma(t_\ell^{(n)}, X(t_\ell^{(n)}))\big| \\ &\qquad + \Big|1/k_n \cdot \sum_{\ell=0}^{k_n-1} E\big|\sigma(t_\ell^{(n)}, X(t_\ell^{(n)}))\big| - \sqrt{6}\cdot C\Big| \\ &\quad \le c \cdot k_n/n + \Big|1/k_n \cdot \sum_{\ell=0}^{k_n-1} E\big|\sigma(t_\ell^{(n)}, X(t_\ell^{(n)}))\big| - \sqrt{6}\cdot C\Big|. \end{aligned}$$

With increasing n the last sum above tends to zero, which finishes the proof. □

We proceed with a comparison of $\widehat{X}_n$ and $\widetilde{X}_n^M$.

Lemma 8.

$$\limsup_{n\to\infty} n \cdot E\|\widehat{X}_n - \widetilde{X}_n^M\|_2^2 \le C/\sqrt{6}.$$

Proof. Let $\mathcal{B}_n$ denote the σ-algebra that is generated by

$$W(t_1^{(n)}), \dots, W(t_{k_n}^{(n)}).$$

Fix $\ell \in \{0, 1, \dots, k_n - 1\}$ and recall that $\widehat{X}_n$ uses

$$\mu_\ell^{(n)} = \big\lceil n/k_n \cdot |\sigma(t_\ell^{(n)}, \widetilde{X}_n^M(t_\ell^{(n)}))|\big\rceil$$

equidistant points

$$\tau_{\ell,j}^{(n)} = t_\ell^{(n)} + j \cdot (k_n \cdot (\mu_\ell^{(n)} + 1))^{-1}, \quad j = 1, \dots, \mu_\ell^{(n)},$$

in the subinterval $]t_\ell^{(n)}, t_{\ell+1}^{(n)}[$.

Note that the adaptive discretization is $\mathcal{B}_n$-measurable and let $\widetilde{W}_n$ denote the piecewise linear interpolation of $W - W(t_\ell^{(n)})$ at these points.

For $t \in [t_\ell^{(n)}, t_{\ell+1}^{(n)}]$ put

$$V_n(t) = 1/2 \cdot \big(\sigma \cdot \sigma^{(0,1)}\big)\big(t_\ell^{(n)}, \widetilde{X}_n^M(t_\ell^{(n)})\big) \times \big((W(t) - W(t_\ell^{(n)}))^2 - (t - t_\ell^{(n)})\big)$$

Then

$$\begin{aligned}&\widetilde{X}_n^M(t) - \widehat{X}_n(t)\\ &\quad = \sigma(t_\ell^{(n)}, \widetilde{X}_n^M(t_\ell^{(n)})) \cdot \big(W(t) - W(t_\ell^{(n)}) - \widetilde{W}_n(t)\big) + V_n(t)\end{aligned}$$

for $t_\ell^{(n)} \le t \le \tau_{\ell,\mu_\ell^{(n)}}^{(n)}$, and

$$\begin{aligned}&|\widetilde{X}_n^M(t) - \widehat{X}_n(t)|\\ &\quad \le \big|\sigma(t_\ell^{(n)}, \widetilde{X}_n^M(t_\ell^{(n)}))\big| \cdot \big|W(t) - W(t_\ell^{(n)}) - \widetilde{W}_n(t)\big|\\ &\qquad + |V_n(t)| + |V_n(t(n)_{\ell+1})|\end{aligned}$$

for $\tau_{\ell,\mu_\ell^{(n)}}^{(n)} < t < t_{\ell+1}^{(n)}$.

Conditioned on $\mathcal{B}_n$ the adaptive discretization is fixed and the process $W - W(t_\ell^{(n)})$ is a Brownian bridge on each of the subintervals $[\tau_{\ell,j}^{(n)}, \tau_{\ell,j+1}^{(n)}]$.

