THE OPTIMAL DISCRETIZATION OF STOCHASTIC DIFFERENTIAL EQUATIONS

NORBERT HOFMANN, THOMAS MÜLLER-GRONbach, AND KLAUS RITTER

Abstract. We study pathwise approximation of scalar stochastic differential equations. The mean squared $L_2$-error and the expected number $n$ of evaluations of the driving Brownian motion are used for the comparison of arbitrary methods. We introduce an adaptive discretization that reflects the local properties of every single trajectory. The corresponding error tends to zero like $c \cdot n^{-1/2}$, where $c$ is the average of the diffusion coefficient in space and time. Our method is justified by the matching lower bound for arbitrary methods that are based on $n$ evaluations on the average. Hence the adaptive discretization is asymptotically optimal. The new method is very easy to implement, and about 7 additional arithmetical operations are needed per evaluation of the Brownian motion. Hereby we can determine the complexity of pathwise approximation of stochastic differential equations. We illustrate the power of our method already for moderate accuracies by means of a simulation experiment.

1. Introduction

Stochastic differential equations can be solved explicitly only in exceptional cases, such that numerical methods must be used in general. In this paper we study pathwise (or strong) approximation for stochastic differential equations, and we analyze numerical methods with respect to their error and computational cost.

Let $W$ denote a one-dimensional Brownian motion on the unit interval, and consider a scalar stochastic differential equation

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

with initial value $X(0)$. A pathwise approximation to the solution $X$ is a stochastic process $\overline{X}$ whose paths are close to the respective paths of $X$.

Every numerical method relies on finite and therefore partial information about the underlying Brownian motion. We assume that $W$ may be observed at adaptively chosen points

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

The choice of $\tau_{k+1}$ may depend in any measurable way on the previously computed values $W(\tau_1), \ldots, W(\tau_k)$. In particular, we do not require $\tau_k < \tau_{k+1}$. The total number

Date: July 16, 1999.

1 Supported by the DFG: GR 876/9-2.
\( \nu \) of observations may be determined by any measurable termination criterion. The discrete data about \( W, a, \) and \( \sigma \) may then be used in any measurable way to produce a pathwise approximation \( \overline{X} \). An adaptive discretization should take into account

(D) the drift and diffusion coefficient,

(T) the particular trajectory,

(E) the error criterion.

Here we use the following error criterion. For a given pair of corresponding trajectories of \( X \) and \( \overline{X} \) we measure the distance in the \( L_2 \)-norm, and we define the error \( e(\overline{X}) \) of the method \( \overline{X} \) by averaging over all trajectories, i.e.,

(2) 
\[
e(\overline{X}) = \left( E(\|X - \overline{X}\|_2^2) \right)^{1/2}.
\]

As this definition depends on the discretization only via \( \overline{X} \), it may serve as a basis for comparing different discretizations. Moreover, we characterize the quality of \( \overline{X} \) globally on \([0, 1]\) and not only on a finite number of points.

As a rough measure for the computational cost of \( \overline{X} \) we use the expected number \( n(\overline{X}) \) of observations of \( W \). A more realistic measure also involves a count of the arithmetical operations performed by \( \overline{X} \).

We address the following question: How large does the cost have to be in order to achieve an error \( e(\overline{X}) \leq \varepsilon \)? A complete answer consists of two parts, upper bounds for specific methods and lower bounds for arbitrary methods. The minimal cost over all methods with \( e(\overline{X}) \leq \varepsilon \) is called the \( \varepsilon \)-complexity of strong approximation of the equation (1).

We introduce an adaptive discretization that reflects the local smoothness of the solution. The discretization is very easy to implement and has the convenient property that \( \tau_k < \tau_{k+1} \). The smoothness of the solution at the point \((t, X(t))\) is determined by

(3) 
\[
E \left( (X(t + \delta) - X(t))^2 \mid X(t) \right) = \sigma^2(t, X(t)) \cdot \delta + o(\delta),
\]

hence \(|\sigma|(t, X(t))\) might be called a conditional Hölder constant. It is reasonable to choose a small step-size if the conditional Hölder constant is large and vice versa. We basically take the step-size proportionally to the inverse of the current value of \(|\sigma|\), and we use Euler and Milstein steps to move along the unit interval from left to right.

Hereby we get a method \( \hat{X}_{h}^{**} \) with input parameter \( h > 0 \) such that

\[
\lim_{h \to 0} n(\hat{X}_{h}^{**})^{1/2} \cdot e(\hat{X}_{h}^{**}) = E \left( \int_0^1 |\sigma|(t, X(t)) \, dt \right) / \sqrt{6},
\]

see Theorem 1. In particular this is an asymptotic upper bound for the methods \( \hat{X}_{h}^{**} \).

The matching asymptotic lower bounds reads

\[
\liminf_{N \to \infty} N^{1/2} \cdot e(X_N) \geq E \left( \int_0^1 |\sigma|(t, X(t)) \, dt \right) / \sqrt{6}
\]
for every sequence of methods $\mathbf{X}_N$ such that
\[ n(\mathbf{X}_N) \leq N, \]
see Theorem 4. Our adaptive method $\hat{\mathbf{X}}^{**}_h$ is therefore asymptotically optimal. The best order of convergence is $1/2$ in terms of $n(\mathbf{X})$, and the best asymptotic constant is given by the mean of the conditional Hölder constant in space and time. Note that the drift $a$ and the initial value $X(0)$ are present implicitly in this constant.

In Section 2 we define three versions of our adaptive method. The fully adaptive method is conceptually very simple, however, we cannot provide an asymptotic analysis. Hence we introduce the simplified adaptive method $\hat{\mathbf{X}}^{**}_h$. These two methods adjust the number $\nu$ of evaluations of the given trajectory of $W$ to the corresponding trajectory of $X$. The third method uses the same number of evaluations for every trajectory, thus providing a deterministic a priori bound for the computation time.

In Section 3 we specify our assumptions regarding the equation. The drift and diffusion coefficients must satisfy Lipschitz conditions, and the initial value must have a finite fourth moment.

Section 4 contains the asymptotic results for our methods. Moreover, we present the asymptotic analysis for the Milstein scheme based on an equidistant discretization. The latter also yields errors of order $1/2$, but its asymptotic constant is not optimal. For instance, this constant depends exponentially on the volatility of a geometric Brownian motion, while the optimal dependence is a linear one.

In Section 5 we formally define arbitrary methods for strong approximation and we present the lower bounds for these methods. In addition to the optimality of $\hat{\mathbf{X}}^{**}_h$ we observe that it is crucial to adjust the number $\nu$ to the given trajectory.

In Section 6 we combine the results from Sections 4 and 5 to determine the $\varepsilon$-complexity. The gap in the upper and lower bounds is only a factor of 8. This is due to the fact that asymptotically the method $\hat{\mathbf{X}}^{**}_h$ only needs 7 additional arithmetical operations per evaluation of $W$. Hence our bounds give a very precise estimate of the actual computation time.

Section 7 is devoted to a simulation experiment for the geometric Brownian motion $dX(t) = 3 \cdot X(t) dW(t)$ with $X(0) = 1$. According to the analysis, our adaptive method reduces the computation time by a factor 900, compared to an equidistant discretization. This is confirmed in the experiments already for moderate accuracies $\varepsilon \leq 2 \cdot 10^{-2}$.

Proofs are given in Section 8, and the appendix contains an upper bound for the error of the Milstein methods, which is useful in our analysis.

Our analysis shows that the local smoothness (3) is the key quantity. This fact was already known in much more generality for problems of reconstruction of stochastic processes, see, e.g., Müller-Gronbach and Ritter (1998) and the references therein. Let us stress two essential differences between both problems.
1) strong approximation for stochastic differential equations,
2) reconstruction of stochastic processes.

Discrete observations of the process $X$ itself are used for the reconstruction of $X$ in 2), while only $W$ is observable in 1). The solution $X$ depends nonlinearly on $W$ in 1), while there is a linear dependence (the identity) in 2). According to (3) one can compute the local smoothness rather accurately in a numerical method. This may also explain why optimal discretizations can be determined for stochastic differential equations, while such discretizations are unknown for ordinary differential equations.

Most of the results for strong (and weak) approximation for stochastic differential equations provide upper bounds for specific methods with unspecified constants. See, e.g., Bouleau and Lépingle (1994), Kloeden and Platen (1995), Milstein (1995), and Talay (1995) for results and numerous references. Only a few papers deal with asymptotic constants. We mention two results that deal with the mean squared error at the point $t = 1$ for autonomous equations.

For the first time, asymptotic constants were derived by Clark and Cameron (1980), who considered the case $\sigma = 1$. They studied a specific method, namely, the conditional expectation of $X(1)$ given the values of $W$ at an equidistant discretization.

The first result concerning adaptivity with respect to $(D)$ is due to Cambanis and Hu (1996), who considered discretizations that are given as quantiles of a fixed density on the unit interval. Cambanis and Hu characterize the asymptotically best choice of the density, which leads to the smallest asymptotic constant. Moreover, they pose several open problems, some of which are settled by the results of this paper. In particular, Theorem 4 solves problem (1.18) from Cambanis and Hu (1996).\(^1\)

The first result concerning adaptivity with respect to $(T)$ is due to Newton (1990), who determined the discretization from passage times of the driving process. This contrasts our approach, since we aim at trajectorial properties of the solution.

In Hofmann et al. (1999a, 1999b) we study pathwise approximation for equations (1) with additive noise, i.e., $\sigma(t, x) = \sigma(t)$. In this case it suffices to use adaption with respect to $(D)$ and $(E)$.

