JOINT CHARACTERISTIC FUNCTION AND SIMULTANEOUS SIMULATION OF ITERATED ITÔ INTEGRALS FOR MULTIPLE INDEPENDENT BROWNIAN MOTIONS

BY MAGNUS WIKTORSSON

Lund University

We consider all two-times iterated Itô integrals obtained by pairing m independent standard Brownian motions. First we calculate the conditional joint characteristic function of these integrals, given the Brownian increments over the integration interval, and show that it has a form entirely similar to what is obtained in the univariate case. Then we propose an algorithm for the simultaneous simulation of the m^2 integrals conditioned on the Brownian increments that achieves a mean square error of order $1/n^2$, where n is the number of terms in a truncated sum. The algorithm is based on approximation of the tail-sum distribution, which is a multivariate normal variance mixture, by a multivariate normal distribution.

1. Introduction. Consider the multidimensional stochastic differential equation (SDE),

$$(1.1) \qquad dX(t) = b(X(t), t)\,dt + \sigma(X(t), t)\,d\mathbf{W}(t),$$

where $X(t)$ is a d-dimensional vector and $\mathbf{W}(t)$ is an m-dimensional vector of independent standard Brownian motions. The functions $b(X(t), t)$ and $\sigma(X(t), t)$ are measurable mappings from $\mathbf{R}^d \times \mathbf{R}$ to $\mathbf{R}^d$ and from $\mathbf{R}^d \times \mathbf{R}$ to $\mathbf{R}^d \times \mathbf{R}^m$, respectively. The above equation is here interpreted in the Itô sense.

Explicit solutions to (1.1) can only be found in a few special cases, so that in general we are confined to computing numerical approximations. Consider a sequence $\{\widetilde{X}^h(t), 0 \le t \le T\}$, for $h > 0$, of numerical approximations of a (strong) solution $\{X(t), 0 \le t \le T\}$, where $\widetilde{X}^h(\cdot)$ is defined for $t = 0, h, 2h, \ldots, T$, where h is called the step size. This sequence is said to converge at rate $O(h^\gamma)$ if

$$E|X(T) - \widetilde{X}^h(T)| = O(h^\gamma) \quad \text{as } h \to 0.$$

For example, the simplest scheme, the Euler one, converges at rate $\gamma = 1/2$.

The dispersion matrix $\sigma(x, t)$ of the SDE (1.1) is said to satisfy the so-called commutativity condition if

$$(1.2) \qquad L^i \sigma_{kj} = L^j \sigma_{ki}, \qquad i, j = 1, \ldots, m, k = 1, \ldots, d,$$

where the differential operator L^i is given by

$$(1.3) \qquad L^i = \sum_{l=1}^{d} \sigma_{li}(x, t) \frac{\partial}{\partial x_l}.$$

In the general case where $\sigma(x, t)$ does not satisfy (1.2), it is not possible to generate numerical approximations converging faster than $O(h^{1/2})$ unless the iterated Itô integrals,

$$(1.4) \qquad I_{ij}(t_n, t_n + h) = \int_{t_n}^{t_n+h} \int_{t_n}^{s} dW_i(u)\, dW_j(s)$$

are included in the numerical scheme [see, e.g., Rümelin (1982)]. Here $\{t_n\}$ are the time points used in the discretisation.

Milshtein (1974) proposed a numerical scheme that converges strongly at rate $O(h)$ if $b \in C^{1,1}(\mathbf{R}^d \times \mathbf{R}^+)$ and $\sigma \in C^{2,1}(\mathbf{R}^d \times \mathbf{R}^+)$. In this scheme the kth component of the approximation is given by

$$\widetilde{X}_k^h(t_n+h) = \widetilde{X}_k^h(t_n) + b_k h + \sum_{i=1}^{m} \sigma_{ki} \Delta W_i(t_n, t_n+h) + \sum_{i=1}^{m} \sum_{j=1}^{m} L^i \sigma_{kj} I_{ij}(t_n, t_n+h),$$

$$\widetilde{X}^h(t_0) = X(t_0),$$

where $\Delta W_i(t_n, t_n + h) = W_i(t_n + h) - W_i(t_n)$.

The purpose of this paper is twofold. First, we derive the conditional joint characteristic function of the iterated Itô integrals given the Brownian increments, and second we propose an algorithm for the simultaneous simulation of the iterated Itô integrals and the Brownian increments. Before proceeding we note that the joint distribution of $\Delta W_i(t, t+h) = W_i(t+h) - W_i(t)$, $\Delta W_j(t, t+h) = W_j(t+h) - W_j(t)$ and $I_{ij}(t, t+h)$ for $i, j = 1, \ldots, m$ does not depend on t, and hereafter we write $\Delta W_i(h)$ for $\Delta W_i(t, t+h)$ and $I_{ij}(h)$ for $I_{ij}(t, t+h)$.

In the case $m = 2$ the conditional characteristic function of $I_{12}(h)$ given $\Delta W_1(h)$ and $\Delta W_2(h)$ is given by

$$(1.5) \qquad \begin{aligned} &\varphi_{I_{12}(h)|\Delta W_1(h),\, \Delta W_2(h)}(t) \\ &\quad = \frac{th/2}{\sinh(th/2)} \exp\left\{ -\frac{\rho^2}{2}((th/2)\coth(th/2) - 1) + \iota tha \right\}, \end{aligned}$$

where $\rho^2 = (\Delta W_1(h)^2 + \Delta W_2(h)^2)/h$, $a = \Delta W_1(h)\Delta W_2(h)/2h$ and ι is the imaginary unit. This expression was derived by Lévy (1951) [see also Talacko (1956), Lévy (1965), pages 329–333]. In Section 3 we show that a similar expression holds true in the multidimensional case.