Hence, by Lemma 4,

$$\begin{aligned}&\int_{t_\ell^{(n)}}^{t_{\ell+1}^{(n)}} E\Big(\big|\sigma(t_\ell^{(n)}, \widetilde{X}_n^M(t_\ell^{(n)}))\\ &\qquad \times (W(t) - W(t_\ell^{(n)}) - \widetilde{W}_n(t))\big|^2 \,\big|\, \mathcal{B}_n\Big)\, dt\\ &\quad = \big|\sigma(t_\ell^{(n)}, \widetilde{X}_n^M(t_\ell^{(n)}))\big|^2 \cdot \frac{1}{6k_n^2 \cdot (\mu_\ell^{(n)} + 1)}\\ &\quad \le \frac{\big|\sigma(t_\ell^{(n)}, \widetilde{X}_n^M(t_\ell^{(n)}))\big|}{6k_n \cdot n},\end{aligned}$$

and we obtain

$$\begin{aligned}&\sum_{\ell=0}^{k_n-1} \int_{t_\ell^{(n)}}^{t_{\ell+1}^{(n)}} E\big|\sigma(t_\ell^{(n)}, \widetilde{X}_n^M(t_\ell^{(n)}))\\ &\qquad \times (W(t) - W(t_\ell^{(n)}) - \widetilde{W}_n(t))\big|^2\, dt\\ &\quad \le 1/6 \cdot \frac{1}{k_n} \cdot \sum_{\ell=0}^{k_n-1} E\big|\sigma(t_\ell^{(n)}, \widetilde{X}_n^M(t_\ell^{(n)}))\big|\\ &\quad \le 1/6 \cdot 1/n^2 \cdot c(\widehat{X}_n).\end{aligned}$$

By the regularity conditions on σ and Lemma 2 we have

$$\sup_{t \in [0,1]} E|V_n(t)|^2 \le c/k_n^2.$$

Thus we conclude

$$\big(E\|\widehat{X}_n - \widetilde{X}_n^M\|_2^2\big)^{1/2} \le 1/\sqrt{6} \cdot 1/n \cdot \big(c(\widehat{X}_n)\big)^{1/2} + c/k_n.$$

Now, observe $k_n = o(n^{-1/2})$ and use Theorem 6(ii) to obtain the desired result. □

Proof of the upper bound in Theorem 6(i).

By Theorem 12,

$$\big(c(\widehat{X}_n)\big)^{1/2} \cdot e(\widehat{X}_n)$$
$$\leq \big(c(\widehat{X}_n)\big)^{1/2} \cdot \big((E\|\widetilde{X}_n^M - \widehat{X}_n\|_2^2)^{1/2} + (E\|X - \widetilde{X}_n^M\|_2^2)^{1/2}\big)$$
$$\leq \big(c(\widehat{X}_n)\big)^{1/2} \cdot \big((E\|\widetilde{X}_n^M - \widehat{X}_n\|_2^2)^{1/2} + c/k_n\big).$$

Theorem 6(ii) implies

$$\limsup_{n\to\infty} 1/k_n \cdot \big(c(\widehat{X}_n)\big)^{1/2} \leq c \cdot \limsup_{n\to\infty} n^{1/2}/k_n = 0.$$

Furthermore, by Theorem 6(ii) and Lemma 8,

$$\limsup_{n\to\infty} \big(c(\widehat{X}_n)\big)^{1/2} \cdot \big(E\|\widetilde{X}_n^M - \widehat{X}_n\|_2^2\big)^{1/2} \leq C,$$

which finishes the proof. □

Obviously, the upper bound from Theorem 6(i) implies the upper bound from Theorem 8, while the lower bound from Theorem 8 yields the lower bound from Theorem 6(i).

References

Literature related to Part I

Arnold, L. (1974). *Stochastic Differential Equations.* Wiley, New York.

Chung, K. L., and Williams, R. (1990). *Introduction to Stochastic Integration.* Birkhäuser, Boston.

Elliot, R. J. (1982). *Stochastic Calculus and Applications.* Springer, Berlin.

Friedman, A. (1975). *Stochastic Differential Equations and Applications, vol. I.* Academic Press, New York.

Ikeda, N., and Watanabe, S. (1981). *Stochastic Differential Equations and Diffusion Processes.* North-Holland, Amsterdam.

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

McKean, H. P. (1969). *Stochastic Integrals.* Academic Press, New York.

Oksendal, B. (1998) *Stochastic Differential Equations.* Springer, Berlin.

Steele, M. (2000). *Stochastic Calculus and Financial Applications.* Springer, Berlin.

Literature related to Parts II, III, IV and V

Bouleau, N., and Lépingle, D. (1994). *Numerical Methods for Stochastic Processes.* Wiley, New York.

Cambanis, S., and Hu, Y. (1996). Exact convergence rate of the Euler-Maruyama scheme, with application to sampling design, Stochastics Stochastics Rep. **59**, 211–240.