We also mention Gaines and Lyons (1997) and Mauthner (1998), who constructed an adaptive method and proved its convergence.

Another problem, which also arises in stochastic analysis, is the approximate computation of stochastic integrals. Wasilkowski and Woźniakowski (1999) and Hertling (1999) determine the complexity of this problem.

\(^1\)The answer is $\gamma = 1$ as well as $F(a, b) = C^2/6$ if only adaption with respect to $(D)$ is used and $F(a, b) = (C^*)^2/6$ in general. Discretizations of varying size are not mentioned by Cambanis and Hu (1996).
2. The Adaptive Method

The smoothness of the solution $X$ of (1) at the point $(t, X(t))$ is determined by the conditional Hölder constant $|\sigma|(t, X(t))$, see (3). Therefore it is reasonable to decrease the step-size $\tau_{\ell+1} - \tau_\ell$ with increasing value of $|\sigma|(\tau_\ell, X(\tau_\ell))$. Hereby we deal with the first two requirements (D) and (T) for an adaptive discretization. The precise relation between the step-size and the conditional Hölder constant depends on the error criterion. For the $L_2$-error (2) we take steps proportionally to $1/|\sigma|(\tau_\ell, X(\tau_\ell))$; see Remark 4 for further discussion. Since $X(\tau_\ell)$ is unknown in general, an approximation must be used.

We take the Milstein scheme to compute approximations at discrete points. For every discretization
\begin{equation}
0 = \tau_0 < \cdots < \tau_m = 1
\end{equation}
this scheme is defined by
\[ \hat{X}(\tau_0) = X(0) \]
and
\[ \hat{X}(\tau_{\ell+1}) = \hat{X}(\tau_\ell) + a(\tau_\ell, \hat{X}(\tau_\ell)) \cdot (\tau_{\ell+1} - \tau_\ell) \]
\[ + \sigma(\tau_\ell, \hat{X}(\tau_\ell)) \cdot (W(\tau_{\ell+1}) - W(\tau_\ell)) \]
\[ + 1/2 \cdot (\sigma \cdot \sigma^{(0,1)})(\tau_\ell, \hat{X}(\tau_\ell)) \cdot (((W(\tau_{\ell+1}) - W(\tau_\ell))^2 - (\tau_{\ell+1} - \tau_\ell)), \]
where $\ell = 0, \ldots, m - 1$. Here $\sigma^{(0,1)}$ denotes the partial derivative of $\sigma$ with respect to the second or state variable. A global approximation $\tilde{X}$ for $X$ on $[0, 1]$ is obtained by piecewise linear interpolation of the data $(\tau_\ell, \hat{X}(\tau_\ell))$. Obviously $\tilde{X}$ depends on $W$ only through its values at the discretization points.

We present three versions of the adaptive method.

2.1. The fully adaptive method $\hat{X}_h$. We choose a basic step-size $h > 0$
and define the adaptive step-size control by $\tau_0 = 0$ and
\[ \tau_{\ell+1} = \tau_\ell + h/|\sigma|(\tau_\ell, \hat{X}(\tau_\ell)) \]
as long as the right-hand side does not exceed one. Otherwise we put $\tau_{\ell+1} = 1$. We refer to the corresponding Milstein approximation by $\hat{X}_h$ in the sequel. See Remark 4 for further discussion.

Obviously the adaptive method $\hat{X}_h$ is very easy to implement. Unfortunately, it seems to be difficult to analyze this method. Therefore we switch to a closely related method $\hat{X}_h^{**}$. Simulation experiments indicate that the fully adaptive method is superior to $\hat{X}_h^{**}$ for moderate basic step sizes. Our analysis shows that $\hat{X}_h$ cannot be better than $\hat{X}_h^{**}$.
asymptotically. In fact, we conjecture that both methods have the same asymptotic performance.

2.2. The simplified adaptive method \( \hat{X}^{**}_h \). For \( h > 0 \) take \( k_h \in \mathbb{N} \) such that

\[
\lim_{h \to 0} k_h \cdot h = 0
\]

and

\[
\lim_{h \to 0} k_h^2 \cdot h = \infty.
\]

The modification \( \hat{X}^{**}_h \) of \( \hat{X}_h \) does not update the step-size in every step but only at the equidistant points

\[
\tau_{i,0} = i/k_h, \quad i = 0, \ldots, k_h - 1.
\]

Suppose that

\[
x_i = \hat{X}^{**}_h(\tau_{i,0})
\]

is already computed, where \( x_i = X(0) \) for \( i = 0 \). Put

\[
\sigma_i = \sigma(\tau_{i,0}, x_i), \quad a_i = a(\tau_{i,0}, x_i),
\]

and define an adaptive discretization of the subinterval \( ]\tau_{i,0}, \tau_{i+1,0}[, \) by

\[
\tau_{i,j+1} = \tau_{i,j} + h/|\sigma_i|
\]

as long as the right-hand side is less than \( \tau_{i+1,0} \). Here \( \tau_{k_h,0} = 1 \). On the subinterval \( ]\tau_{i,0}, \tau_{i+1,0}[ \) we use the Euler method

\[
\hat{X}^{**}_h(\tau_{i,j+1}) = \hat{X}^{**}_h(\tau_{i,j}) + a_i \cdot (\tau_{i,j+1} - \tau_{i,j}) + \sigma_i \cdot (W(\tau_{i,j+1}) - W(\tau_{i,j})),
\]

without updating the drift and diffusion coefficient. The approximate solution \( x_{i+1} = \hat{X}^{**}_h(\tau_{i+1,0}) \) at the right endpoint is computed by a single Milstein step of length \( 1/k_h \) starting at the left endpoint \( \tau_{i,0} \) with initial value \( x_i \). Globally we use piecewise linear interpolation. In Theorem 1 we present the asymptotic analysis of the error \( e(\hat{X}^{**}_h) \) with \( h \) tending to zero.

Note that on each subinterval \( ]\tau_{i,0}, \tau_{i+1,0}[ \) we apply the Euler method to the stochastic differential equation

\[
dX(t) = a_i \, dt + \sigma_i \, dW_i(t)
\]

with initial value \( X(\tau_{i,0}) = x_i \) and driving Brownian motion \( W_i(t) = W(t) - W(\tau_{i,0}) \). For this equation the Euler and the Milstein method coincide, and they yield the exact solution at the discretization points.

It is known that a Milstein step of length \( 1/k_h \) yields a squared error of order \( 1/k_h^2 \). Even for the original equation, Euler steps of order \( h \) only yield a squared error of order \( h \). The use of the large Milstein steps is therefore crucial for the good approximation of the diffusion coefficient at \( \tau_{i+1,0} \).
The number of observations of $W$ that are used by $\hat{X}_h^{**}$ is roughly given by $1/h \cdot S$, where

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

is the average of the conditional Hölder constant for the particular trajectory. The quantity $S$ may heavily depend on the trajectory, see Figure 2 in Section 7, and consequently, there is no a priori bound on the computation time available for the user. If all approximations have to be computed in the same amount of time, we suggest to use the following version $\hat{X}_n^*$ of the adaptive method. We stress, however, that one has to pay a price for this property, see Theorems 1 and 2, the second equation from Theorem 4 as well as Example 1.

### 2.3. The adaptive method $\hat{X}_n^*$ with at most $n$ discretization points.

For $n \in \mathbb{N}$ take $k_n \in \mathbb{N}$ such that

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

and

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

In contrast to $\hat{X}_h^{**}$ and $\hat{X}_h$, the method $\hat{X}_n^*$ does not progress from left to right. At first the Milstein approximation

$$x_i = \hat{X}_n^*(\tau_{i,0})$$

is computed at all equidistant points

$$\tau_{i,0} = i/k_n, \quad i = 0, \ldots, k_n - 1.$$  

Thereafter the discretization is refined adaptively in the following way. Put

$$\sigma_i = \sigma(\tau_{i,0}, x_i)$$

and let

$$\mu_i = \begin{cases} 
\lfloor (n - k_n) \cdot |\sigma_i|/\sum_{\ell=0}^{k_n-1} |\sigma_\ell| \rfloor, & \text{if } \sum_{\ell=0}^{k_n-1} |\sigma_\ell| > 0 \\
\lfloor (n - k_n)/k_n \rfloor, & \text{otherwise.}
\end{cases}$$

The adaptive discretization of the subinterval $]\tau_{i,0}, \tau_{i+1,0}[$ is given by

$$\tau_{i,j+1} = \tau_{i,j} + 1/(k_n \cdot (\mu_i + 1))$$

as long as the right-hand side is less than $\tau_{i+1,0}$. Here $\tau_{k_n,0} = 1$. Now $\hat{X}_n^*$ proceeds in the same way as $\hat{X}_h^{**}$. The asymptotic behavior of the error $e(\hat{X}_n^*)$ is determined in Theorem 2.
Note that the number of Euler steps on $[\tau_i, \tau_{i+1}]$ is given by $\mu_i$. Therefore $\hat{X}_n^*$ uses at most $n$ observations of $W$ for every trajectory. Observe that $\hat{X}_n^*$ takes steps of size roughly given by $1/n \cdot S/|\sigma(\tau_i, X(\tau_i, 0))|$. 

3. Assumptions

Throughout this paper we assume that the drift and diffusion coefficients $a, \sigma : [0, 1] \times \mathbb{R} \to \mathbb{R}$ and the initial value $X(0)$ have the following properties.