There is no simple way to simultaneously simulate the iterated Itô integrals and Brownian increments exactly. In the case $m = 2$, Gaines and Lyons (1994) proposed an algorithm for exact simulation of ρ^2 (see above) and the single iterated Itô integral $I_{12}(h)$, based on Marsaglia's "rectangle-wedge-tail" method. This method is complicated to implement, however, and occasionally requires

numerical inversion of the joint characteristic function of ρ^2 and $I_{12}(h)$. In higher dimensions, Kloeden, Platen and Wright (1992) suggested a simulation algorithm essentially based on truncation of an infinite series representation of the iterated Itô integrals. In order to accomplish a convergence rate of h for a numerical scheme approximating an SDE, the mean square error (MSE) in the approximation of the iterated Itô integrals must be negligible compared to the discretisation error of the numerical scheme. More precisely, an MSE of Ch^3 for some positive constant C is required [Kloeden and Platen (1995), Corollary 10.6.5]. Hence it is important to have an algorithm that simulates the iterated Itô integrals with small MSE in short time. The algorithm of Kloeden, Platen and Wright (1992) has an MSE of order h^2/n, where n is the number of terms in the truncated sum. In Section 4 we show that a slight modification of this algorithm yields an MSE of order h^2/n^2. Hence, with this improved convergence rate n needs to be proportional to $h^{-1/2}$ rather than h^{-1}, resulting in a considerable speed-up of the simulation. This can also be viewed in a different way. Assume that we want to simulate an SDE with a mean error $E|X^h(T) - X(T)| \leq \varepsilon$; how much work is required to accomplish this? If we measure work by the number of Gaussian random variables that needs to be simulated, we obtain $\mathbf{W}_{\mathrm{KPW}}(\varepsilon) \sim \varepsilon^{-2}$ for Milshtein combined with the Kloeden, Platen and Wright algorithm and $\mathbf{W}_{\mathrm{NA}}(\varepsilon) \sim \varepsilon^{-3/2}$ for Milshtein combined with our new algorithm. The notation $\mathbf{W}_{\mathrm{M}}(\varepsilon) \sim \varepsilon^{-\alpha}$ means that as $\varepsilon \to 0$ the number of Gaussian variables needed to achieve the accuracy ε for the method M is $O(\varepsilon^{-\alpha})$. If we compare this with the Euler method which has $\mathbf{W}_{\mathrm{EULER}}(\varepsilon) \sim \varepsilon^{-2}$, it is evident that there is no gain in using Milshtein combined with the Kloeden, Platen and Wright method since it requires no less (in practice even more) work than the Euler method to obtain the same accuracy. The Euler method is also easier to implement and faster to execute provided that the evaluations of the drift and dispersion functions are not too time-consuming compared to generation of the Gaussian random variables. This clearly shows why it is crucial to have a convergence rate faster than h^2/n in the approximation of the iterated Itô integrals.

Before closing this section we give some general notation used throughout the paper. The matrix I_n is an $n \times n$ identity matrix, $0_{n \times m}$ is an $n \times m$ matrix of zeros and 0_n is a column-vector of n zeros. Furthermore A^T will denote the transpose of A. The imaginary unit $(\sqrt{-1})$ will be denoted by ι.

2. Representation of the iterated Itô integrals. The iterated Itô integrals are closely related to the so-called Lévy stochastic area integrals, denoted by $A_{ij}(h)$ and defined by

$$A_{ij}(h) = \frac{I_{ij}(h) - I_{ji}(h)}{2}$$

for $i, j = 1, \ldots, m$. These integrals have a nice geometric interpretation; $A_{ij}(h)$ would, if the Brownian motion had finite variation, equal the signed area enclosed by the two-dimensional Brownian motion $\{\mathbf{W}(t)\} = \{(W_i(t), W_j(t))\}$ from 0 to
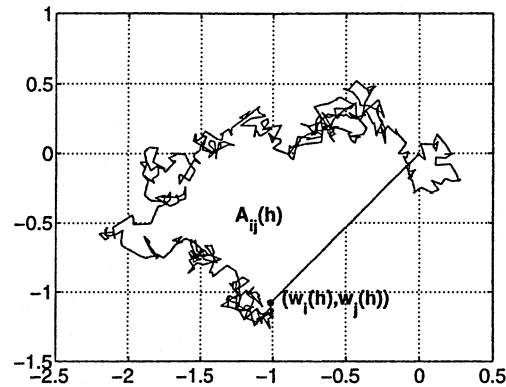
FIG. 1.   *Illustration of Lévy's stochastic area integral.*

h and the chord connecting $\mathbf{W}(h)$ and $\mathbf{W}(0) = (0, 0)$ (see Figure 1). We can think of it as a stochastic generalization of area.

We now state some useful relations between $I_{ij}(h)$, $A_{ij}(h)$, $\Delta W_i(h)$ and $\Delta W_j(h)$ for $i \neq j$:

$$I_{ij}(h) + I_{ji}(h) = \Delta W_i(h)\Delta W_j(h) \quad \text{a.s.,}$$

$$I_{ij}(h) = \frac{\Delta W_i(h)\Delta W_j(h)}{2} + A_{ij}(h) \quad \text{a.s.,}$$

(2.1) $$A_{ji}(h) = -A_{ij}(h),$$

$$I_{ii}(h) = \frac{\Delta W_i(h)^2 - h}{2} \quad \text{a.s.,}$$

$$A_{ii}(h) = 0.$$

Kloeden, Platen and Wright (1992) gave the following simultaneous representation of $I_{ij}(h)$, $\Delta W_i(h)$ and $\Delta W_j(h)$ for $i, j = 1, \ldots, m$:

$$I_{ij}(h) = \frac{\Delta W_i(h)\Delta W_j(h) - h\delta_{ij}}{2} + A_{ij}(h),$$

$$A_{ij}(h) = \frac{h}{2\pi} \sum_{k=1}^{\infty} \frac{1}{k} \left\{ X_{ik}\left(Y_{jk} + \sqrt{\frac{2}{h}}\Delta W_j(h) \right) \right.$$

$$\left. - X_{jk}\left(Y_{ik} + \sqrt{\frac{2}{h}}\Delta W_i(h) \right) \right\},$$

where $\Delta W_i(h) \sim N(0, h)$, $X_{ik} \sim N(0, 1)$ and $Y_{ik} \sim N(0, 1)$, $i = 1, \ldots, m$, $k = 1, 2, \ldots$ are all independent.

