

ON THE COMPLEXITY OF STOCHASTIC INTEGRATION

G. W. WASILKOWSKI AND H. WOŹNIAKOWSKI

ABSTRACT. We study the complexity of approximating stochastic integrals with error ε for various classes of functions. For Ito integration, we show that the complexity is of order ε^{-1} , even for classes of very smooth functions. The lower bound is obtained by showing that Ito integration is not easier than Lebesgue integration in the average case setting with the Wiener measure. The upper bound is obtained by the Milstein algorithm, which is almost optimal in the considered classes of functions. The Milstein algorithm uses the values of the Brownian motion and the integrand. It is bilinear in these values and is very easy to implement. For Stratonovich integration, we show that the complexity depends on the smoothness of the integrand and may be much smaller than the complexity of Ito integration.

1. INTRODUCTION

One of the most frequently studied subjects in computational complexity of continuous problems is the complexity of integration. Integration is understood as approximation of Lebesgue integrals. This problem is studied for various classes of functions in the worst case, average case and probabilistic settings, see, e.g., [7].

Probably, the main reason why so many complexity results have been established for Lebesgue integration is that this is a linear problem. There are a number of general results for linear problems which can be readily applied for Lebesgue integration. For instance, it is known that a linear algorithm is optimal and that adaptive information does not help, see, e.g., [7]. There are no such general results for nonlinear problems.

In this paper, we study complexity of *stochastic* integration for various classes of functions. We will be mostly dealing with Ito integrals, but will also remark on Stratonovich integrals. We restrict ourselves to scalar stochastic integration, and usually assume that the integrands depend only on the current values of the Brownian motion.

Stochastic integration is *not* a linear problem, since stochastic integrals depend nonlinearly on the Brownian motion. Furthermore, the natural setting for stochastic integration seems to be the worst case setting with respect to integrands and the average case setting with respect to Brownian motion. This corresponds to the

Received by the editor October 19, 1998 and, in revised form, April 6, 1999.

2000 *Mathematics Subject Classification.* Primary 60H05; Secondary 60H10.

Key words and phrases. Ito integrals, Stratonovich integrals, optimal algorithms, complexity.

The authors were partially supported by the National Science Foundation under Grants CCR-9729971 and CCR-9731858, respectively. This work was completed while the second author was a member of the Mathematical Sciences Research Institute at Berkeley in Fall 1998.

mixed setting, which is not usually studied for Lebesgue integration, see however [9].

Stochastic integration is intimately related to Brownian motion. That is why it is probably not surprising that Ito integration cannot be easier than Lebesgue integration in the average case setting for the class of continuous functions equipped with the classical Wiener measure. This holds for any class of integrands which contains a function which is linear in the first argument and quadratic in the second argument, see Theorem 1. The average case complexity for the Wiener measure is known, and is proportional to ε^{-1} if we want to guarantee an average error of at most ε . This shows that the complexity of Ito integration is at least of order ε^{-1} . This holds even if we consider very smooth integrands or even if we have the complete information about the integrand. For Lebesgue integration the complexity usually depends on smoothness and decreases with increasing smoothness.

The complexity bound ε^{-1} is sharp for classes of functions with uniformly bounded first derivatives with respect to the first argument, and uniformly bounded second derivatives with respect to the second argument. For such classes, the Milstein algorithm is almost optimal; it samples the Brownian motion at equally spaced points, and computes the integrand at sample points obtained through the Brownian motion. The algorithm is bilinear and easy to implement. It does not require any precomputation and its cost is proportional to ε^{-1} , see Theorem 2.

In this way, we obtain bounds on the complexity of Ito integration, see Theorem 3. These bounds are tight with respect to the error parameter ε^{-1} , the bounds on the corresponding derivatives, and on the length T of the interval in the stochastic integral. They are linear in all the parameters except T , for which the dependence is proportional to $T^{3/2}$. This result should be compared to a recent result on the complexity of approximating stochastic differential equations. As shown in [3], solving such equations is more difficult than approximating stochastic integrals, since their complexity is proportional to ε^{-2} .

For the class with only uniformly bounded first derivatives with respect to both arguments, we derive an algorithm that solves the problem with cost of order ε^{-2} . We do not know if this bound is sharp, although we believe this to be the case.¹ The bound of order ε^{-2} also holds for Ito integration for more general Lipschitz integrands that may depend on the whole Brownian motion.

We also briefly discuss Stratonovich integration. Unlike the Ito case, the complexity of Stratonovich integration is *not* related to the complexity of Lebesgue integration in the average case setting for the Wiener measure. We show that the complexity of Stratonovich integration *does* depend on smoothness of integrands, and may be much smaller than the complexity of Ito integration. This shows once more an essential difference between Ito and Stratonovich integrals, despite an apparent similarity in their definitions. Finally, we show how almost optimal algorithms for Ito and Stratonovich integration can be used for a nonlinear Lebesgue integration problem.

2. BASIC FACTS AND DEFINITIONS

In what follows, B will denote the *Brownian motion*, i.e., the random process distributed according to the classical Wiener measure. To stress this fact, we will write B_t instead of $B(t)$ to denote the value of B at time t . Recall that B is a

¹This has been recently proven by P. Hertling.

zero-mean Gaussian process which has continuous sample paths with covariance function

$$\mathbb{E}(B_t B_y) = \min\{t, y\} \quad \text{for} \quad t, y \in \mathbb{R}_+.$$

Here and elsewhere, \mathbb{E} denotes the expectation with respect to the Wiener measure.