(A) Both, $a$ and $\sigma$ are differentiable with respect to the state variable. Moreover, there exists a constant $K > 0$ such that $f = a$ and $f = \sigma$ satisfy

$$|f(t, x) - f(t, y)| \leq K \cdot |x - y|,$$$$
$$|f(s, x) - f(t, x)| \leq K \cdot (1 + |x|) \cdot |s - t|,$$$$
$$|f^{(0,1)}(t, x) - f^{(0,1)}(t, y)| \leq K \cdot |x - y|$$

for all $s, t \in [0, 1]$ and $x, y \in \mathbb{R}$.

(B) The initial value $X(0)$ is independent of $W$ and

$$E(X(0))^4 < \infty.$$ 

Note that (A) yields the linear growth condition

$$|f(t, x)| \leq c \cdot (1 + |x|).$$

Moreover, $f^{(0,1)}$ is bounded and

$$|f(t, x) - f(t, y) - f^{(0,1)}(t, y)(x - y)| \leq c \cdot (x - y)^2.$$ 

Given the above properties, a pathwise unique strong solution of the equation (1) with initial value $X(0)$ exists. In particular the conditions assure that

$$\sup_{t \in [0,1]} E(X(t))^4 < \infty.$$ 

4. Analysis of the Adaptive Method

To every equation (1) we associate the constant

$$C^{**} = E\left(\int_0^1 |\sigma(t, X(t))| \, dt\right).$$

Let $n(\hat{X}_h^{**})$ denote the expected number of observations of $W$ that are used by $\hat{X}_h^{**}$. We relate this quantity to the error $e(\hat{X}_h^{**})$ as $h$ tends to zero.
Theorem 1. The adaptive method $\hat{X}_h^*$ satisfies

$$\lim_{h \to 0} n(\hat{X}_h^{**})^{1/2} \cdot e(\hat{X}_h^{**}) = C^{**}/\sqrt{6}$$

for every equation (1).

For the analysis of $\hat{X}_n^*$ we introduce the constant

$$C^* = \left( E \left( \int_0^1 |\sigma(t, X(t))| \, dt \right)^2 \right)^{1/2}.$$

Recall that $\hat{X}_n^*$ uses at most $n$ observations of $W$ for every trajectory.

Theorem 2. The adaptive method $\hat{X}_n^*$ satisfies

$$\lim_{n \to \infty} n^{1/2} \cdot e(\hat{X}_n^*) = C^*/\sqrt{6}$$

for every equation (1).

Clearly, one can hardly justify the use of an adaptive method, if it is not superior to good methods that are based on an equidistant discretization. Here we take the equidistant Milstein scheme $\hat{X}_n^e$ with step-size $1/n$ for the comparison. Note, however, that we will establish much stronger optimality properties of the adaptive method in Section 5.

We define

$$C^e = \left( \int_0^1 E \left( \sigma^2(t, X(t)) \right) \, dt \right)^{1/2}$$

for the analysis of the method $\hat{X}_n^e$.

Theorem 3. The equidistant Milstein scheme $\hat{X}_n^e$ satisfies

$$\lim_{n \to \infty} n^{1/2} \cdot e(\hat{X}_n^e) = C^e/\sqrt{6}$$

for every equation (1).

Note that the order of convergence is $1/2$ for all the above methods and

$$C^{**} \leq C^* \leq C^e$$

with strict inequality in most cases. See Remark 2 for a characterization of equality.

Example 1. The constants from Theorems 1–3 can be determined analytically only in exceptional cases, e.g., for linear equations or equations with additive noise. Here we consider the linear equation

$$dX(t) = b X(t) \, dW(t)$$

with initial value $X(0) = 1$. The solution is the geometric Brownian motion

$$X(t) = \exp(-b^2/2 \cdot t + b \cdot W(t)).$$
with drift zero. Straightforward calculations yield

\[ C^{**} = b, \]
\[ C^* = \left( 2 \cdot (\exp(b^2) - b^2 - 1) \right)^{1/2}/b, \]
\[ C^e = \left( \exp(b^2) - 1 \right)^{1/2}. \]

For \( \hat{X}^e_n \) and \( \hat{X}^*_n \) the asymptotic constant depends exponentially on the parameter \( b \). For \( \hat{X}^{**}_b \) we only have a linear dependence on \( b \).

5. LOWER BOUNDS AND OPTIMALITY OF THE ADAPTIVE METHOD

Theorems 1–3 determine the asymptotic performance of specific methods for arbitrary equations. These methods are based on a realization of the initial value \( X(0) \) and on a finite number of observations of a trajectory of the driving Brownian motion \( W \). The choice of the adaptive schemes will now be justified by lower bounds that hold for arbitrary methods of the above form and arbitrary equations.

Fix \( a \) and \( \sigma \), and consider the corresponding equation (1). Formally a general method is then defined by mappings

\[ \psi_k : \mathbb{R}^k \to [0, 1], \]
\[ \chi_k : \mathbb{R}^{k+1} \to \{ \text{STOP, GO} \}, \]
\[ \phi_k : \mathbb{R}^{k+1} \to L^2([0, 1]) \]

for \( k \in \mathbb{N} \). The sequential observation of a trajectory \( w \) starts at the knot \( \psi_1(x) \), which may depend on the realization \( x \) of the initial value. After \( k \) steps we have observed the data

\[ \Psi_k(x, w) = (x, y_1, \ldots, y_k), \]

where

\[ y_1 = w(\psi_1(x)), \ldots, y_k = w(\psi_k(x, y_1, \ldots, y_{k-1})). \]

A decision to stop or to further evaluate \( w \) is made after each step, and the total number of observations of \( w \) is given by

\[ \nu(x, w) = \min\{ k \in \mathbb{N} : \chi_k(\Psi_k(x, w)) = \text{STOP} \}. \]

If \( \nu(x, w) < \infty \) then the data

\[ \Psi(x, w) = \Psi_{\nu(x,w)}(x, w) \]

are used to construct the approximation \( \phi_{\nu(x,w)}(\Psi(x, w)) \).

We only assume measurability of the mappings \( \psi_k, \chi_k, \) and \( \phi_k \). For simplicity, these mappings are defined on the whole spaces \( \mathbb{R}^k \) and \( \mathbb{R}^{k+1} \), respectively. Obviously only
the case \( \nu(X(0), W) < \infty \) with probability one is of practical interest. Then we end up with the method
\[
\overline{X} = \phi_{\nu(X(0), W)}(\Psi(X(0), W)).
\]
As previously we relate the error \( e(\overline{X}) \) to the expected number
\[
n(\overline{X}) = E(\nu(X(0), W))
\]of evaluations of \( W \).

Let \( X^{**} \) denote the class of all methods of the above form, and put
\[
X^{**}_N = \{ \overline{X} \in X^{**} : n(\overline{X}) \leq N \}
\]for \( N \in \mathbb{N} \). Clearly \( \hat{X}_h, \hat{X}^{**}_h \in X^{**} \) for the adaptive methods from Section 2.1 and 2.2. The quantity
\[
e^{**}(N) = \inf\{e(\overline{X}) : \overline{X} \in X^{**}_N\}
\]
is the minimal error that can be obtained by methods that use at most \( N \) sequential observations of \( W \) on the average.

As a subclass \( X^* \subset X^{**} \) we consider all methods that use the same number of observations for all trajectories. Formally this means that the mappings \( \chi_k \) are constant and \( \nu = \min\{k \in \mathbb{N} : \chi_k = \text{STOP}\} \). We put
\[
X^*_N = \{ \overline{X} \in X^* : n(\overline{X}) \leq N \}
\]as well as
\[
e^*(N) = \inf\{e(\overline{X}) : \overline{X} \in X^*_N\}.
\]
Recall that the adaptive method \( \hat{X}_n^* \) from Section 2.3 uses at most \( n \) observations for each trajectory. Hence \( \hat{X}_n^* \in X^*_n \) by a suitable definition of the mappings \( \psi_k, \chi_k, \) and \( \phi_k \).

The subclass \( \overline{X} \subset X^* \) consists of all methods that use the same observation points for every trajectory. Formally the mappings \( \psi_k \) and \( \chi_k \) are constant, such that \( \Psi(x, w) = (x, w(\psi_1), \ldots, w(\psi_\nu)) \). We put
\[
X_N = \{ \overline{X} \in X : n(\overline{X}) \leq N \}
\]as well as
\[
e(N) = \inf\{e(\overline{X}) : \overline{X} \in X_N\}.
\]
Note that \( \hat{X}_n^c \in X_n \) for the Milstein scheme with constant step-size \( 1/n \).

For the analysis of \( e(N) \) we define
\[
C = \int_0^1 \left( E\left( \sigma^2(t, X(t)) \right) \right)^{1/2} dt.
\]
Theorem 4. The minimal errors satisfy
\[
\lim_{N \to \infty} N^{1/2} \cdot e^{**}(N) = C^{**}/\sqrt{6}, \\
\lim_{N \to \infty} N^{1/2} \cdot e^{*}(N) = C^{*}/\sqrt{6}, \\
\lim_{N \to \infty} N^{1/2} \cdot e(N) = C/\sqrt{6}
\]
for every equation (1).

Theorems 1 and 2 match with the first two estimates from Theorem 4.

Corollary 1. The methods \(\hat{X}^{**}_N\) and \(\hat{X}^{*}_N\) are asymptotically optimal in the respective classes \(X^{**}_N\) with \(N = n(\hat{X}^{**}_h)\) and \(X^{*}_n\).