If we let $\mathbf{I}(h)$ and $\mathbf{A}(h)$ be the matrices where element i, j equals $I_{ij}(h)$ and $A_{ij}(h)$, respectively, we can rewrite this representation in matrix form as

$$\mathbf{I}(h) = \frac{\Delta \mathbf{W}(h) \Delta \mathbf{W}(h)^T - h I_m}{2} + \mathbf{A}(h),$$

(2.2)

$$\mathbf{A}(h) = \frac{h}{2\pi} \sum_{k=1}^{\infty} \frac{1}{k} \left\{ \mathbf{X}_k (\mathbf{Y}_k + \sqrt{2/h} \Delta \mathbf{W}(h))^T - (\mathbf{Y}_k + \sqrt{2/h} \Delta \mathbf{W}(h)) \mathbf{X}_k^T \right\},$$

where now $\Delta \mathbf{W}(h) \sim N(0, h I_m)$, $\mathbf{X}_k \sim N(0_m, I_m)$ and $\mathbf{Y}_k \sim N(0_m, I_m)$, $k = 1, 2, \ldots$ are all independent. Indeed, $\Delta \mathbf{W}(h) = (\Delta W_1(h), \ldots, \Delta W_m(h))^T$, $\mathbf{X}_k = (X_{1k}, \ldots, X_{mk})^T$ and $\mathbf{Y}_k = (Y_{1k}, \ldots, Y_{mk})^T$.

3. Conditional joint characteristic function of the stochastic area integrals. Recall that m is the number of independent Brownian motions. Let θ_{ij} be the variable which corresponds to the random variable $A_{ij}(h)$ in the joint characteristic function. From (2.1) it follows that we only need to calculate the characteristic function of the random variables $A_{ij}(h), i < j$, since the other stochastic area integrals depend in a deterministic way on these. We denote this set of $A_{ij}(h)$'s by $\widetilde{A}(h)$. Now define $\tilde{\theta}$ as the $m \times m$ upper triangular matrix, with zeros on the diagonal, given by

$$\tilde{\theta}_{ij} = \begin{cases} \theta_{ij}, & \text{for } i < j, \\ 0, & \text{otherwise.} \end{cases}$$

Let $\varphi_{\widetilde{A}(h)|\Delta \mathbf{W}(h)}(\tilde{\theta})$ be the conditional joint characteristic function of $\widetilde{A}(h)$ given $\Delta \mathbf{W}(h)$.

THEOREM 3.1. *The conditional joint characteristic function of $\widetilde{A}(h)$ given $\Delta \mathbf{W}(h)$ can be written as*

$$\varphi_{\widetilde{A}(h)|\Delta \mathbf{W}(h)}(\tilde{\theta})$$

(3.1)

$$= \det\{\operatorname{sinch}(\Theta(\tilde{\theta}))\}^{-1/2}$$

$$\times \exp\left(-\operatorname{tr}\left(\frac{\Delta \mathbf{W}(h) \Delta \mathbf{W}(h)^T}{2h} \{\cosh(\Theta(\tilde{\theta})) \operatorname{sinch}(\Theta(\tilde{\theta}))^{-1} - I_m\}\right)\right),$$

where $\operatorname{sinch}(x) = \sinh(x)/x$, $\Theta(\tilde{\theta}) = \{-h(\tilde{\theta} - \tilde{\theta}^T)^2/2\}^{1/2}$ and where the hyperbolic functions and the square root should be interpreted in the matrix sense.

Before giving the proof we note that the characteristic function $\widetilde{A}(h)$ has a form that is similar to what is obtained for $m = 2$ [cf. (1.5)]. The first factor does not depend on $\Delta \mathbf{W}(h)$ and it follows by taking $\Delta \mathbf{W}(h) = 0_m$ that it is itself a characteristic function. In the univariate case $m = 2$ it is the characteristic function of a logistic random variable and so its density is known; for $m > 2$ all marginals are of course still logistic, but the joint distribution involves dependencies and we have not been able to find a closed form for the joint density. The second factor is also a characteristic function itself when $m = 2$

but its density admits no simple closed form expression. When $m > 2$ we do not even know if this factor is a characteristic function. The factor $\exp(\iota t h a)$ in (1.5) comes from the second relation in (2.1) and so does not appear in $\varphi_{\widetilde{A}(h)|\Delta\mathbf{W}(h)}$.

PROOF. First note that

$$\varphi_{\widetilde{A}(h)|\Delta\mathbf{W}(h)}(\tilde{\theta}) = E\left[\exp\left(\iota\sum_{i<j}\theta_{ij}A_{ij}(h)\Big|\Delta\mathbf{W}(h)\right)\right]$$

$$= E[\exp\{\iota\ \mathrm{tr}(\tilde{\theta}^T\mathbf{A}(h))\}\mid \Delta\mathbf{W}(h)].$$

Using (2.2) it is clear that

$$\mathrm{tr}(\tilde{\theta}^T\mathbf{A}(h)) = \frac{h}{2\pi}\sum_{k=1}^{\infty}\frac{1}{k}\left\{\mathbf{X}_k^T(\tilde{\theta} - \tilde{\theta}^T)(\mathbf{Y}_k + \sqrt{2/h}\Delta\mathbf{W}(h))\right\}.$$

Now since $\{\mathbf{X}_k\}_1^{\infty}$ and $\{\mathbf{Y}_k\}_1^{\infty}$ are two independent i.i.d. sequences of random vectors it follows that

$$\varphi_{\widetilde{A}(h)|\Delta\mathbf{W}(h)}(\tilde{\theta}) = \prod_{k=1}^{\infty}\varphi\left(\frac{h}{2\pi k}\tilde{\theta}\right),$$

where $\varphi(\tilde{\theta}) = E[\exp\{\iota\mathbf{X}_1^T(\tilde{\theta} - \tilde{\theta}^T)(\mathbf{Y}_1 + \sqrt{2/h}\Delta\mathbf{W}(h))\}\mid \Delta\mathbf{W}(h)]$. In order to simplify the notation we write $\bar{\theta}$ for $\tilde{\theta} - \tilde{\theta}^T$.