For a given positive number T and a function

$$f : [0, T] \times \mathbb{R} \rightarrow \mathbb{R},$$

consider the problem of approximating the *Ito integral*

$$\mathcal{I}(f, B) = \int_0^T f(t, B_t) dB_t.$$

Since the formal definition of $\mathcal{I}(f, B)$ is quite long, we will only list some properties of \mathcal{I} . The reader unfamiliar with the definition of the Ito integral is referred to, e.g., [1, 4, 6]. Here we only stress that $\mathcal{I}(f, B) \in \mathbb{R}$ is a random variable since B is a random process. Furthermore, \mathcal{I} depends linearly on f and, in general, nonlinearly on B .

In the references mentioned above, one can also find conditions on f for the Ito integral to exist. Here we only mention that the integral exists for continuous functions f , and that in this paper we will impose stronger conditions on f .

For the readers familiar with stochastic differential equations, we add that the Ito integral $\mathcal{I}(f, B)$ is the solution of the system of stochastic differential equations at time T given by

$$(1) \quad \begin{aligned} dY_t &= dB_t, \\ dX_t &= f(t, Y_t) dB_t, \end{aligned}$$

with the initial condition $Y_0 = X_0 = 0$. Then Y is a Brownian motion and $X_T = \mathcal{I}(f, B)$. This obviously allows us to use known algorithms for stochastic differential equations, such as the Euler or Milstein algorithms, for approximating the Ito integrals.

We begin with the first important property of the Ito integral:

$$(2) \quad \int_a^b g(t) dB_t = g(t)B_t \Big|_a^b - \int_a^b g'(t)B_t dt \quad \text{if} \quad g \in C^1([a, b])$$

with the latter integral being the ordinary Lebesgue integral. We stress that (2) holds because the integrand g does not depend on B . In particular,

$$(3) \quad \int_a^b 1 dB_t = B_b - B_a.$$

We also have

$$(4) \quad \int_a^b B_t dB_t = \frac{1}{2} (B_b^2 - B_a^2) - \frac{1}{2}(b - a).$$

For twice continuously differentiable functions $g : \mathbb{R} \rightarrow \mathbb{R}$ we have

$$(5) \quad g(B_t) = g(B_\alpha) + \int_\alpha^t g'(B_s) dB_s + \frac{1}{2} \int_\alpha^t g''(B_s) ds,$$

which is a simplified version of the Ito formula, see, e.g., [6, Theorem 4.2].

The next formula, called *Ito isometry*, deals with the variance (with respect to B) of the Ito integral:

$$(6) \quad \mathbb{E} \left(\left(\int_a^b f(t, B_t) dB_t \right)^2 \right) = \int_a^b \mathbb{E} \left((f(t, B_t))^2 \right) dt.$$

As already mentioned, we want to approximate $\mathcal{I}(f, B)$. In this paper we assume that we can evaluate f and (depending on its regularity) the partial derivatives $f^{(k,j)}$ at points $(a_i, b_i) \in [0, T] \times \mathbb{R}$. We can also evaluate B at a finite number of $t_j \in \mathbb{R}$. We also allow for *adaptive* choice of evaluation points, i.e., t_i , a_i and b_i may depend on the computed values of B and/or f at the previous points. In particular, we can have $b_i = B_{a_i}$. For formal definitions and discussions on algorithms, we refer to [7]. Here we only state informally that any algorithm $\mathcal{A}(f, B)$ for approximating $\mathcal{I}(f, B)$ is of the form

$$\mathcal{A}(f, B) = \psi \left(f^{(i_1, j_1)}(a_1, b_1), \dots, f^{(i_k, j_k)}(a_k, b_k), B_{t_1}, \dots, B_{t_\ell} \right)$$

for some function ψ . Let $n = k + \ell$ denote the total number of evaluations of the function f and its derivatives, as well as the process B , used by \mathcal{A} . Let Ψ_n denote the class of all algorithms of this form.

We now define the worst case error of \mathcal{A} . Let \mathcal{F} be a class of functions f . Then the *worst case error* (with respect to f) of the algorithm \mathcal{A} is defined by

$$\text{error}(\mathcal{A}, \mathcal{F}) = \sup_{f \in \mathcal{F}} \sqrt{\mathbb{E}((\mathcal{I}(f, B) - \mathcal{A}(f, B))^2)}.$$

Given the total number n of evaluations of f and B , we want to know the n th minimal error

$$\text{error}(n, \mathcal{F}) = \inf_{\mathcal{A} \in \Psi_n} \text{error}(\mathcal{A}, \mathcal{F}).$$

We also want to obtain an optimal (or almost optimal) algorithm, i.e., an algorithm from Ψ_n whose error equals (or is close to) the n th minimal error.

In the next section we will prove that the n th minimal error is proportional to n^{-1} for a large family of input classes \mathcal{F} . We also derive an almost optimal algorithm that can be implemented at cost proportional to n .

In this way, we obtain the complexity of stochastic integration, $\text{comp}(\varepsilon, \mathcal{F})$, in the class \mathcal{F} , which is defined as the minimal cost of computing an approximation with error at most ε . See [7] for the precise definition. Specifically, we will prove that

$$\text{comp}(\varepsilon, \mathcal{F}) = \Theta(\varepsilon^{-1})$$

for all classes $\mathcal{F} = C^{r_1, r_2}$ with $r_1 \geq 1$ and $r_2 \geq 2$. Here C^{r_1, r_2} denotes the class of functions which are continuously differentiable r_1 times with respect to the first variable and r_2 times with respect to the second variable, and whose sup-norms of derivatives are uniformly bounded.