Example 2. For the geometric Brownian motion from Example 1 we obtain
\[
C = 2/b \cdot (\exp(b^2/2) - 1).
\]

Remark 1. The conditional Hölder constant \(|\sigma(t, X(t))|\) describes the smoothness of \(X\) locally in time and space. The constants \(C^{**}\) and \(C^{*}\) are based on the average \(S\) of the conditional Hölder constant along a trajectory, see (7). We have
\[
C^{**} = E(S), \quad C^{*} = (E(S^2))^{1/2}.
\]

Due to (3),
\[
E(X(t + \delta) - X(t))^2 = \alpha^2(t) \cdot \delta + o(\delta)
\]
with
\[
(11) \quad \alpha(t) = \left( E\left(\sigma^2(t, X(t))\right) \right)^{1/2}.
\]
Hence \(\alpha(t)\) describes the smoothness of \(X\) only locally in time. The constants \(C\) and \(C^e\) are based on \(\alpha\). We have
\[
C = \int_0^1 \alpha(t) \, dt, \quad C^e = \left( \int_0^1 \alpha^2(t) \, dt \right)^{1/2}.
\]

Remark 2. Clearly
\[
C^{**} \leq C^{*} \leq C \leq C^e,
\]
and \(C = C^e\) iff \(\alpha\) is constant. Furthermore, \(C^{*} = C\) iff there exist \(t_0 \in [0, 1]\) and \(\gamma \in C([0, 1])\) such that, with probability one,
\[
(12) \quad \forall \, t \in [0, 1]: \quad |\sigma|(t, X(t)) = \gamma(t) \cdot |\sigma|(t_0, X(t_0)).
\]
Note that \(\gamma\) is then determined by
\[
\gamma = \alpha/(E(\sigma^2(t_0, X(t_0)))^{1/2}
\]
if \(C > 0\).

Finally, the Markov property of \(X\) implies that \(C^{**} = C^{*}\) iff, with probability one,
\[
(13) \quad \forall \, t \in [0, 1]: \quad |\sigma|(t, X(t)) = \alpha(t).
\]
Obviously (13) implies (12), and both conditions are equivalent in the case of a nonzero and deterministic $\sigma(0, X(0))$. It would be nice to know whether this equivalence holds for general equations (1).

Note that (13) is satisfied for equations with additive noise, i.e., $\sigma^{(0,1)} = 0$ implies $C^{**} = C$. Moreover, the Milstein scheme and the Euler scheme coincide in this situation. Equations with additive noise were studied in Hofmann et al. (1999a).

**Remark 3.** So far we have not presented an asymptotically optimal method in the class $X_n$, since $C = C^e$ only in exceptional cases. In order to close this gap, we introduce a method $\hat{X}_n$ that is adaptive only with respect to the equation. The method $\hat{X}_n$ coincides with $\hat{X}_n^*$ up to the fact that $\sigma_0, \ldots, \sigma_{k-1}$ are replaced by $\alpha(\tau_0,0), \ldots, \alpha(\tau_{k-1},0)$ in the definition of the numbers $\mu_i$. For every equation (1),

\[
\lim_{n \to \infty} n^{1/2} \cdot e(\hat{X}_n) = C/\sqrt{6},
\]

such that $\hat{X}_n$ is asymptotically optimal in $X_n$.

We stress that $\hat{X}_n$ is not easy to implement, as it requires knowledge of $\alpha$.

**Remark 4.** The error criterion does influence the optimal step-size control. If small errors

\[
e(\bar{X}) = \left( E \|X - \bar{X}\|_\infty^q \right)^{1/q}, \quad 1 \leq q < \infty,
\]

with respect to the $L_\infty$-norm on $[0,1]$ are needed, then steps of length $h/\sigma^2(\tau_\ell, \hat{X}(\tau_\ell))$ seem to be appropriate. At least for (systems of) stochastic differential equations with additive noise, this choice leads to asymptotically optimal methods. See Hofmann et al. (1999b).

For practical purposes, the definition of the step-size control in Section 2.1 must be extended, as $|\sigma(\tau_\ell, \hat{X}(\tau_\ell))$ might be too small or even zero. We suggest to use

\[
\tau_{\ell+1} = \tau_\ell + \min \left( h/|\sigma(\tau_\ell, \hat{X}(\tau_\ell)), h^{2/3} \right)
\]

for errors in $L_2$-norm. See Hofmann et al. (1999a) for a motivation of this choice in case of equations with additive noise. For errors in $L_\infty$-norm $h^{2/3}$ should be replaced by $h \cdot \ln h^{-1}$, see Hofmann et al. (1999b).

6. The Complexity of Stochastic Differential Equations

The $\varepsilon$-complexity $\text{comp}(\varepsilon)$ of a numerical problem is the minimal computational cost to solve the problem with error at most $\varepsilon$. See Traub, Wasilkowski, and Woźniakowski (1988). In this paper the problem is pathwise approximation for a stochastic differential equation, and the error is the mean squared $L_2$-distance. For a particular method and trajectory the cost is determined by the following quantities
1) the number of evaluations of $W$,
2) the number of evaluations of the drift and diffusion coefficients,
3) the computational cost to evaluate the mappings $\psi_k$, $\chi_k$, and $\phi_k$.

For every method $X$ we let $c(X)$ denote the expected cost over all trajectories. Clearly
\[ n(X) \leq c(X), \]
but often $c(X)$ is much larger than $n(X)$. For instance, ordinary differential equations or optimization problems are solved in 3) or weak approximations for stochastic differential equations are computed. See Remark 3 for a method of the latter kind.

Now we look at the adaptive method $\hat{X}_h^{**}$. By Lemma 6, $n(\hat{X}_h^{**})$ is of order $1/h$, and $k_h = o(1/h)$ by (5). Hence the cost from 2) are irrelevant asymptotically. We need one arithmetical operation to obtain the Brownian increment and one more operation to accumulate the Brownian increment for the Milstein step. Moreover, three arithmetical operations are needed per Euler step. Finally the discretization point is updated and a comparison is made. It is reasonable to assume that an arithmetical operation is not more expensive than an evaluation of $W$. We thus conclude that
\[ \lim_{h \to \infty} c(\hat{X}_h^{**})/n(\hat{X}_h^{**}) = 8. \]

Hence the cost of $\hat{X}_h^{**}$ is determined by its cost from 1), up to a very small multiplicative constant. Together with Theorems 1 and 4 this yields the following result.

**Theorem 5.** The $\varepsilon$-complexity of every equation (1) satisfies
\[ \liminf_{\varepsilon \to 0} \varepsilon^2 \cdot \text{comp}(\varepsilon) \geq (C^{**})^2 / 6 \]
and
\[ \limsup_{\varepsilon \to 0} \varepsilon^2 \cdot \text{comp}(\varepsilon) \leq (C^{**})^2 \cdot 4 / 3. \]

**Remark 5.** Although the methods $\hat{X}_h^{**}$ and $\hat{X}_n^*$ are adaptive, they are very well suited for parallel processing. In fact, only the evaluation of $W$ at the points $\tau_{i,0}$ must be done sequentially. Thereafter the Euler steps can be performed in parallel on all subintervals.

### 7. Simulation Experiments

So far we have presented an asymptotic analysis where $n$, the (expected) number of evaluations of $W$, tends to infinity. Let us now illustrate the practical relevance of these results for moderate size of $n$ or $\varepsilon$, respectively, by means of a simulation experiment. We take the geometric Brownian motion from Example 1 with
\[ b = 3. \]

Note that the asymptotic constants are known in this case.
For the fully adaptive method $\hat{X}_h$ the error $e$ as well as $n$ are shown in Table 1 for several values of $h$. Both quantities are determined by extensive simulations (simul.). For comparison we take the simplified adaptive method $\hat{X}_h^{**}$, where we use the asymptotic formulas (asymp.) for $e$ and $n$, see Theorem 1 and Lemma 6. Table 1 supports our conjecture that Theorem 1 also holds for $\hat{X}_h$ instead of $\hat{X}_h^{**}$.

For several methods or classes of methods we show how $n$ depends on the error, which varies within the reasonable range $[1 \cdot 10^{-3}, 2 \cdot 10^{-2}]$. See Figure 1. Solid or dotted lines are computed by means of asymptotic formulas, and $\star$ corresponds to the simulated values from Table 1. Finally + and $\circ$ are computed by use of explicit formulas for $e(\hat{X}_n^{**})$ and $n(\hat{X}_n^{**})$ as well as $e(\hat{X}_n)$ and $n(\hat{X}_n)$. Such explicit formulas can be derived in the particular case of the geometric Brownian motion.

The fully adaptive method gets very close to the asymptotic lower bound from Theorem 5 (complexity). The simplified adaptive method is not quite as good as the fully adaptive one. A significant loss of performance occurs if we require that $n$ is fixed for all trajectories; the second asymptotic formula from Theorem 4 is used here. For equidistant discretizations the asymptotic formula from Theorem 3 yields a very precise approximation for the error $e(\hat{X}_n)$.

We conclude in particular that the computation time increases by a factor close to

$$(\exp(9) - 1)/9 = 900.2\ldots,$$

if we wish to achieve the same error by an equidistant discretization instead of a fully adaptive one. The absolute values of $n$ demonstrate the practical relevance of this fact.