To compute $\varphi(\tilde{\theta})$, first calculate the conditional characteristic function given $\mathbf{Y}_1$,

$$\varphi_{|\mathbf{Y}_1}(\tilde{\theta}) = E[\exp\{\iota\mathbf{X}_1^T\bar{\theta}(\mathbf{Y}_1 + \sqrt{2/h}\Delta\mathbf{W}(h))\}\mid \mathbf{Y}_1, \Delta\mathbf{W}(h)]$$

$$= \exp\{-(\mathbf{Y}_1 + \sqrt{2/h}\Delta\mathbf{W}(h))^T\bar{\theta}\bar{\theta}^T(\mathbf{Y}_1 + \sqrt{2/h}\Delta\mathbf{W}(h))/2\}.$$

The random variable $Q = (\mathbf{Y}_1 + \sqrt{2/h}\Delta\mathbf{W}(h))^T\bar{\theta}\bar{\theta}^T(\mathbf{Y}_1 + \sqrt{2/h}\Delta\mathbf{W}(h))$ is a quadratic form in $\mathbf{Y}_1 + \sqrt{2/h}\Delta\mathbf{W}(h)$ with matrix $\bar{\theta}\bar{\theta}^T$. Thus $\varphi(\tilde{\theta}) = E_{\mathbf{Y}_1}\exp(-Q/2)$, which is the moment generating function of Q evaluated at the point $-1/2$. From Mathai and Provost [(1992), Theorem 3.2a.1] we get

$$E_{\mathbf{Y}_1}\exp(tQ) = \det(I_m - 2t\bar{\theta}\bar{\theta}^T)^{-1/2}$$

$$\times \exp\left(-\tfrac{1}{2}\sqrt{2/h}\Delta\mathbf{W}(h)^T\{I_m - (I_m - 2t\bar{\theta}\bar{\theta}^T)^{-1}\}\sqrt{2/h}\Delta\mathbf{W}(h)\right).$$

Hence

$$\varphi(\tilde{\theta}) = \det(I_m + \bar{\theta}\bar{\theta}^T)^{-1/2}$$

$$\times \exp\left(-\tfrac{1}{2}\sqrt{2/h}\Delta\mathbf{W}(h)^T\{I_m - (I_m + \bar{\theta}\bar{\theta}^T)^{-1}\}\sqrt{2/h}\Delta\mathbf{W}(h)\right),$$

and thus

$$\varphi_{\widetilde{A}(h)|\Delta\mathbf{W}(h)}(\tilde{\theta}) = \prod_{k=1}^{\infty} \det\left(I_m + \frac{c^2}{k^2}\bar{\theta}\bar{\theta}^T\right)^{-1/2}$$

$$\times \exp\left(-\frac{1}{2}\sqrt{2/h}\Delta\mathbf{W}(h)^T\left\{I_m - \left(I_m + \frac{c^2}{k^2}\bar{\theta}\bar{\theta}^T\right)^{-1}\right\}\sqrt{2/h}\Delta\mathbf{W}(h)\right)$$

$$= \det\left(\prod_{k=1}^{\infty}\left(I_m + \frac{c^2}{k^2}\bar{\theta}\bar{\theta}^T\right)^{-1}\right)^{1/2}$$

$$\times \exp\left(-\frac{1}{h}\Delta\mathbf{W}(h)^T\sum_{k=1}^{\infty}\left\{I_m - \left(I_m + \frac{c^2}{k^2}\bar{\theta}\bar{\theta}^T\right)^{-1}\right\}\Delta\mathbf{W}(h)\right),$$

where $c = h/(2\pi)$.

If $a \in \mathbf{R}^+$ then

$$\prod_{k=1}^{\infty}(1 + a/k^2)^{-1} = \frac{\pi\sqrt{a}}{\sinh(\pi\sqrt{a})}$$

and

$$\sum_{k=1}^{\infty}(1 - (1 + a/k^2)^{-1}) = \frac{1}{2}\left\{\cosh(\pi\sqrt{a})\frac{\pi\sqrt{a}}{\sinh(\pi\sqrt{a})} - 1\right\};$$

for $a = 0$ the functions by continuity equal one and zero, respectively. Since $\bar{\theta}\bar{\theta}^T$ is a symmetric and nonnegative definite matrix it then follows, by the spectral lemma for normal matrices [see, e.g., Golub and van Loan (1996) Theorem 11.1.3], that for $\bar{\theta}\bar{\theta}^T$ positive definite,

$$\prod_{k=1}^{\infty}\left(I_m + \frac{c^2}{k^2}\bar{\theta}\bar{\theta}^T\right)^{-1} = \frac{h}{2}\sqrt{\bar{\theta}\bar{\theta}^T}\sinh\left(\frac{h}{2}\sqrt{\bar{\theta}\bar{\theta}^T}\right)^{-1}$$

and

$$\sum_{k=1}^{\infty}\left\{I_m - \left(I_m + \frac{c^2}{k^2}\bar{\theta}\bar{\theta}^T\right)^{-1}\right\}$$

$$= \frac{1}{2}\left\{\cosh\left(\frac{h}{2}\sqrt{\bar{\theta}\bar{\theta}^T}\right)\frac{h}{2}\sqrt{\bar{\theta}\bar{\theta}^T}\sinh\left(\frac{h}{2}\sqrt{\bar{\theta}\bar{\theta}^T}\right)^{-1} - I_m\right\},$$

where the square root and the hyperbolic functions should be interpreted in the matrix sense. If $\bar{\theta}$ is singular some extra precaution is necessary and in fact $\bar{\theta}$ is always singular if m is odd. This follows from that $\bar{\theta}$ is skew symmetric and since skew-symmetric real matrices only have eigenvalues with zero real part, at least one of the eigenvalues must be zero. To avoid problems in the singular case we simply use the matrix version of the function $\text{sinch}(x)$ instead, where $\text{sinch}(x) = \sinh(x)/x$. This leads to the representation (3.2) and the proof is complete. $\square$

Note that the function $1/\operatorname{sinch}(\cdot)$ is well behaved on the entire real line and is in fact analytic in the strip $-\pi < \operatorname{Im}(x) < \pi$. This implies that the characteristic function is analytic in the region where all eigenvalues of $(h/2)\sqrt{\bar{\theta}\bar{\theta}^T}$ have their imaginary parts in $(-\pi, \pi)$.