3. LOWER BOUNDS

In this section we provide a simple lower bound on the n th minimal error. This bound holds if \mathcal{F} contains a function which is linear in the first argument and quadratic in the second argument.

Theorem 1. *If \mathcal{F} contains a function $f(t, y) = a + bt + cy^2$ for some a, b, c with $b + c \neq 0$, then*

$$\text{error}(n, \mathcal{F}) \geq \frac{T^{3/2}|b + c|}{2\sqrt{3}(n + 3/2)}.$$

Proof. From (2) applied to $g(t) = a + bt$ we get

$$\mathcal{I}(g, B) = \int_0^T (a + bt) dB_t = (a + bT)B_T - b \int_0^T B_t dt.$$

Let $h(y) = cy^2$. Since $h(y) = (cy^3/3)'$, then (5) with $g(y) = cy^3/3$ yields

$$\mathcal{I}(h, B) = \frac{c}{3}B_T^3 - c \int_0^T B_t dt.$$

Since $f(t, y) = g(t) + h(y)$, we have

$$(7) \quad \mathcal{I}(f, B) = (a + bT)B_T + \frac{c}{3}B_T^3 - (b + c) \int_0^T B_t dt.$$

Since $b + c \neq 0$, this implies that approximating $\mathcal{I}(f, B)$ is equivalent to the problem of approximating scalar integrals in the average case setting with respect to the classical Wiener measure. Note that

$$\int_0^T B_t dt = T^{3/2} \int_0^1 W_t dt,$$

where $W_t = B_{Tt}/\sqrt{T}$ is also the Brownian motion. The problem of approximating $\int_0^1 W_t dt$ was studied in [5, 8]. For nonadaptive choice of evaluation points, the n th minimal average error is equal to $1/((2n + 1)\sqrt{3})$, see [5]. Adaptive choice of evaluation points may save at most one evaluation, see [8], and therefore the n th minimal average error is no less than $1/((2n + 3)\sqrt{3})$. This completes the proof. \square

Theorem 1 states that the n th minimal error cannot go to zero faster than n^{-1} as long as the function f with $b + c \neq 0$ belongs to the class \mathcal{F} . In terms of solving the system (1) of stochastic differential equations this means that the order of strong convergence is at most one. This result may be somehow surprising, since there are algorithms for solving (1) with the order of strong convergence greater than one as long as the coefficients of the system are sufficiently regular, which obviously holds for the function f . The algorithms with the order of strong convergence greater than one require, however, exact values of certain integrals with Brownian motion, see [4]. In our setting, this is not allowed, since we assumed that only values of Brownian motion can be obtained. In fact, under the additional assumption that integrals of Brownian motion are available, we can compute (7) exactly with two evaluations, and the proof breaks down.

4. OPTIMALITY OF THE MILSTEIN ALGORITHM

We now show that the Milstein algorithm is almost optimal for many classes \mathcal{F} consisting of functions that have continuous partial derivatives. The Milstein algorithm is used for the solution of general systems of stochastic differential equations, see [4]. For the system (1), the Milstein algorithm takes the following form:

$$(8) \quad \mathcal{A}_n(f, B) = \sum_{i=1}^n A_{i,n}(f, B),$$

where

$$A_{i,n}(f, B) = f(a_i, B_{a_i})(B_{a_{i+1}} - B_{a_i}) + \frac{1}{2} \frac{\partial f(a_i, y)}{\partial y} \Big|_{y=B_{a_i}} \left((B_{a_{i+1}} - B_{a_i})^2 - \frac{T}{n} \right)$$

with equally spaced points

$$a_i = \frac{i-1}{n} T, \quad i = 1, 2, \dots, n+1.$$

This algorithm uses n values of B , $b_i = B_{a_i}$ for $i = 2, 3, \dots, n+1$ (note that $b_1 = B_0 = 0$), n values of f , $f(a_i, b_i)$ for $i = 1, 2, \dots, n$, and n values of partial derivatives $\partial f(a_i, y)/\partial y$ with $y = b_i$ for $i = 1, 2, \dots, n$. Hence, the total number of function evaluations used by \mathcal{A}_n equals $3n$. That is, $\mathcal{A}_n \in \Psi_{3n}$. Moreover, the algorithm \mathcal{A}_n requires at most $8n+1$ arithmetic operations, and does not require precomputation of any number.

It is known that the order of strong convergence of the Milstein algorithm is one as long as the function f is sufficiently regular. This corresponds to the error bound of order n^{-1} for the Ito integrals. The regularity assumptions on f are presented in a number of papers. For instance, in [4] f is assumed, in particular, to be three times differentiable with respect to the second argument. In [2], only two times differentiability with respect to the second argument is assumed. In what follows we present a relatively short proof of the error bound with the explicit constants and with the minimal regularity assumptions on f . From this we conclude almost optimality of the Milstein algorithms with respect to the order of convergence as well as with respect to other parameters of the problem.