Finally we consider again the fully adaptive method $\hat{X}_h$. We show that the number $\nu(1, w)$ of evaluations of the trajectory $w$ strongly depends on $w$. To this end relative frequencies are presented in Figure 2 in the case $h = 2 \cdot 10^{-4}$. Recall that the mean $n(\hat{X}_n^{**})$ of $\nu$ is about 15 000, see Table 1. The median of $\nu$ is about 4 450. Two percent of the trajectories required less than 1000 evaluations. Less than 0.1% of the trajectories

<table>
<thead>
<tr>
<th>$h$</th>
<th>$e$</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$5 \cdot 10^{-4}$</td>
<td>$1.82 \cdot 10^{-2}$</td>
<td>6009</td>
</tr>
<tr>
<td>$4 \cdot 10^{-4}$</td>
<td>$1.59 \cdot 10^{-2}$</td>
<td>7624</td>
</tr>
<tr>
<td>$2 \cdot 10^{-4}$</td>
<td>$1.04 \cdot 10^{-2}$</td>
<td>14981</td>
</tr>
<tr>
<td>$1 \cdot 10^{-4}$</td>
<td>$7.09 \cdot 10^{-3}$</td>
<td>29843</td>
</tr>
<tr>
<td>$6 \cdot 10^{-5}$</td>
<td>$5.55 \cdot 10^{-3}$</td>
<td>51297</td>
</tr>
<tr>
<td>$3 \cdot 10^{-5}$</td>
<td>$3.81 \cdot 10^{-3}$</td>
<td>100122</td>
</tr>
</tbody>
</table>

Table 1. Error and expected number of evaluations for $\hat{X}_h$ and $\hat{X}_h^{**}$
required more than $10^6$ evaluations. The maximal number $\nu(1, w)$ that occurred in the simulation was about $1.41 \cdot 10^7$, and the empirical deviation of $\nu(1, w)$ was $9.2 \cdot 10^4$.

8. Proofs

Instead of estimating $X - \overline{X}$ directly, we introduce processes $\hat{X}_m$ and consider $X - \hat{X}_m$ as well as $\hat{X}_m - \overline{X}$ separately. For $m \in \mathbb{N}$ let

$$t_\ell = \ell/m, \quad \ell = 0, \ldots, m.$$
Since the latter are defined by quantiles with respect to a fixed density. Here we replace Theorems 1, 2, 4 and (14).

Lemma 1. \( \left( \sum \right) \)

Let \( A \) denote the interval \( [t_{\ell}, t_{\ell+1}] \). Note that \( \tilde{X} \) coincides with the Milstein scheme at the discretization points \( t_\ell \). Complete knowledge of a trajectory of \( \tilde{X} \) from conditions (A) as well as on \( a(t, \tilde{X}(t)) \), \( \sigma(t, \tilde{X}(t)) \) and \( (W(t) - W(t_\ell)) \) is needed to generate the corresponding trajectory of \( X \). The error is then dominated by \( E(X(t) - \tilde{X}_n(t))^2 \). In this way we prove Theorems 1, 2, 4 and (14).

For Theorem 3 we apply a well known technique for regular sequences of discretizations. The latter are defined by quantiles with respect to a fixed density. Here we replace \( \tilde{X} \) by the linear interpolation \( \tilde{X}_n \) of the solution at the discretization. The error is then dominated by \( E(X(t) - \tilde{X}_n(t))^2 \).

In the sequel we let \( c \) denote unspecified positive constants, which only depend on the constant \( K \) from conditions (A) as well as on \( a(0, 0), \sigma(0, 0), \) and \( E(X(0))^4 \).

8.1. Approximation of Brownian bridges. Let \( B \) denote a Brownian bridge on an interval \( [S, T] \) with mean zero and variance \((T - t) \cdot (t - S)/(T - S)\) at \( t \). Let \( \tilde{B}_m \) denote the piecewise linear interpolation of \( B \) at the points \((T - S) \cdot \ell/(m + 1)\) where \( \ell = 1, \ldots, m \). Then straightforward calculations yield

\[
E \left( \int_{S}^{T} \left( B(t) - \tilde{B}_m(t) \right)^2 \, dt \right) = \frac{(T - S)^2}{6 \cdot (m + 1)}.
\]

8.2. Proof of the lower bounds in Theorem 4. Consider an arbitrary sequence of methods \( \tilde{X}_N \in \mathbb{X}_N^{**} \). Take a sequence of positive integers \( k_N \) such that

\[
\lim_{N \to \infty} N^{1/2} / k_N = \lim_{N \to \infty} k_N / N = 0.
\]

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

\[
t_\ell = \ell / k_N, \quad \ell = 0, \ldots, k_N.
\]

Let \( \mathcal{A} \) denote the \( \sigma \)-algebra that is generated by \( \Psi(X(0), W) \). Moreover, put \( U_\ell = (t_\ell, \tilde{X}_{k_N}(t_\ell)) \).

Lemma 1.

\[
\liminf_{N \to \infty} N \cdot e(\tilde{X}_N)^2 \geq \liminf_{N \to \infty} N \cdot \sum_{\ell=0}^{k_N-1} \int_{t_\ell}^{t_{\ell+1}} E \left( \sigma^2(U_\ell) \cdot (W(t) - E(W(t) \mid \mathcal{A}))^2 \right) \, dt.
\]
Proof. Theorem 6 implies
\[ \int_0^1 E(X(t) - \bar{X}(t))^2 \, dt \leq c/k_N^2 = o(N^{-1}). \]

It is therefore sufficient to analyze \( \bar{X} \).

Clearly
\[ E \left( \bar{X}_k(t) - \bar{X}_N(t) \right)^2 \geq E \left( \bar{X}_k(t) - \bar{X}_N(t) \right)^2. \]

Let \( t \in [t_\ell, t_{\ell+1}] \). Since \( U_\ell \) is \( \mathcal{A} \)-measurable we get
\[
\bar{X}_k(t) - E(\bar{X}_k(t) | \mathcal{A}) = \sigma(U_\ell) \cdot (W(t) - E(W(t) | \mathcal{A})) + 1/2 \cdot \left( \sigma \cdot \sigma^{(0,1)}(U_\ell) \cdot (W(t) - E(W(t) | \mathcal{A}))^2 - E((W(t) - W(t_\ell))^2 | \mathcal{A})) \right).
\]

By boundedness of \( \sigma^{(0,1)} \) and linear growth of \( \sigma \),
\[
E \left( (\sigma \cdot \sigma^{(0,1)})(U_\ell) \cdot ((W(t) - W(t_\ell))^2 - E((W(t) - W(t_\ell))^2 | \mathcal{A})) \right) \leq c \cdot E(\sigma^2(U_\ell) \cdot (W(t) - W(t_\ell))^4) \leq c \cdot \left( 1 + E \left( \bar{X}_k(t_\ell) \right) \right) \cdot E(\sigma^2(W(t) - W(t_\ell))^4) \leq c/k_N^2,
\]

where the last estimate follows from Lemma 10. We conclude that
\[
\left( \int_0^1 E(\bar{X}_k(t) - \bar{X}_N(t))^2 \, dt \right)^{1/2} \geq \left( \sum_{\ell=0}^{k_N-1} \int_{t_\ell}^{t_{\ell+1}} E(\sigma^2(U_\ell) \cdot (W(t) - E(W(t) | \mathcal{A}))^2) \, dt \right)^{1/2} - c/k_N.
\]

This completes the proof, since \( k_N^{-1} = o(1) \).

According to Lemma 1 the error of every method \( \bar{X}_N \) is bounded from below by an error for weighted \( L_2 \)-approximation of the Brownian motion. The weight function is a stochastic process itself, and it is close to \( \sigma^2(t, X(t)) \). The set of observation points is determined by \( \bar{X}_N \), namely,
\[ D(X(0), W) = \{ \psi_1(X(0)), \ldots, \psi_{\nu(W)}(X(0), \Psi_{\nu(W)}(X(0), W) - 1)(X(0), W) \}. \]

Let
\[ d_\ell(X(0), W) = \# (D(X(0), W) \cap [t_\ell, t_{\ell+1}]) + 1, \]
where \# denotes the cardinality of a set. Observe that \(d_\ell(X(0), W)\) is \(\mathfrak{A}\)-measurable. Put

\[
A_{kN}(X(0), W) = \sum_{\ell=0}^{kN-1} \sigma^2(t_\ell, X(t_\ell)) d_\ell(X(0), W).
\]

**Lemma 2.**

\[
\lim \inf_{N \to \infty} \frac{N}{kN - 1} \sum_{\ell=0}^{kN-1} \int_{t_\ell}^{t_{\ell+1}} E \left( \sigma^2(U_\ell) \cdot (W(t) - E(W(t) \mid \mathfrak{A}))^2 \right) dt \geq \lim \inf_{N \to \infty} N/ \left( 6 k_N^2 \right) \cdot E \left( A_{kN}(X(0), W) \right).
\]

**Proof.** Clearly

\[
E \left( \sigma^2(U_\ell) \cdot (W(t) - E(W(t) \mid \mathfrak{A}))^2 \mid \mathfrak{A} \right) = \sigma^2(U_\ell) \cdot E \left( (W(t) - E(W(t) \mid \mathfrak{A}))^2 \mid \mathfrak{A} \right).
\]

Conditioned on \(\mathfrak{A}\), the discretization \(D(X(0), W)\) is fixed and the process \(W(t) - E(W(t) \mid \mathfrak{A})\) is a Brownian bridge on each of the corresponding subintervals. Hence

\[
\int_{t_\ell}^{t_{\ell+1}} E \left( (W(t) - E(W(t) \mid \mathfrak{A}))^2 \mid \mathfrak{A} \right) dt \geq \frac{1}{6 k_N^2 \cdot d_\ell(X(0), W)},
\]

due to (16), and consequently

\[
\int_{t_\ell}^{t_{\ell+1}} E \left( \sigma^2(U_\ell) \cdot (W(t) - E(W(t) \mid \mathfrak{A}))^2 \right) dt \geq 1/ \left( 6 k_N^2 \right) \cdot E \left( \frac{\sigma^2(U_\ell)}{d_\ell(X(0), W)} \right).
\]

By (A),

\[
|\sigma^2(U_\ell) - \sigma^2(t_\ell, X(t_\ell))| \leq c \cdot |\tilde{X}_{kN}(t_\ell) - X(t_\ell)| \cdot (1 + |\tilde{X}_{kN}(t_\ell)| + |X(t_\ell)|).
\]

Theorem 6, Lemma 10 and (10) yield

\[
(17) \quad E \left| \sigma^2(U_\ell) - \sigma^2(t_\ell, X(t_\ell)) \right| \leq c/k_N.
\]

Since \(k_N^2 = o(N^{-1})\) and \(d_\ell(X(0), W) \geq 1\), the lemma follows. \(\Box\)

Now we consider the three different classes \(X^{**}, X^*,\) and \(X\) of methods separately.