We further remark that the conditional joint characteristic function of all the stochastic area integrals can in fact be easily obtained from the one calculated above, since exactly the same calculations go through with $\tilde{\theta}$ replaced by a matrix that has arbitrary real-valued elements in all positions.

4. Simulation of the iterated Itô integrals. Because the Brownian increments $\Delta \mathbf{W}(h)$ can easily be simulated exactly while the iterated Itô integrals cannot, we consider an algorithm that first simulates $\Delta \mathbf{W}(h)$ and then simulates approximations $\widehat{I}_{ij}(h)$ of $I_{ij}(h)$ conditional on the realized $\Delta \mathbf{W}(h)$. Furthermore, since the iterated Itô integrals and stochastic area integrals only differ by products of the Brownian increments, it is enough to simulate approximations $\widehat{A}_{ij}(h)$ of $A_{ij}(h)$. Then $\widehat{I}_{ij}(h)$ can easily be constructed. Because of the relations (2.1) we only need to approximate $A_{ij}(h)$ for $i < j$; see the previous section.

The MSE of interest is $\max_{ij} E(I_{ij}(h) - \widehat{I}_{ij}(h))^2$; see Kloeden and Platen [(1995), Corollary 10.6.5]. It turns out, however, to be more convenient to work with the sum of squared errors which is of course larger. Thus the error we consider is

$$E \sum_{i<j}(I_{ij}(h) - \widehat{I}_{ij}(h))^2 = E \sum_{i<j}(A_{ij}(h) - \widehat{A}_{ij}(h))^2.$$

The representation (2.2) immediately suggests a simulation algorithm; truncate the sum after n terms. This algorithm is essentially the one proposed by Kloeden, Platen and Wright (1992) and it has an MSE of order h^2/n. In order to improve on this rate, a careful analysis of the discarded tail-sum is needed. We shall show that this sum asymptotically has a multivariate Gaussian distribution, and that approximating it with a Gaussian random vector yields a convergence rate of order h^2/n^2.

To carry out this proof, some more notation is needed, and we also need to formalize the operation of picking out elements with indices $i < j$. We now show how to do this.

The Kronecker tensor product between two matrices A and B will be denoted $A \otimes B$ and is defined as

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix}.$$

If A is an $m \times n$ matrix and B is an $p \times q$ matrix, then $A \otimes B$ is an $mp \times nq$ matrix. The operation vec is defined as the $mn \times 1$ matrix obtained by stacking

the columns of a matrix on top of each other, that is,

$$\text{vec}(A) = \begin{bmatrix} A_{.1} \\ A_{.2} \\ \vdots \\ A_{.n} \end{bmatrix},$$

for an $m \times n$ matrix A.

The representation (2.2) can now be written with column vectors as

$$\text{vec}(\mathbf{I}(h)^T) = \frac{\Delta\mathbf{W}(h) \otimes \Delta\mathbf{W}(h) - \text{vec}(hI_m)}{2} + \text{vec}(\mathbf{A}(h)^T),$$

$$(4.1) \qquad \text{vec}(\mathbf{A}(h)^T) = \frac{h}{2\pi} \sum_{k=1}^{\infty} \frac{1}{k} \Bigg\{ \mathbf{X}_k \otimes \left(\mathbf{Y}_k + \sqrt{\frac{2}{h}} \Delta\mathbf{W}(h) \right)$$

$$- \left(\mathbf{Y}_k + \sqrt{\frac{2}{h}} \Delta\mathbf{W}(h) \right) \otimes \mathbf{X}_k \Bigg\}.$$

Let P_m be the $m^2 \times m^2$ permutation matrix which swaps rows i and $j = 1 + m((i-1) \bmod m) + (i-1) \text{ div } m$ for $i = 1, 2, \ldots, m^2$. Then $P_m(\mathbf{X}_k \otimes \mathbf{Y}_k) = \mathbf{Y}_k \otimes \mathbf{X}_k$. From the definition of P_m we see that it is symmetric and thus P_m is its own inverse. We rewrite the representation of $\text{vec}(\mathbf{I}(h)^T)$ as

$$\text{vec}(\mathbf{I}(h)^T) = \frac{\Delta\mathbf{W}(h) \otimes \Delta\mathbf{W}(h) - \text{vec}(hI_m)}{2} + \text{vec}(\mathbf{A}(h)^T),$$

$$(4.2)$$

$$\text{vec}(\mathbf{A}(h)^T) = \frac{h}{2\pi} \sum_{k=1}^{\infty} \frac{1}{k} (P_m - I_{m^2}) \Bigg\{ \left(\mathbf{Y}_k + \sqrt{\frac{2}{h}} \Delta\mathbf{W}(h) \right) \otimes \mathbf{X}_k \Bigg\}.$$

We now want to pick out the M-dimensional subset of $A_{ij}(h)$'s corresponding to $i < j$, where $M = m(m-1)/2$. Thus define

$$\widetilde{A}(h) = [A_{12}(h), \ldots, A_{1m}(h), \ldots, A_{l,l+1}(h), \ldots, A_{lm}(h), \ldots, A_{m-1,m}(h)]^T.$$

This column vector can also be written

$$\widetilde{A}(h) = K_m \, \text{vec}(\mathbf{A}(h)^T),$$

where K_m is an $M \times m^2$ matrix which picks out elements $2, \ldots, m, \ldots,$ $(k-1)m + k + 1, \ldots, km, \ldots, M$ from $\text{vec}(\mathbf{A}(h)^T)$. The matrix K_m is thus

$$K_m = \begin{bmatrix} 0_{m-1\times 1} & I_{m-1} & 0_{m-1\times m(m-1)} \\ 0_{m-2\times m+2} & I_{m-2} & 0_{m-2\times m(m-2)} \\ \vdots & \vdots & \vdots \\ 0_{m-k\times(k-1)m+k} & I_{m-k} & 0_{m-k\times m(m-k)} \\ \vdots & \vdots & \vdots \\ 0_{1\times(m-2)m+m-1} & 1 & 0_{1\times m} \end{bmatrix}.$$