Theorem 2. *Let $f : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ satisfy the following conditions: $\partial f/\partial x$, $\partial f/\partial y$, and $\partial^2 f/\partial x^2$ are continuous,*

$$(9) \quad |f(t, y) - f(z, y)| \leq L|t - z|, \quad \forall t, z \in [0, T], \quad \forall y \in \mathbb{R},$$

and

$$(10) \quad \sup_{t \in [0, T], y \in \mathbb{R}} \left| \frac{\partial^2}{\partial y^2} f(t, y) \right| \leq K$$

for some constants K and L . Then

$$(11) \quad \mathbb{E}(\mathcal{I}(f, B) - \mathcal{A}_n(f, B))^2 \leq \frac{T^3}{n^2} \left(\frac{2}{3} L^2 + K^2 \right).$$

Proof. Due to (3) and (4), we have

$$\begin{aligned} \int_{a_i}^{a_{i+1}} \left(\int_{a_i}^t 1 dB_s \right) dB_t &= \int_{a_i}^{a_{i+1}} (B_t - B_{a_i}) dB_t \\ &= \frac{1}{2} ((B_{a_{i+1}} - B_{a_i})^2 - (a_{i+1} - a_i)). \end{aligned}$$

This and (3) imply that

$$A_{i,n}(f, B) = \int_{a_i}^{a_{i+1}} \left(f(a_i, B_{a_i}) + \frac{\partial f(a_i, y)}{\partial y} \Big|_{y=B_{a_i}} \int_{a_i}^t 1 dB_s \right) dB_t.$$

Consequently,

$$(12) \quad \mathcal{A}_n(f, B) = \int_0^T \hat{f}(t, B) dB_t$$

with

$$\widehat{f}(t, B) = \sum_{i=1}^n \chi_{(a_i, a_{i+1}]}(t) \left(f(a_i, B_{a_i}) + \frac{\partial f(a_i, y)}{\partial y} \Big|_{y=B_{a_i}} \int_{a_i}^t 1 dB_s \right),$$

where $\chi_{(a_i, a_{i+1}]}$ is the characteristic (indicator) function of $(a_i, a_{i+1}]$.

Note that $\widehat{f}(t, B)$ depends only on the values of B_x for $x \leq t$. For such functions, the Ito isometry (6) is still valid, and therefore we have

$$\begin{aligned} \mathbb{E}(\mathcal{I}(f, B) - \mathcal{A}_n(f, B))^2 &= \mathbb{E} \left(\int_0^T (f(t, B_t) - \widehat{f}(t, B)) dB_t \right)^2 \\ &= \int_0^T \mathbb{E} \left(f(t, B_t) - \widehat{f}(t, B) \right)^2 dt = \sum_{i=1}^n e_i^2 \end{aligned}$$

with

$$e_i^2 = \int_{a_i}^{a_{i+1}} \mathbb{E} \left(f(t, B_t) - \widehat{f}(t, B) \right)^2 dt.$$

Note that

$$e_i^2 \leq 2(e_{i,1}^2 + e_{i,2}^2)$$

with

$$\begin{aligned} e_{i,1}^2 &= \int_{a_i}^{a_{i+1}} \mathbb{E} \left((f(t, B_t) - f(a_i, B_t))^2 \right) dt, \\ e_{i,2}^2 &= \int_{a_i}^{a_{i+1}} \mathbb{E} \left((f(a_i, B_t) - \widehat{f}(t, B))^2 \right) dt. \end{aligned}$$

Due to (9), we get

$$(13) \quad e_{i,1}^2 \leq \int_{a_i}^{a_{i+1}} L^2(t - a_i)^2 dt = \frac{T^3 L^2}{3n^3}.$$

To estimate $e_{i,2}^2$, we use the Ito formula (5). Applying it to $g(y) = f(a_i, y)$, we get

$$f(a_i, B_t) = f(a_i, B_{a_i}) + \int_{a_i}^t \frac{\partial}{\partial y} f(a_i, y) \Big|_{y=B_s} dB_s + \frac{1}{2} \int_{a_i}^t \frac{\partial^2}{\partial y^2} f(a_i, y) \Big|_{y=B_s} ds.$$

Then

$$\begin{aligned} e_{i,2}^2 &= \int_{a_i}^{a_{i+1}} \mathbb{E} \left(\int_{a_i}^t \left(\frac{\partial f(a_i, y)}{\partial y} \Big|_{y=B_s} - \frac{\partial f(a_i, y)}{\partial y} \Big|_{y=B_{a_i}} \right) dB_s \right. \\ &\quad \left. + \frac{1}{2} \int_{a_i}^t \frac{\partial^2 f(a_i, y)}{\partial y^2} \Big|_{y=B_s} ds \right)^2 dt \\ &\leq 2 \int_{a_i}^{a_{i+1}} \mathbb{E} \left(\frac{1}{2} \int_{a_i}^t \frac{\partial^2}{\partial y^2} f(a_i, y) \Big|_{y=B_s} ds \right)^2 dt \\ &\quad + 2 \int_{a_i}^{a_{i+1}} \mathbb{E} \left(\int_{a_i}^t \left(\frac{\partial}{\partial y} f(a_i, y) \Big|_{y=B_s} - \frac{\partial}{\partial y} f(a_i, y) \Big|_{y=B_{a_i}} \right) dB_s \right)^2 dt. \end{aligned}$$

Note that

$$\mathbb{E} \left(\left(\int_{a_i}^t \frac{1}{2} \frac{\partial^2}{\partial y^2} f(a_i, y) \Big|_{y=B_s} ds \right)^2 \right) \leq \frac{1}{4} K^2 (t - a_i)^2$$

due to (10). Therefore

$$2 \int_{a_i}^{a_{i+1}} \mathbb{E} \left(\frac{1}{2} \int_{a_i}^t \frac{\partial^2}{\partial y^2} f(a_i, y) \Big|_{y=B_s} ds \right)^2 dt \leq \frac{1}{6} K^2 (a_{i+1} - a_i)^3 = \frac{T^3 K^2}{6n^3}.$$