**Lemma 3.** If \(X_N \in X_N\) for every \(N\) then

\[
\lim \inf_{N \to \infty} N/k_N^2 \cdot E \left( A_{kN}(X(0), W) \right) \geq C^2.
\]

**Proof.** By definition of \(X_N\), the numbers \(d_\ell(X(0), W)\) are constant with

\[
\sum_{\ell=0}^{kN-1} d_\ell \leq N.
\]
The Cauchy-Schwarz inequality yields

$$\frac{N}{k_N^2} \cdot E(A_{k_N}(X(0), W)) \geq 1/k_N^2 \cdot \sum_{\ell=0}^{k_N-1} \frac{\sigma^2(t_{\ell}, X(t_{\ell}))}{d_{\ell}} \cdot \sum_{\ell=0}^{k_N-1} d_{\ell}$$

$$\geq 1/k_N^2 \cdot \left( \sum_{\ell=0}^{k_N-1} \left( E(\sigma^2(t_{\ell}, X(t_{\ell}))) \right)^{1/2} \right)^2.$$

The right-hand side tends to $C^2$. \hfill \Box

**Lemma 4.** If $X_N \in X_N^*$ for every $N$ then

$$\liminf_{N \to \infty} \frac{N}{k_N^2} \cdot E(A_{k_N}(X(0), W)) \geq (C^*)^2.$$

**Proof.** By definition of $X_N^*$,

$$\sum_{\ell=0}^{k_N-1} d_{\ell}(X(0), W) \leq N.$$

The Cauchy-Schwarz inequality yields

$$N \cdot A_{k_N}(X(0), W) \geq \sum_{\ell=0}^{k_N-1} \frac{\sigma^2(t_{\ell}, X(t_{\ell}))}{d_{\ell}(X(0), W)} \cdot \sum_{\ell=0}^{k_N-1} d_{\ell}(X(0), W)$$

$$\geq \left( \sum_{\ell=0}^{k_N-1} |\sigma(t_{\ell}, X(t_{\ell}))| \right)^2.$$

Thus

$$\liminf_{N \to \infty} \frac{N}{k_N^2} \cdot E(A_{k_N}(X(0), W)) \geq E \left( \lim_{N \to \infty} \frac{1}{k_N} \cdot \sum_{\ell=0}^{k_N-1} |\sigma(t_{\ell}, X(t_{\ell}))| \right)^2 = (C^*)^2$$

by Fatou’s Lemma and by continuity of $X$ and $\sigma$. \hfill \Box

**Lemma 5.** If $X_N \in X_N^{**}$ for every $N$ then

$$\liminf_{N \to \infty} \frac{N}{k_N^2} \cdot E(A_{k_N}(X(0), W)) \geq (C^{**})^2.$$

**Proof.** By definition of $X_N^{**}$,

$$\sum_{\ell=0}^{k_N-1} E(d_{\ell}(X(0), W)) \leq N.$$
The Cauchy-Schwarz inequality yields
\[
N \cdot E(A_{k_N}(X(0), W)) \geq \sum_{\ell=0}^{k_N-1} E\left( \frac{\sigma^2(t_\ell, X(t_\ell))}{d_\ell(X(0), W)} \right) \cdot \sum_{\ell=0}^{k_N-1} E(d_\ell(X(0), W))
\]
\[
\geq \left( \sum_{\ell=0}^{k_N-1} E\left( \frac{\sigma^2(t_\ell, X(t_\ell))}{d_\ell(X(0), W)} \right) \right)^{1/2} \cdot \left( E(d_\ell(X(0), W)) \right)^{1/2}
\]
\[
\geq \left( \sum_{\ell=0}^{k_N-1} E(|\sigma|)(t_\ell, X(t_\ell))) \right)^2.
\]
Thus
\[
\liminf_{N \to \infty} \frac{N}{k_N^2} \cdot E(A_{k_N}(X(0), W)) \geq \left( \lim_{N \to \infty} \frac{1}{k_N} \cdot \sum_{\ell=0}^{k_N-1} E(|\sigma|)(t_\ell, X(t_\ell))) \right)^2 = (C^{**})^2,
\]
as claimed.

We combine Lemma 1 and 2 with Lemma 3–5 to obtain the lower bounds in Theorem 4. Moreover, if \( N \) is chosen appropriately then the first two lower bounds from this theorem yield the lower bounds in Theorems 1 and 2.

8.3. Proof of the upper bounds in Theorems 1 and 2 and (14). First we consider the method \( \hat{X}^{**}_h \). Let
\[
t_\ell = \tau_{\ell,0} = \ell/k_h
\]
and consider the corresponding process \( \hat{X}_{k_h} \). Observe that \( \hat{X}_{k_h} \) and \( \hat{X}^{**}_h \) coincide at the points \( t_\ell \). Recall that \( \sigma_\ell = \sigma(t_\ell, X_k(t_\ell)) \).

Asymptotically, the expected number of observations \( n(\hat{X}^{**}_h) \) increases like \( 1/h \cdot C^{**} \).

**Lemma 6.**
\[
\lim_{h \to 0} h \cdot n(\hat{X}^{**}_h) = \lim_{h \to 0} \frac{1}{k_h} \cdot \sum_{\ell=0}^{k_h-1} E(|\sigma_\ell|) = C^{**}.
\]

**Proof.** Note that
\[
n(\hat{X}^{**}_h) \leq k_h + 1/(h \cdot k_h) \cdot \sum_{\ell=0}^{k_h-1} E(|\sigma_\ell|).
\]
Hence \( n(\hat{X}^{**}_h) < \infty \) due to (A) and Lemma 10. Moreover,
\[
E |\sigma_\ell - \sigma(t_\ell, X(t_\ell))| \leq c/k_h
\]
due to (A) and Theorem 6. Clearly \( \lim_{h \to 0} k_h = \infty \) by (6). Use (5) to conclude that
\[
\limsup_{h \to 0} h \cdot n(\hat{X}^{**}_h) \leq \limsup_{h \to 0} \frac{1}{k_h} \cdot \sum_{\ell=0}^{k_h-1} E(|\sigma_\ell|) \leq C^{**}.
\]
The lower bounds are established in a similar way.

Because of Lemma 6 the upper bound in Theorem 1 reads
\[ \limsup_{h \to 0} h^{-1} \cdot e(\hat{X}_k^*)^2 \leq C^*/6. \]

Theorem 6 and (6) imply
\[ E\left( \int_0^1 (X(t) - \hat{X}_k(t))^2 \, dt \right) \leq c/k^2 h = o(h). \]

Hence the upper bound in Theorem 1 is a consequence of the following estimate.

Lemma 7.
\[ \limsup_{h \to 0} h^{-1} \cdot E\left( \int_0^1 (X_{\gamma h}(t) - \hat{X}_k^*(t))^2 \, dt \right) \leq C^*/6. \]

Proof. Fix \( \ell \in \{0, \ldots, k - 1\} \) and let \( \mathcal{B} \) denote the \( \sigma \)-algebra that is generated by \( X(0), W(t_1), \ldots, W(t_\ell) \). Define
\[ d_\ell = \min(1, \lceil |\sigma_\ell|/(h \cdot k) \rceil) \]
and
\[ V_h(t) = 1/2 \cdot (\sigma \cdot (0,1)) (t, X_{\gamma h}(t)) \cdot ((W(t) - W(t_\ell))^2 - (t - t_\ell)) \]
for \( t \in [t_\ell, t_{\ell+1}] \). The adaptive discretization yields points
\[ t_\ell = \tau_{\ell,0} < \tau_{\ell,1} < \cdots < \tau_{\ell,d_\ell - 1} < \tau_{\ell,d_\ell} = t_{\ell+1}, \]
which are measurable with respect to \( \mathcal{B} \). Let \( \tilde{W}_h \) denotes the piecewise linear interpolation of \( W - W(t_\ell) \) at these points.