One may verify that K_m and P_m satisfy the following relations:

$$K_m K_m^T = I_M,$$

$$K_m^T K_m = \text{diag}(0, 1_{m-1}^T, \ldots, 0_k^T, 1_{m-k}^T, \ldots, 0_{m-1}^T, 1, 0_m^T),$$

(4.3) $$K_m P_m K_m^T = 0_{M \times M},$$

$$K_m I_{m^2} K_m^T = I_M,$$

$$(I_{m^2} - P_m) K_m^T K_m (I_{m^2} - P_m) = I_{m^2} - P_m,$$

where $\text{diag}(x^T)$ is the diagonal matrix with x^T on its diagonal. Given $\widetilde{A}(h)$, the other stochastic area integrals can be easily be generated from $\widetilde{A}(h)$ using relations (2.1). More precisely,

$$\text{vec}(\mathbf{A}(h)^T) = (I_{m^2} - P_m) K_m^T \widetilde{A}(h).$$

It follows from (4.2) that we can write $\widetilde{A}(h)$ as

$$\widetilde{A}(h) = \frac{h}{2\pi} \sum_{k=1}^{\infty} \frac{1}{k} K_m (P_m - I_{m^2}) \left\{ \left(\mathbf{Y}_k + \sqrt{\frac{2}{h}} \Delta \mathbf{W}(h) \right) \otimes \mathbf{X}_k \right\}.$$

We now split this sum into two by defining

$$\widetilde{A}^{(n)}(h) = \frac{h}{2\pi} \sum_{k=1}^{n} \frac{1}{k} K_m (P_m - I_{m^2}) \left\{ \left(\mathbf{Y}_k + \sqrt{\frac{2}{h}} \Delta \mathbf{W}(h) \right) \otimes \mathbf{X}_k \right\}$$

and

$$\varepsilon_n = \frac{h}{2\pi} \sum_{k=n+1}^{\infty} \frac{1}{k} K_m (P_m - I_{m^2}) \left\{ \left(\mathbf{Y}_k + \sqrt{\frac{2}{h}} \Delta \mathbf{W}(h) \right) \otimes \mathbf{X}_k \right\}.$$

It easy to see from these definitions that given $\Delta \mathbf{W}(h)$, ε_n and $\widetilde{A}^{(n)}(h)$ are conditionally independent. We proceed by examining the tail-sum ε_n in closer detail.

The term $(P_m - I_{m^2})\{(\mathbf{Y}_k + \sqrt{2/h}\Delta \mathbf{W}(h)) \otimes \mathbf{X}_k\}$ in the above sum can for each k, given $\mathbf{Y}_k$ and $\Delta \mathbf{W}(h)$, be seen as a conditionally Gaussian column vector with conditional mean 0_{m^2} and conditional covariance matrix $\Sigma(\mathbf{Y}_k)$, where

$$\Sigma(\mathbf{Y}_k) = (I_{m^2} - P_m)\left((\mathbf{Y}_k + \sqrt{2/h}\Delta \mathbf{W}(h)) \otimes I_m \right)$$

(4.4) $$\times \left((\mathbf{Y}_k + \sqrt{2/h}\Delta \mathbf{W}(h))^T \otimes I_m \right)(I_{m^2} - P_m)$$

$$= (I_{m^2} - P_m)\left((\mathbf{Y}_k + \sqrt{2/h}\Delta \mathbf{W}(h))(\mathbf{Y}_k + \sqrt{2/h}\Delta \mathbf{W}(h))^T \otimes I_m \right)$$

$$\times (I_{m^2} - P_m).$$

Hence, given $\mathbf{Y} = \{\mathbf{Y}_k\}_1^\infty$ and $\Delta\mathbf{W}(h)$, the conditional distribution of ε_n is Gaussian with mean 0_M and covariance matrix $(h/2\pi)^2 \sum_{k=n+1}^\infty \widetilde{\Sigma}(\mathbf{Y}_k)/k^2$, where $\widetilde{\Sigma}(\mathbf{Y}_k) = K_m\Sigma(\mathbf{Y}_k)K_m^T$. Write

$$\varepsilon_n = \frac{h}{2\pi}\left(\sum_{k=n+1}^\infty \frac{\widetilde{\Sigma}(\mathbf{Y}_k)}{k^2}\right)^{1/2}\mathbf{G}_n,$$

where $\mathbf{G}_n$ is the random vector

$$\mathbf{G}_n = \frac{2\pi}{h}\left(\sum_{k=n+1}^\infty \widetilde{\Sigma}(\mathbf{Y}_k)/k^2\right)^{-1/2}\varepsilon_n.$$

The conditional distribution of $\mathbf{G}_n$ given $\mathbf{Y}$ and $\Delta\mathbf{W}(h)$ is thus a standard Gaussian distribution $N(0_M, I_M)$, that is, the conditional distribution does not depend on $\mathbf{Y}$ and $\Delta\mathbf{W}(h)$. Hence $\mathbf{G}_n$ is a standard Gaussian vector independent of $\mathbf{Y}$ and $\Delta\mathbf{W}(h)$. From the above calculations it is evident that the random vector ε_n is a Gaussian variance mixture with random covariance matrix $\sum_{k=n+1}^\infty \widetilde{\Sigma}(\mathbf{Y}_k)/k^2$.

We remark that in the scalar case ($m = 2$), the distribution of the mixing random variance is infinitely divisible, which is equivalent to saying that ε_n has a so-called class G distribution. This was shown in Rydén and Wiktorsson (1999). For the definition of class G distributions, see, for example, Rosinski (1990). In the case $m > 2$ it holds that each element in the random covariance matrix is infinitely divisible, but we have not been able to prove that ε_n has a multivariate class G distribution.