Applying the Ito isometry (6), as well as (10), we get

$$\begin{aligned} & 2 \int_{a_i}^{a_{i+1}} \mathbb{E} \left(\int_{a_i}^t \left(\frac{\partial}{\partial y} f(a_i, y) \Big|_{y=B_s} - \frac{\partial}{\partial y} f(a_i, y) \Big|_{y=B_{a_i}} \right) dB_s \right)^2 dt \\ &= 2 \int_{a_i}^{a_{i+1}} \int_{a_i}^t \mathbb{E} \left(\frac{\partial}{\partial y} f(a_i, y) \Big|_{y=B_s} - \frac{\partial}{\partial y} f(a_i, y) \Big|_{y=B_{a_i}} \right)^2 ds dt \\ &\leq 2K^2 \int_{a_i}^{a_{i+1}} \int_{a_i}^t \mathbb{E}(B_s - B_{a_i})^2 ds dt \\ &= 2K^2 \int_{a_i}^{a_{i+1}} \int_{a_i}^t (s - a_i) ds dt = \frac{1}{3} K^2 (a_{i+1} - a_i)^3 = \frac{T^3 K^2}{3n^3}. \end{aligned}$$

Therefore $e_{i,2}^2 \leq T^3 K^2 / (2n^3)$. This and (13) yield

$$\sum_{i=1}^n e_i^2 \leq 2n \left(\frac{T^3 L^2}{3n^3} + \frac{T^3 K^2}{2n^3} \right) = \frac{T^3}{n^2} \left(\frac{2}{3} L^2 + K^2 \right),$$

which completes the proof. \square

Remark 1 (Only Function Values). The Milstein algorithm \mathcal{A}_n uses values of partial derivatives

$$D_i = \frac{\partial f(a_i, y)}{\partial y} \Big|_{y=B_{a_i}}, \quad i = 1, \dots, n.$$

With a rather insignificant increase in the error bound, one can modify \mathcal{A}_n so that the resulting algorithm does not require derivative values. Indeed, let

$$\tilde{\mathcal{A}}_n(f, B) = \sum_{i=1}^n \tilde{A}_{i,n}(f, B),$$

where $\tilde{A}_{i,n}$ differs from $A_{i,n}$ in that D_i is replaced by

$$\tilde{D}_{i,h} = \frac{f(a_i, B_{a_i} + h) - f(a_i, B_{a_i})}{h}$$

for some positive h , say $h \in (0, T/n]$. It is easy to see that with such a change the error bound will change to

$$\mathbb{E} \left(\mathcal{I}(f, B) - \tilde{\mathcal{A}}_n(f, B) \right)^2 \leq \frac{T^3}{n^2} \left(\frac{2}{3} L^2 + K^2 \right) + 2 \sum_{i=1}^n e_{i,3}^2$$

with

$$e_{i,3}^2 = \frac{1}{4} \mathbb{E} \left((D_i - \tilde{D}_{i,h}) \left((B_{a_{i+1}} - B_{a_i})^2 - \frac{T}{n} \right) \right)^2.$$

Due to (10), $|D_i - \tilde{D}_{i,h}| \leq Kh/2 \leq KT/(2n)$ and

$$e_{i,3}^2 \leq \frac{K^2 T^2}{4n^2} \mathbb{E} \left((B_{a_{i+1}} - B_{a_i})^2 - \frac{T}{n} \right)^2 = \frac{K^2 T^4}{2n^4}.$$

This yields

$$(14) \quad \mathbb{E} \left(\mathcal{I}(f, B) - \tilde{\mathcal{A}}_n(f, B) \right)^2 \leq \frac{T^3}{n^2} \left(\frac{2}{3} L^2 + K^2 + \frac{T}{2n} K^2 \right).$$

5. COMPLEXITY

From Theorems 1 and 2 we easily obtain bounds on the complexity of stochastic integration for the class $\mathcal{F} = \mathcal{F}_{L,K}$. In what follows, by $f^{i,j}$ we mean $\partial^{i+j} f / (\partial t^i \partial y^j)$, and $\|f^{i,j}\|$ denotes the sup-norm with respect to $t \in [0, T]$ and $y \in \mathbb{R}$. Let

$$\mathcal{F}_{L,K} = \{f : [0, T] \times \mathbb{R} \rightarrow \mathbb{R} : \|f^{1,0}\| \leq L, \|f^{0,2}\| \leq K\}$$

for some nonnegative L and K with $L^2 + K^2 > 0$. We will also assume that any arithmetic operation has unit cost, and that the cost of one evaluation of f , its partial derivative, and B is \mathbf{c} . Usually, $\mathbf{c} \gg 1$.

Theorem 3. *For the class $\mathcal{F}_{L,K}$ the n th minimal error, $n \geq 3$, satisfies*

$$\frac{T^{3/2}}{2\sqrt{3}(n+3/2)} \left(L + \frac{K}{2} \right) \leq \text{error}(n, \mathcal{F}) \leq \frac{3T^{3/2}}{n-2} \sqrt{\frac{2}{3} L^2 + K^2},$$

and the complexity of stochastic integration satisfies

$$\left(\frac{T^{3/2}}{2\sqrt{3}\varepsilon} \left(L + \frac{K}{2} \right) - \frac{3}{2} \right) \mathbf{c} \leq \text{comp}(\varepsilon, \mathcal{F}) \leq \left(3 + \frac{3T^{3/2}}{\varepsilon} \sqrt{\frac{2}{3} L^2 + K^2} \right) (\mathbf{c} + 8).$$

Moreover, the Milstein algorithm \mathcal{A}_n from Theorem 2 or its modification $\tilde{\mathcal{A}}_n$ from Remark 1 is almost optimal.