Note that
\[ X_{\gamma h}(t) - \hat{X}_k^*(t) = \sigma_\ell \cdot (W(t) - W(t_\ell) - \tilde{W}_h(t)) + V_h(t) \]
if \( t_\ell \leq t \leq \tau_{\ell,d_\ell - 1} \), and
\[ |X_{\gamma h}(t) - \hat{X}_k^*(t)| \leq |\sigma_\ell| \cdot |W(t) - W(t_\ell) - \tilde{W}_h(t)| + |V_h(t)| + |V_h(t_{\ell+1})| \]
if \( \tau_{\ell,d_\ell - 1} < t < t_{\ell+1} \). Conditioned on \( \mathcal{B} \) the discretization is fixed and the process \( W(t) - W(t_\ell) - \tilde{W}_h(t) \) is a Brownian Bridge on each of the subintervals \([\tau_{\ell,j}, \tau_{\ell,j+1}]\). Hence
\[ \int_{t_\ell}^{t_{\ell+1}} E\left( \sigma_\ell^2 \cdot \left( W(t) - W(t_\ell) - \tilde{W}_h(t) \right)^2 \mid \mathcal{B} \right) \, dt \leq (d_\ell - 1) \cdot h^2/6 + c \cdot h^2 \leq |\sigma_\ell| \cdot h/(6 \cdot k) + c \cdot h^2 \]
by (16), and we obtain
\[
\begin{align*}
&h^{-1} \cdot \sum_{\ell=0}^{k_h-1} \int_{t_\ell}^{t_{\ell+1}} E \left( \sigma_\ell^2 \cdot \left( W(t) - W(t_\ell) - \tilde{W}_h(t) \right)^2 \right) dt \\
&\quad \leq 1/(6 \cdot k_h) \cdot \sum_{\ell=0}^{k_h-1} E(\|\sigma_\ell\|) + c \cdot h,
\end{align*}
\]
which tends to $C^{**}/6$, see Lemma 6.

It remains to observe that
\[
\sup_{t \in [0, 1]} E(|V_h(t)|) \leq c/k_h = o(h^{1/2})
\]
follows from (A), Lemma 10, and (6). \qed

Now we analyze the method $\hat{X}_n$. Let
\[
t_\ell = \tau_{\ell,0} = \ell/k_n
\]
and consider the corresponding process $\hat{X}_n$. Due to Theorem 6 and (9) the upper bound in Theorem 2 is a consequence of the following estimate.

Lemma 8.
\[
\limsup_{n \to \infty} n \cdot E \left( \int_0^1 \left( \hat{X}_{k_n}(t) - \hat{X}_n^*(t) \right)^2 dt \right) \leq C^*/6.
\]

Proof. We proceed as in the proof of Lemma 7. Let $t \in [t_\ell, t_{\ell+1}]$ and define $V_n$ by the right-hand side of (18) with $k_n$ instead of $k_h$. Then $E(|V_n(t)|) = o(n^{-1/2})$ uniformly in $t \in [0, 1]$. Let $\mathcal{B}$ denote the $\sigma$-algebra that is generated by $X(0), W(t_1), \ldots, W(t_\ell)$, and let $\tilde{W}_n$ denote the piecewise linear interpolation of $W - W(t_\ell)$ at the points $\tau_{\ell,0}, \ldots, \tau_{\ell+1,0}$. We use (16) to obtain
\[
\int_{t_\ell}^{t_{\ell+1}} E \left( \sigma_\ell^2 \cdot \left( W(t) - W(t_\ell) - \tilde{W}_n(t) \right)^2 \mid \mathcal{B} \right) dt \leq \frac{|\sigma_\ell| \cdot \sum_{\ell=0}^{k_n-1} |\sigma_\ell|}{6 \cdot k_n^2 \cdot (n - k_n)}.
\]
Thus
\[
n \cdot \sum_{\ell=0}^{k_n-1} \int_{t_\ell}^{t_{\ell+1}} E \left( \sigma_\ell^2 \cdot \left( W(t) - W(t_\ell) - \tilde{W}_n(t) \right)^2 \right) dt \leq \frac{E \left( \sum_{\ell=0}^{k_n-1} |\sigma_\ell| \right)^2}{6 \cdot k_n^2} \cdot \frac{n}{n - k_n}.
\]
Due to (A), Theorem 5, and (8) the right-hand side tends to $C^*/6$. \qed

Finally we turn to the method $\hat{X}_n$, see Remark 3. Recall that $\hat{X}_n$ coincides with the method $\hat{X}_n^*$ except for the deterministic choice of the numbers $\mu_\ell$. In view of the
arguments given above, it thus suffices to show that

$$\limsup_{n \to \infty} \frac{n}{k_n^2} \cdot \sum_{\ell=0}^{k_n-1} E(\sigma_\ell^2)/(\mu_\ell + 1) \leq C.$$  

Proof. Observing (17) and the definition of $\mu_\ell$ we have

$$E(\sigma_\ell^2)/(\mu_\ell + 1) \leq \frac{\alpha_\ell^2 + c/k_n}{\mu_\ell + 1} \leq \frac{\alpha_\ell \cdot k_n}{n/(n - k_n) + c/k_n}.$$

Hence

$$\frac{n}{k_n^2} \cdot \sum_{\ell=0}^{k_n-1} E(\sigma_\ell^2)/(\mu_\ell + 1) \leq \left(\frac{1}{k_n} \cdot \sum_{\ell=0}^{k_n-1} \alpha_\ell\right)^2 \cdot n/(n - k_n) + c \cdot n/k_n^2.$$  

Now use (8) and (9) to obtain (19).

Obviously the upper bounds from Theorems 1 and 2 imply the upper bounds in the first two estimates from Theorem 4, and the upper bound from (14) matches the upper bound in the last estimate from Theorem 4.

8.4. Proof of Theorem 3. Consider fixed points

$$0 = \tau_0^{(n)} < \cdots < \tau_n^{(n)} = 1,$$

and let $\tilde{X}_n$ denote the piecewise linear interpolation of $X$ at these points. Obviously $\tilde{X}_n$ is an implementable numerical scheme only in exceptional cases, as it requires exact knowledge of $X$ at discrete points. The Milstein scheme $\hat{X}$ that is based on the same discretization satisfies

$$\sup_{t \in [0,1]} E \left(\hat{X}(t) - \tilde{X}_n(t)\right)^2 \leq \sup_{\ell=0,\ldots,n} E \left(\hat{X}(\tau_\ell^{(n)}) - \tilde{X}_n(\tau_\ell^{(n)})\right)^2 \leq c \cdot (\Delta_{\text{max}}^{(n)})^2$$

due to Theorem 6.

Let $\gamma$ denote a continuous and strictly positive function on $[0,1]$, and suppose that

$$\int_0^{\tau_n} \gamma(t) \, dt = \ell/n \cdot \int_0^1 \gamma(t) \, dt.$$  

In this case the discretizations (20) form a so called regular sequence. For strong approximation of stochastic differential equations regular sequences were introduced in Cambanis and Hu (1996). We are interested in particular in $\gamma = 1$ and $\gamma = \alpha$, see (11). Clearly $\Delta_{\text{max}}^{(n)} = O(1/n)$ for every regular sequence. Thus it suffices to estimate the difference between $X$ and $\tilde{X}_n$. We can use known techniques to determine the error of piecewise linear interpolation based on a regular sequence.

To this end we need the following fact about the local smoothness of $X$.
Lemma 9. Let
\[ \Gamma = E((X(s_2) - X(s_1)) \cdot (X(s_4) - X(s_3))) \]
and
\[ \Delta = \max(s_2 - s_1, s_4 - s_3), \]
where \( s_1 \leq s_2 \) and \( s_3 \leq s_4 \). Then
\[ |\Gamma| \leq c \cdot \Delta^2 \]
if \( s_2 \leq s_3 \), and
\[ |\Gamma - \alpha^2(s_1) \cdot \Delta| \leq c \cdot \Delta^2 \]
if \( s_1 = s_3 \) and \( s_2 = s_4 \).

Proof. We use
\[ X(t) - X(s) = Y(t) - Y(s) + V(t) - V(s), \]
where
\[ Y(t) = \int_0^t a(u, X(u)) \, du \]
and
\[ V(t) = \int_0^t \sigma(u, X(u)) \, dW(u). \]
Assume that \( s_2 \leq s_3 \). Then
\[ E((V(s_2) - V(s_1)) \cdot (V(s_4) - V(s_3))) = 0, \]
since \( V \) is an \( L_2 \)-martingale. Together with (A) and (10) this yields the first estimate. Moreover,
\[ E(V(t) - V(s))^2 = \int_s^t E(\sigma^2(u, X(u))) \, du. \]
and
\[ E(Y(t) - Y(s))^2 \leq c \cdot (t - s)^2. \]
Using (A) we get the second estimate. \( \square \)

Lemma 9 implies
\[ \lim_{n \to \infty} n^{1/2} \cdot \left( E\|X - \tilde{X}_n\|^2 \right)^{1/2} = 1/\sqrt{6} \cdot \left( \int_0^1 \alpha^2/\gamma(t) \, dt \right)^{1/2} \cdot \left( \int_0^1 \gamma(t) \, dt \right)^{1/2}. \]
See Su and Cambanis (1993). Take \( \gamma = 1 \) to obtain Theorem 3. Suppose that \( \alpha(t) > 0 \)
for all \( t \in [0, 1] \). Then \( \gamma = \alpha \) yields asymptotically optimal methods in the class \( \mathbb{X}_n \), see Theorem 4. Obviously, the regular sequence generated by \( \alpha \) and the discretizations used by the method \( \tilde{X}_n \) from Remark 3 almost coincide. Both discretizations are not easily implemented, as they require knowledge of \( \alpha \).
Appendix

Choose fixed points

\[ 0 = t_0 < \cdots < t_m = 1 \]

and consider the process \( \tilde{X} = \tilde{X}_m \) for this discretization, see (15). Put

\[ \Delta_\ell = t_{\ell+1} - t_\ell. \]

We derive a uniform upper bound for \( E \left( X(t) - \tilde{X}(t) \right)^2 \) in terms of

\[ \Delta_{\text{max}} = \max_{\ell=0, \ldots, m-1} \Delta_\ell. \]

We rely on this estimate in the analysis of numerical methods. Recall that \( \tilde{X} \) itself cannot be used as a numerical method for the global approximation of \( X \).