Clark, J. M. C., and Cameron, R. J. (1980). The maximum rate of convergence of discrete approximations for stochastic differential equations. In W. H. Fleming and L. G. Gorostiza (eds.), *Lecture Notes in Control and Information Sciences* **25**, Springer, Berlin.

Clark, J. M. C. (1982). An efficient approximation for a class of stochastic differential equations. In B. Grigelionis (ed.), *Lecture Notes in Control and Information Sciences* **42**, Springer, Berlin.

Faure, O. (1992). *Simulation du mouvement brownien et des diffusions.* Thèse ENPC, Paris.

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

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

Hofmann, N., Müler-Gronbach, T., and Ritter, K. (2000a). Optimal approximation of stochastic differential equations by adaptive step-size control, Math. Comp. **69**, 1017–1034.

Hofmann, N., Müler-Gronbach, T., and Ritter, K. (2000b). Step-size control for the uniform approximation of systems of stochastic differential equations with additive noise, Ann. Appl. Prob. **10**, 616–633.

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

Hofmann, N., Müller-Gronbach, T. und K. Ritter (2002). Linear vs. Standard Information for Scalar Stochastic Differential Equations. J. Complexity **18**, 394–414.

Hofmann, N. und T. Müler-Gronbach (2004). On the Global Error of Ito-Taylor Schemes for Strong Approximation of Stochastic Differential Equations. To appear in J. Complexity .

Kloeden, P. and Platen, E. (1995). *Numerical Solution of Stochastic Differential Equations.* Springer, Berlin.

Maruyama, G. (1955). Continuous Markov Processes and stochastic equations, Ren. Circolo Math. Palermo **4**, 48–90.

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

Milstein, G. N. (1974). Approximate integration of stochastic differential equations, Theory Probab. Appl. **19**, 557–562.

Milstein, G. N. (1995). *Numerical Integration of Stochastic Differential Equations*. Kluwer, Dordrecht.

Müller-Gronbach, T. (2002a). *Strong Approximation of Systems of Stochastic Differential Equations.* Habilitationsschrift, Technische Universität Darmstadt.

Müller-Gronbach, T. (2002b) Optimal Uniform Approximation of Systems of Stochastic Differential Equations, Ann. Appl. Prob. **12**, 664–690.

Müller-Gronbach, T. (2004). Optimal pointwise approximation of SDE's based on Brownian motion at discrete points. To appear in Ann. Appl. Prob. **14**, No. 4.

Newton, N. J. (1986). An asymptotically efficient difference formula for solving stochastic differential equations, Stochastics Stochastics Rep. **19**, 175–206.

Newton, N. J. (1991). Asymptotically efficient Runge-Kutta methods for a class of Itô and Stratonovich equations, SIAM J. Appl. Math. **51**, 542–567.

Pardoux, E. and Talay, D. (1985). Discretization and simulation of stochastic differential equations, Acta Appl. Math. **3**, 23–47.

Revuz, D. and Yor, M. (1991). *Continuous Martingales and Brownian Motion*, Springer, Berlin.

Ritter, K. (2000). *Average-Case Analysis of Numerical Problems.* Lecture Notes in Mathematics **1733**, Springer, Berlin.

Rümelin, W. (1982). Numerical treatment of stochastic differential equations, SIAM J. Numer. Anal. **19**, 604–613.

Rydén, T. and Wiktorsson, M. (2001). On the simulation of iterated Itô-integrals, Stochastic Processes Appl. **91**, 151–168.

Talay, D. (1983). Résolution trajectorielle et analyse numérique des équations différentieles stochastiques, Stochastics **9**, 275–306.

Talay, D. (1995). Simulation of stochastic differential systems. In *Probabilistic Methods in Applied Physics* (P. Krée, W. Wedig, eds.) Lecture Notes in Physics **451**, Springer, Berlin.

Traub, J. F., Wasilkowski, G. W., and Woźniakowski, H. (1988). *Information-based Complexity*, Academic Press, New York.

Wagner, W. and Platen, E. (1978). Approximation of Itô integral equations, Preprint ZIMM, Akad. der Wiss. der DDR, Berlin.

Yamoto, Y. (1979). Stochastic differential equations and nilpotent Lie algebras, Z. Wahrscheinlichkeitstheorie verw. Gebiete **47**, 213–229.