Proof. The upper bounds on the n th minimal error and complexity follow from Theorem 2 applied for the algorithm $\mathcal{A}_{\lfloor n/3 \rfloor}$, since then we use $3\lfloor n/3 \rfloor \leq n$ evaluations of the function. (We also estimated $1/\lfloor n/3 \rfloor$ by $3/(n-2)$.) The lower bound on the n th minimal error is obtained from Theorem 1 applied to $f(t, y) = Lt + y^2 K/2$, i.e., for $a = 0$, $b = L$ and $c = K/2$. The lower bound on the complexity easily follows from the lower bound on the n th minimal error. \square

The essence of Theorem 3 is that the complexity depends linearly on ε^{-1} , L and K . The dependence on the length of the interval T of stochastic integration is a little more significant, since the complexity is proportional to $T^{3/2}$.

6. FINAL COMMENTS

We end this paper with a number of remarks concerning other classes of functions and Stratonovich stochastic integration.

Remark 2 (Smooth Classes). We stress that the complexity of stochastic integration as a function of ε^{-1} does not decrease for classes of smooth integrands. Indeed, let

$$\mathcal{F}(r_1, r_2) = \{f : [0, T] \times \mathbb{R} : \|f^{i,j}\| \leq 1, i = 0, 1, \dots, r_1, j = 0, 1, \dots, r_2\}$$

for some integers $r_1 \geq 1$ and $r_2 \geq 2$, and $T \geq 1$. Then the complexity is proportional to ε^{-1} even for large r_1 and r_2 . Indeed, the function $f(t, y) = t/T$ belongs to $\mathcal{F}(r_1, r_2)$ for all considered values of r_1 and r_2 . Theorem 1 then yields a lower bound of order ε^{-1} on the complexity. The algorithm \mathcal{A}_n works for the class $\mathcal{F}(r_1, r_2)$, since $\mathcal{F}(r_1, r_2) \subset \mathcal{F}(1, 2)$. This yields an upper bound on the complexity of order ε^{-1} .

Remark 3 (Lipschitz Class). We have so far assumed that integrands f are twice continuously differentiable with respect to the second argument. It is also of interest to consider the class of once differentiable functions. In particular, let us consider the Lipschitz class of functions with respect to two arguments,

$$\mathcal{F}_{\text{Lip}} = \{f : |f(t, y) - f(z, y)| \leq L|t - z|, |f(t, y) - f(t, w)| \leq K|y - w|, \\ \forall t, z \in [0, T], y, w \in \mathbb{R}\}.$$

Then the Milstein algorithm \mathcal{A}_n is not well defined, since the partial derivatives of f with respect to y may not exist. We may modify this algorithm just by dropping the terms with the partial derivatives. Then we obtain the Euler algorithm, see [4],

$$\mathcal{A}_n(f, B) = \sum_{i=1}^n f(a_i, B_{a_i})(B_{a_{i+1}} - B_{a_i})$$

with the same sample points $a_i = (i-1)T/n$.

It is known that $\text{error}(n, \mathcal{F}_{\text{Lip}}) = O(n^{-1/2})$. More precisely, one can show along the lines of the proof of Theorem 2 that

$$\text{error}(\mathcal{A}_n, \mathcal{F}_{\text{Lip}}) \leq \left(\frac{2L^2T^3}{3n^2} + \frac{K^2T^2}{n} \right)^{1/2}.$$

This yields an upper bound of order ε^{-2} on the complexity. We do not know if this bound is sharp, although we conjecture that this is the case.² Observe that Theorem 1 gives only a lower bound of order n^{-1} on the n th minimal error, which does not match the error of \mathcal{A}_n .

We also remark that the error of the algorithm \mathcal{A}_n is of a similar form for the following Hölder class of integrands $\mathcal{F}^{(\alpha, \beta)}$:

$$\mathcal{F}^{(\alpha, \beta)} = \{f : |f(t, y) - f(z, y)| \leq L|t - z|^\alpha, |f(t, y) - f(t, w)| \leq K|y - w|^\beta, \\ \forall t, z \in [0, T], y, w \in \mathbb{R}\}$$

for $\alpha, \beta \in (0, 1]$. Then

$$\text{error}(\mathcal{A}_n, \mathcal{F}^{(\alpha, \beta)}) \leq \sqrt{2} \left(\frac{L^2T^{2\alpha+1}}{(2\alpha+1)n^{2\alpha}} + \frac{K^2T^{\beta+1}c_\beta}{(\beta+1)n^\beta} \right)^{1/2}$$

with $c_\beta = (2\pi)^{-1/2} \int_{-\infty}^{\infty} |t|^{2\beta} e^{-t^2/2} dt = \Gamma(\beta + 1/2) 2^\beta / \sqrt{\pi}$.

Remark 4 (Generalized Class). We have considered stochastic Ito integration for integrands of the form $f(t, B_t)$. That is, the only dependence on the Brownian motion is through its value at the point t . Obviously, it is also of interest to consider more general forms of integrands. For instance, consider the functions

²As already mentioned, this conjecture has been established by P. Hertling.

of the form $f(t, B)$ which may depend on the whole Brownian motion. The Ito integral

$$\mathcal{I}(f, B) = \int_0^T f(t, B) dB_t$$

is well defined only if the dependence on B in $f(t, B)$ is restricted to $B|_{[0, t]}$ as, for instance, $f(t, B) = g\left(t, \int_0^t B_s ds\right)$ for some function g . Assuming this property, we consider the class

$$\mathcal{F} = \{f : |f(t, B) - f(z, B)| \leq L \|B\| |t - z|^\alpha + K |B_{g(t)} - B_{g(z)}|\}$$

for some nonnegative L, K, α , and a function $g : [0, T] \rightarrow [0, T]$ which satisfies the condition $|g(t) - g(z)| \leq M |t - z|^\beta$ for some nonnegative M and β . Here $\|B\| = \max_{t \in [0, T]} |B_t|$.