We shall now examine the asymptotic properties of the random covariance matrix. Let Σ_n denote the normalized version, that is,

$$\Sigma_n = \frac{1}{a_n}\sum_{k=n+1}^\infty \frac{\widetilde{\Sigma}(\mathbf{Y}_k)}{k^2},$$

where $a_n = \sum_{k=n+1}^\infty 1/k^2$. Define $\Sigma_\infty = E_{\mathbf{Y}_1}\widetilde{\Sigma}(\mathbf{Y}_1)$. Taking the expectation in (4.4) and using (4.3) it follows that

$$(4.5) \quad \Sigma_\infty = 2I_M + \frac{2}{h}K_m(I_{m^2} - P_m)\bigl(I_m \otimes \Delta\mathbf{W}(h)\Delta\mathbf{W}(h)^T\bigr)(I_{m^2} - P_m)K_m^T.$$

We will show below that Σ_n converges in mean square sense to the constant matrix Σ_∞ as $n \to \infty$. This implies that $(2\pi/h)a_n^{-1/2}\varepsilon_n$ converges weakly to a Gaussian vector with zero mean and covariance matrix Σ_∞. This property gives rise to the following improved simulation algorithm for $\text{vec}(\mathbf{I}(h)^T)$.

1. Simulate $\Delta\mathbf{W}(h)$ from $N(0_m, \sqrt{h}I_m)$.
2. First approximate the stochastic area integrals as

$$\widetilde{A}^{(n)}(h) = \frac{h}{2\pi} \sum_{k=1}^{n} \frac{1}{k} K_m(P_m - I_{m^2})\left\{\left(\mathbf{Y}_k + \sqrt{\frac{2}{h}}\Delta\mathbf{W}(h)\right) \otimes \mathbf{X}_k\right\},$$

where $\mathbf{X}_k \sim N(0_m, I_m)$ and $\mathbf{Y}_k \sim N(0_m, I_m)$.
3. Simulate $\mathbf{G}_n \sim N(0_M, I_M)$ and add the tail-sum approximation:

$$\widetilde{A}^{(n)'}(h) = \widetilde{A}^{(n)}(h) + \frac{h}{2\pi} a_n^{1/2}\sqrt{\Sigma_\infty}\mathbf{G}_n.$$

4. Finally define the approximation $\text{vec}(\mathbf{I}(h)^T)^{(n)'}$ of $\text{vec}(\mathbf{I}(h)^T)$ as

$$\text{vec}(\mathbf{I}(h)^T)^{(n)'} = \frac{\Delta\mathbf{W}(h) \otimes \Delta\mathbf{W}(h) - \text{vec}(h I_{m^2})}{2}$$

$$+ (I_{m^2} - P_m)K_m^T \widetilde{A}^{(n)'}(h).$$

The following result gives a bound on the maximal conditional MSE given $\Delta\mathbf{W}(h)$. We also give an explicit expression for the square root of the asymptotic covariance matrix Σ_∞.

THEOREM 4.1.   (i) *The maximal conditional MSE for the approximation of the iterated Itô integrals given* $\Delta\mathbf{W}(h)$, *for* $n \geq 1$, *satisfies*

$$\max_{i,\,j} E\Big[\big|I_{ij}(h) - I_{ij}^{(n)'}(h)\big|^2 \mid \Delta\mathbf{W}(h)\Big]$$

(4.6)
$$\leq \sum_{i<j} E\Big[\big|I_{ij}(h) - I_{ij}^{(n)'}(h)\big|^2 \mid \Delta\mathbf{W}(h)\Big]$$

$$\leq \frac{h^2}{24\pi^2 n^2}m(m-1)(m + 4|\Delta\mathbf{W}(h)|^2/h).$$

(ii) *The matrix square root* $\sqrt{\Sigma_\infty}$ *can be explicitly written as*

(4.7)
$$\sqrt{\Sigma_\infty} = \frac{\Sigma_\infty + 2\sqrt{1 + |\Delta\mathbf{W}(h)|^2/h}I_M}{\sqrt{2}(1 + \sqrt{1 + |\Delta\mathbf{W}(h)|^2/h})}.$$

As mentioned in the introduction, the MSE in the simulation of the iterated Itô integrals should be no larger than Ch^3. We then see that it is enough to choose

(4.8)
$$n \geq \frac{\sqrt{m(m-1)(m + 4|\Delta\mathbf{W}(h)|^2/h)/(24\pi^2)}}{\sqrt{Ch}}$$

in the approximation of the iterated Itô integral.

Taking the expectation over $\Delta\mathbf{W}(h)$ of the right-hand side of (4.6), it follows that

$$\max_{i,\,j} E\Big[\big|I_{ij}(h) - I_{ij}^{(n)'}(h)\big|^2\Big] \leq \frac{5h^2}{24\pi^2 n^2}m^2(m-1).$$

Hence, there are two ways of choosing n. Either we first simulate $\Delta\mathbf{W}(h)$ and then choose n according to (4.8), or we fix n beforehand according to

$$(4.9) \qquad n \geq \frac{\sqrt{5m^2(m-1)/(24\pi^2)}}{\sqrt{Ch}},$$

and thus do not take $\Delta\mathbf{W}(h)$ into account when selecting n. By Jensen's inequality, the mean of the right-hand side of (4.8) is smaller than the right-hand side of (4.9), so that the first way of choosing n yields a smaller n on the average.

PROOF. The first part is an immediate consequence of Theorem 4.3 below. The second part follows by direct calculation. First let $a = \sqrt{1 + |\Delta\mathbf{W}(h)|^2/h}$ to simplify the notation. Then, with A as in the right-hand side of (4.7),

$$A^2 = \frac{\Sigma_\infty^2 + 4a\Sigma_\infty + 4a^2 I_M}{2(1+a)^2}$$

$$= \frac{\Sigma_\infty^2 - (2 + 2a^2)\Sigma_\infty + 4a^2 I_M}{2(1+a)^2} + \frac{2 + 2a^2 + 4a}{2 + 2a^2 + 4a}\Sigma_\infty$$

$$= \frac{\Sigma_\infty^2 - (2 + 2a^2)\Sigma_\infty + 4a^2 I_M}{2(1+a)^2} + \Sigma_\infty.$$