Faure (1992) presents the upper bound in the special case of constant step-size \( \Delta_{\text{max}} = 1/m \). Our proof is completely different from his approach. Moreover, it is based on weaker assumptions (see Section 3). In particular we do not need the existence of second derivatives with respect to the state variable.

As previously, \( c \) denotes unspecified positive constants, which only depend on the constant \( K \) from (A) as well as on \( a(0,0), \sigma(0,0), \) and \( E(X(0))^4 \).

Lemma 10.

\[ \sup_{t \in [0,1]} E \left( \tilde{X}(t) \right)^4 \leq c \]

and

\[ \sup_{t \in [t_\ell, t_{\ell+1}]} E \left( \tilde{X}(t) - \tilde{X}(t_\ell) \right)^4 \leq c \cdot \Delta_\ell^2. \]

Proof. Put \( \|Y\| = (E|Y|^4)^{1/4} \) and define

\[ Z(t) = \sum_{\ell=0}^{m-1} \sigma(t_\ell, \tilde{X}(t_\ell)) \cdot 1_{[t_\ell, t_{\ell+1}]}(t) \]

as well as

\[ g(t) = \sup_{0 \leq s \leq t} \|\tilde{X}(s)\|. \]

Let \( t \in [t_\ell, t_{\ell+1}] \). Then

\[ \|\tilde{X}(t) - \tilde{X}(t_\ell)\| \leq c \cdot (1 + \|\tilde{X}(t_\ell)\|) \cdot \Delta_\ell^{1/2}, \quad (21) \]
such that $g(1) < \infty$ follows from $\|X(0)\| < \infty$. Furthermore,

\[
\|\tilde{X}(t)\| \\
\leq c \cdot \sum_{j=0}^{\ell-1} (1 + \|\tilde{X}(t_j)\|) \cdot \Delta_j + c \cdot (1 + \|\tilde{X}(t_\ell)\|) \cdot (t - t_\ell) + \left\| \int_0^t Z(s) \, dW(s) \right\|
\]

\[
\leq c \cdot \left(1 + \int_0^t g(s) \, ds\right) + \left\| \int_0^t Z(s) \, dW(s) \right\|.
\]

Observe that

\[
\|Z(t)\| \leq c \cdot \left(1 + \sup_{j=0,\ldots,\ell} \|\tilde{X}(t_j)\|\right) \leq c \cdot (1 + g(t)).
\]

A well-known martingale moment inequality yields

\[
\left\| \int_0^t Z(s) \, dW(s) \right\|^4 \leq 36 \cdot t \cdot \int_0^t \|Z(s)\|^4 \, ds.
\]

We conclude that

\[
\|\tilde{X}(t)\| \leq c \cdot \left(1 + \left(\int_0^t g^4(s) \, ds\right)^{1/4}\right)
\]

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

\[
g^4(t) \leq c \cdot \left(1 + \int_0^t g^4(s) \, ds\right),
\]

and Gronwall’s Lemma yields

\[
g(t) \leq c.
\]

This completes the proof of the first estimate. The second estimate is an immediate consequence of (21) and the first estimate. \(\square\)

**Theorem 6.**

\[
\sup_{t \in [0,1]} E \left( \left( X(t) - \tilde{X}(t) \right)^2 \right) \leq c \cdot \Delta_{\max}^2.
\]

**Proof.** Put $U_\ell = (t_\ell, \tilde{X}(t_\ell))$. By definition, $X(t) = X(0) + A(t) + B(t)$ where

\[
A(t) = \int_0^t a(s, X(s)) \, ds
\]

and

\[
B(t) = \int_0^t \sigma(s, X(s)) \, dW(s).
\]

Similarly, $\tilde{X}(t) = X(0) + \tilde{A}(t) + \tilde{B}(t)$ where

\[
\tilde{A}(t) = \int_0^t \sum_{\ell=0}^{m-1} a(U_\ell) \cdot 1_{[t_\ell, t_{\ell+1})}(s) \, ds
\]
and
\[ \hat{B}(t) = \int_0^t \sum_{\ell=0}^{m-1} \sigma(U_\ell) \cdot \left( 1 + \sigma^{(0,1)}(U_\ell) \cdot (W(s) - W(t_\ell)) \right) \cdot 1_{[t_\ell, t_{\ell+1}]}(s) \, dW(s). \]

For estimation of \( A - \hat{A} \) we define
\[ Z(t) = \sum_{\ell=0}^{m-1} V_\ell(t) \cdot 1_{[t_\ell, t_{\ell+1}]}(t), \]
where
\[ V_\ell(t) = a^{(0,1)}(U_\ell) \cdot (\hat{X}(t) - \hat{X}(t_\ell) - a(U_\ell) \cdot (t - t_\ell)). \]

Let \( t \in [t_\ell, t_{\ell+1}] \). Clearly
\[ a(t, X(t)) = a(U_\ell) - V_\ell(t) = a(t, X(t)) - a(t_\ell, X(t)) + a(t_\ell, X(t)) - a(U_\ell) - a^{(0,1)}(U_\ell) \cdot (\hat{X}(t) - \hat{X}(t_\ell)) + a^{(0,1)}(U_\ell) \cdot a(U_\ell) \cdot (t - t_\ell). \]

Lemma 10 and (10) yield
\[ E \left( a(t, X(t)) - a(U_\ell) - V_\ell(t) \right)^2 \leq c \cdot \left( \Delta^2 + E(X(t) - \hat{X}(t))^2 \right), \]
and therefore
\[ \int_0^t \sum_{\ell=0}^{m-1} E \left( a(s, X(s)) - a(U_\ell) - V_\ell(s) \right)^2 \cdot 1_{[t_\ell, t_{\ell+1}]}(s) \, ds \]
\[ \leq c \cdot \left( \Delta_{\text{max}}^2 + \int_0^t E(X(s) - \hat{X}(s))^2 \, ds \right). \]

Note that
\[ V_\ell(t) = (a^{(0,1)} \cdot \sigma)(U_\ell) \cdot \int_{t_\ell}^t \left( 1 + \sigma^{(0,1)}(U_\ell) \cdot (W(u) - W(t_\ell)) \right) \, dW(u). \]

If \( s \in [t_\ell, t_{\ell+1}] \), too, with \( s \leq t \), then
\[ E \left( Z(s) \cdot Z(t) \right) = E \left( V_\ell(s) \cdot V_\ell(t) \right) \]
\[ = \int_{t_\ell}^t E \left( (a^{(0,1)} \cdot \sigma)(U_\ell) \cdot (1 + \sigma^{(0,1)}(U_\ell) \cdot (W(u) - W(t_\ell))) \right)^2 \, du \]
\[ \leq c \cdot E \left( \sigma^2(U_\ell) \right) \cdot (s - t_\ell) \leq c \cdot (s - t_\ell) \]
by Lemma 10. Otherwise \( E(Z(s) \cdot Z(t)) = 0 \). We conclude that
\[ \int_0^t \int_0^t E \left( Z(s) \cdot Z(u) \right) \, ds \, du \leq c \cdot \Delta_{\text{max}}^2. \]
Combining (22) and (23) we get

\[ E\left( A(t) - \tilde{A}(t) \right)^2 \leq 2 \cdot \left( E\left( A(t) - \int_0^t \sum_{\ell=0}^{m-1} \left( a(U_{\ell}) - V_{\ell}(s) \right) \cdot 1_{[t_{\ell},t_{\ell+1}]}(s) ds \right)^2 + E\left( \int_0^t Z(s) ds \right)^2 \right) \]

\[ \leq c \cdot \left( \Delta_{\text{max}}^2 + \int_0^t E\left( X(s) - \tilde{X}(s) \right)^2 ds \right). \]

For estimation of \( B - \tilde{B} \) we define

\[ R_\ell(t) = \sigma(t, X(t)) - \sigma(U_{\ell}) \cdot (1 + \sigma^{(0,1)}(U_{\ell}) \cdot (W(t) - W(t_{\ell}))). \]

Lemma 10 and (10) yield

\[ E(R_\ell(t))^2 \leq c \cdot (\Delta_{\ell}^2 + E(X(t) - \tilde{X}(t))^2) \]

for \( t \in [t_\ell, t_{\ell+1}] \). Hereby

\[ E\left( B(t) - \tilde{B}(t) \right)^2 \leq \int_0^t \sum_{\ell=0}^{m-1} E(R_\ell(s))^2 \cdot 1_{[t_\ell,t_{\ell+1}]}(s) ds \leq c \cdot \left( \Delta_{\text{max}}^2 + \int_0^t E(X(s) - \tilde{X}(s))^2 ds \right). \]

By means of (24) and (25) we conclude that

\[ E(X(t) - \tilde{X}(t))^2 \leq c \cdot \left( \Delta_{\text{max}}^2 + \int_0^t E(X(s) - \tilde{X}(s))^2 ds \right) \]

for all \( t \in T \). It remains to apply Gronwall’s Lemma.

References


E-mail address: hofmann@mi.uni-erlangen.de

E-mail address: gronbach@math.fu-berlin.de

E-mail address: klaus.ritter@fmi.uni-passau.de