For example, observe that $f(t, B) = B(t/a)$ with $a \geq 1$ belongs to the class \mathcal{F} with arbitrary L, α and $K = 1$, $g(t) = t/a$ for which $\beta = 1$ and $M = 1/a$. On the other hand, the function $f(t, B) = \int_0^t B_s ds$ belongs to the class \mathcal{F} with $L = \alpha = 1$ and arbitrary K and g .

We assume that we can compute the values of $f(t, B)$ and again consider the Euler algorithm

$$\mathcal{A}_n(f, B) = \sum_{i=1}^n f(a_i, B)(B_{a_{i+1}} - B_{a_i}).$$

It is easy to check that the upper bound on the error of \mathcal{A}_n satisfies

$$\text{error}(n, \mathcal{F}) \leq \left(\frac{2L^2 \mathbb{E}(\|B\|^2) T^{2\alpha+1}}{(2\alpha+1)n^{2\alpha}} + \frac{2K^2 M^2 T^{\beta+1}}{\beta+1} \frac{T^{\beta+1}}{n^\beta} \right)^{1/2} = O\left(n^{-\min(\alpha, \beta/2)}\right).$$

This yields an upper bound on the complexity of order $\varepsilon^{-1/\min(\alpha, \beta/2)}$. It would be interesting to know whether this bound is sharp.

Remark 5 (Stratonovich Stochastic Integration). The Stratonovich stochastic integral is defined differently than the Ito integral. Roughly speaking, see [4], both are defined as the mean-square limit of the sums

$$\sum_{j=1}^n ((1-\lambda)f(t_{j,n}, B) + \lambda f(t_{j+1,n}, B)) (B_{t_{j+1,n}} - B_{t_{j,n}}),$$

where the evaluation points $t_{j,n}$ are ordered, $t_{1,n} = 0 < t_{2,n} < \dots < t_{n+1,n} = T$, and $\max_j (t_{j+1,n} - t_{j,n})$ goes to zero as n approaches infinity. Here $\lambda \in [0, 1]$. For $\lambda = 0$ we obtain the Ito integral, whereas for $\lambda = 1/2$ we obtain the Stratonovich integral, which is denoted by

$$\mathcal{I}_S(f, B) = \int_0^T f(t, B) \circ dB_t.$$

Despite similar definitions, the complexity of Stratonovich integration may be quite different from the complexity of Ito integration. To illustrate this difference we consider the simplified case in which the function $f(t, B) = h(B_t)$ depends only on the value of the Brownian motion at t , and the function h belongs to the class

$$\mathcal{F}_{r,\alpha} = \{h : \mathbb{R} \rightarrow \mathbb{R} \mid h^{(r)} \text{ is continuous and } \sup_{x \in \mathbb{R}} |h^{(r)}(x)(1 + |x|)^\alpha| \leq 1\}$$

for an integer r and $\alpha \leq 0$.

For the Stratonovich integral, see [4, p. 101], we have

$$\int_0^T h(B_t) \circ dB_t = \int_0^{B_T} h(t) dt.$$

Hence, the integral depends only on the value of B_T . The approximation of such Stratonovich integrals corresponds to approximation of the function h and then to integration over the interval $[0, B_T]$.

It is proven in [10] that there exist sample points x_i and piecewise polynomial functions a_i such that the algorithm $M_n(h, x) = \sum_{i=1}^n h(x_i) a_i(x)$ approximates h with error

$$|h(x) - M_n(h, x)| \leq C n^{-r} \exp(|x|), \quad \forall x \in \mathbb{R}, \quad \forall h \in \mathcal{F}_{r,\alpha}, \quad \forall n,$$

where C is a positive constant. Define the algorithm

$$(15) \quad \mathcal{A}_n(h, B) = \sum_{i=1}^n h(x_i) \int_0^{B_T} a_i(x) dx.$$

This algorithm uses n function values of h and one function value of B . Furthermore, it requires only $O(n)$ arithmetic operations. It is easy to check that

$$\begin{aligned} \text{error}_S(\mathcal{A}_n, \mathcal{F}_{r,\alpha}) &:= \sup_{h \in \mathcal{F}_r} \sqrt{\mathbb{E}(\mathcal{I}_S(h, B) - \mathcal{A}_n(h, B))^2} \\ &= O(|B_T| \exp(|B_T|) n^{-r}) = O\left(n^{-r} \int_0^\infty t e^{(t-t^2/(2T))} dt\right) = O(n^{-r}). \end{aligned}$$

This yields an upper bound $O(\varepsilon^{-1/r})$ on the complexity $\text{comp}_S(\varepsilon, \mathcal{F}_{r,\alpha})$ of Stratonovich integration. In fact, this bound is sharp. This follows from the fact that the complexity of approximation of the functions h from the class $\mathcal{F}_{r,\alpha}$ is of order $\varepsilon^{-1/r}$, see [10]. Furthermore, integration over $[0, B_T]$ is not easier than integration over the interval $[0, a]$ if $|B_T| \geq a > 0$ holds. The last inequality holds with a positive probability. Hence,

$$\text{comp}_S(\varepsilon, \mathcal{F}_{r,\alpha}) = \Theta(\varepsilon^{-1/r}).$$

This proves that the complexity of Stratonovich integration does depend on the smoothness, and decreases with increasing smoothness. Since the complexity of Ito integration does not depend on smoothness, this shows that the complexities of Stratonovich and Ito integration may be quite different.

It is easy to find the complexity of Ito integration for the class $\mathcal{F}_{r,0}$. For $r = 2$ the complexity of Ito integration is of order ε^{-1} , since the function $h(y) = y^2/2$ belongs to $\mathcal{F}_{r,0}$, and the algorithm \mathcal{A}_n given by (8) can be applied. For $r > 2$, the complexity of Ito integration is infinite, since the functions $h(y) = cy^2$ for arbitrary c belong to $\mathcal{F}_{r,0}$ and the error of any algorithm must be infinite due to Theorem 1.

Of course, the last negative result can be avoided if we restrict the class to $\mathcal{F}_{r,\alpha} \cap \mathcal{F}_{2,0}$. Then the second derivatives are bounded, and the complexity of Ito integration is of order ε^{-1} independently of the smoothness parameter r . This restriction of the class does not lower the complexity of Stratonovich integration, and it remains of order $\varepsilon^{-1/r}$.

Remark 6 (Application of Ito and Stratonovich Integration). We now show how algorithms for Ito and Stratonovich integration can be used for nonlinear Lebesgue integration in the average case setting.

Let $r \geq 2$. For a function $h \in \mathcal{F}_{r,\alpha} \cap \mathcal{F}_{2,0}$ we want to approximate the Lebesgue integral

$$\mathcal{I}_h(B) = \int_0^T h'(B_t) dt.$$

We assume that we can compute the values of h, h' and B . The error of an algorithm is defined as before, i.e., the worst case with respect to h from $h \in \mathcal{F}_{r,\alpha} \cap \mathcal{F}_{2,0}$ and the average case with respect to B . Observe that if h' is not linear, then we have nonlinear dependence on B . That is why this is a nonlinear integration problem. We are interested in the complexity $\text{comp}^{\text{non}}(\varepsilon)$ of this nonlinear problem.

This problem can be solved by using the relation between $\mathcal{I}_h(B)$, Ito and Stratonovich integrals. From [4, p. 101] we have

$$\mathcal{I}_h(B) = 2\mathcal{I}_S(h, B) - 2\mathcal{I}(h, B).$$

Since we know the complexity and almost optimal algorithms for Ito and Stratonovich integration, we can easily see that

$$\text{comp}^{\text{non}}(\varepsilon) = \Theta(\varepsilon^{-1}).$$

Furthermore to achieve the cost bound proportional to ε^{-1} it is enough to use the algorithm (15) for Stratonovich integration and the algorithm (8) for Ito integration.

The same problem can be also studied for the class $\mathcal{F}_{1,0}$. Then the same relation with Ito and Stratonovich integrals holds, and we may apply algorithms discussed before to get an upper bound on the complexity of the nonlinear problem of order ε^{-2} . We do not know whether this bound is sharp.

ACKNOWLEDGMENTS

We are grateful to P. Hertling, N. Hofmann, T. Müller-Gronbach, E. Novak, K. Ritter, J. F. Traub and A. G. Werschulz for valuable comments on our paper. We are also thankful for constructive comments from the referee.

REFERENCES

- [1] K. L. Chung and R. J. Williams, *Introduction to Stochastic Integration*, (2nd ed.), Birkhäuser, Boston, 1990. MR **92d**:60057
- [2] O. Faure, Simulation du mouvement brownien et des diffusions, *These ENPC*, Paris, 1992.
- [3] N. Hofmann, T. Müller-Gronbach, and K. Ritter, Optimal approximation of stochastic differential equations by adaptive step-size control, *Math. Comp.*, Posted on May 20, 1999, PII S 0025-5718(99)01177-1 (to appear in print).
- [4] P. E. Kloeden and E. Platen, *Numerical Solutions of Stochastic Differential Equations*, Springer-Verlag, Berlin, 1992. MR **94b**:60069
- [5] D. Lee, Approximation of linear operators on a Wiener space, *Rocky Mount. J. Math.* **16** (1986), 641-659. MR **88f**:41053
- [6] B. Øksendal, *Stochastic Differential Equations*, (3rd ed.), Springer-Verlag, New York, 1992. MR **94g**:60112
- [7] J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski, *Information-Based Complexity*, Academic Press, New York, 1988. MR **90f**:68085
- [8] G. W. Wasilkowski, Information of varying cardinality, *J. Complexity*, **2** (1986), 204-228. MR **88m**:65099
- [9] G. W. Wasilkowski and H. Woźniakowski, Mixed settings for linear problems, *J. of Complexity*, **5** (1989), 457-465. MR **91b**:65065
- [10] G. W. Wasilkowski and H. Woźniakowski, Weighted approximation over \mathbb{R} , to appear in *J. Approx. Theory*.

DEPARTMENT OF COMPUTER SCIENCE, UNIVERSITY OF KENTUCKY, LEXINGTON, KY 40506,
USA

E-mail address: `greg@cs.uky.edu`

DEPARTMENT OF COMPUTER SCIENCE, COLUMBIA UNIVERSITY, NEW YORK, NY 10027, USA,
AND INSTITUTE OF APPLIED MATHEMATICS AND MECHANICS, UNIVERSITY OF WARSAW, UL. BA-
NACHA 2, 02-097 WARSZAWA, POLAND

E-mail address: `henryk@cs.columbia.edu`