Hence we need to show that

$$\Sigma_\infty^2 - (2 + 2a^2)\Sigma_\infty + 4a^2 I_M = 0_{M \times M},$$

which is the same as saying that Σ_∞ has minimal polynomial $(x-2)(x-2a^2)$. This is, further, the same as saying that Σ_∞ only has two different eigenvalues, namely 2 and $2 + 2|\Delta\mathbf{W}(h)(h)|^2/h$, and since $\Sigma_\infty = 2I_M + B$, where B is the nonnegative definite matrix $K_m(I_{m^2} - P_m)((2/h)\Delta\mathbf{W}(h)\Delta\mathbf{W}(h)^T \otimes I_m)(I_{m^2} - P_m)K_m^T$, it is enough to show that $B^2 = (2|\Delta\mathbf{W}(h)|^2/h)B$. This is equivalent to that B has minimal polynomial $x(x-2|\Delta\mathbf{W}(h)|^2/h)$. Writing W for $\sqrt{2/h}\Delta\mathbf{W}(h)$ to simplify the notation and using (4.3), it follows that

$$B^2 = K_m(I_{m^2} - P_m)\Big(WW^T \otimes I_m\Big)\Big(WW^T \otimes I_m\Big)(I_{m^2} - P_m)K_m^T$$

$$\quad - K_m(I_{m^2} - P_m)\Big(WW^T \otimes I_m\Big)P_m\Big(WW^T \otimes I_m\Big)(I_{m^2} - P_m)K_m^T$$

$$= K_m(I_{m^2} - P_m)\Big(|W|^2 WW^T \otimes I_m\Big)(I_{m^2} - P_m)K_m^T$$

$$\quad + K_m(I_{m^2} - P_m)\Big(WW^T \otimes I_m\Big)P_m\Big(WW^T \otimes I_m\Big)P_m(I_{m^2} - P_m)K_m^T.$$

Now using the definition of P_m it follows that

$$B^2 = |W|^2 B + K_m(I_{m^2} - P_m)(WW^T \otimes WW^T)(I_{m^2} - P_m)K_m^T.$$

The last term is zero since

$$(WW^T \otimes WW^T) = (W \otimes W)(W^T \otimes W^T),$$

and from the definition of P_m it is evident that $(I_{m^2} - P_m)(W \otimes W) = 0_{m^2}$. This concludes the proof. $\square$

In the following the usual operator norm will be denoted $\|B\|$ and the Frobenius norm of a $p \times q$ matrix B, defined as $(\sum_{i=1}^{p} \sum_{j=1}^{q} B_{ij}^2)^{1/2}$, will be denoted $\|B\|_F$.

THEOREM 4.2.   *Conditional on $\Delta\mathbf{W}(h)$, Σ_n converges in mean square sense to Σ_∞. Moreover,*

$$E_{\mathbf{Y}}\|\Sigma_n - \Sigma_\infty\|_F^2 \le \frac{d_m}{3n} \quad \text{for } n \ge 1,$$

where $d_m = m(m-1)(m + 4|\Delta\mathbf{W}(h)|^2/h)$.

PROOF.   First observe that

$$E_{\mathbf{Y}}\Sigma_n = \frac{1}{a_n}\left(\sum_{k=n+1}^{\infty} E_{\mathbf{Y}}\frac{\widetilde{\Sigma}(\mathbf{Y}_k)}{k^2}\right) = E_{\mathbf{Y}}\widetilde{\Sigma}(\mathbf{Y}_1) = \Sigma_\infty.$$

Hence, by the definition of the Frobenius norm,

$$E_{\mathbf{Y}}\|\Sigma_n - \Sigma_\infty\|_F^2 = \sum_{p,\,q=1}^{M} E_{\mathbf{Y}}(\Sigma_{n,\,pq} - E_{\mathbf{Y}}\Sigma_{n,\,pq})^2 = \sum_{p,\,q=1}^{M} V_{\mathbf{Y}}(\Sigma_{n,\,pq}),$$

and since $\{\widetilde{\Sigma}(\mathbf{Y}_k)\}$ is an i.i.d. sequence of matrices we obtain

$$E_{\mathbf{Y}}\|\Sigma_n - \Sigma_\infty\|_F^2 = \sum_{p,\,q=1}^{M} \frac{1}{a_n^2}\left(\sum_{k=n+1}^{\infty} \frac{V_{\mathbf{Y}}(\widetilde{\Sigma}(\mathbf{Y}_k)_{pq})}{k^4}\right)$$

$$= \frac{b_n}{a_n^2}\sum_{p,\,q=1}^{M} V_{\mathbf{Y}}(\widetilde{\Sigma}(\mathbf{Y}_1)_{pq}) = \frac{b_n}{a_n^2}d_m,$$

where $b_n = \sum_{k=n+1}^{\infty} 1/k^4$ and $d_m = \sum_{p,\,q=1}^{M} V_{\mathbf{Y}}(\widetilde{\Sigma}(\mathbf{Y}_1)_{pq})$. Approximating the sums a_n and b_n by integrals yields

$$\frac{b_n}{a_n^2} = \frac{\sum_{k=n+1}^{\infty} 1/k^4}{\left(\sum_{k=n+1}^{\infty} 1/k^2\right)^2} \le \frac{\int_{n+1/2}^{\infty} 1/x^4 dx}{\left(\int_{n+3/4}^{\infty} 1/x^2 dx\right)^2} = \frac{(n+3/4)^2}{3(n+1/2)^3} \le \frac{1}{3n} \quad \text{for } n \ge 1.$$

We now turn to the calculation of d_m. Recall the definition (4.4) of $\Sigma(\mathbf{Y}_k)$. To simplify the notation we drop the index k on $\mathbf{Y}_k$ and define

$$Q = \left(\mathbf{Y} + \sqrt{2/h}\Delta\mathbf{W}(h)\right)\left(\mathbf{Y} + \sqrt{2/h}\Delta\mathbf{W}(h)\right)^T \otimes I_m$$

$$- E_{\mathbf{Y}}\left(\mathbf{Y} + \sqrt{2/h}\Delta\mathbf{W}(h)\right)\left(\mathbf{Y} + \sqrt{2/h}\Delta\mathbf{W}(h)\right)^T \otimes I_m$$

$$= \left(\mathbf{Y}\mathbf{Y}^T + \mathbf{Y}\sqrt{2/h}\Delta\mathbf{W}(h)^T + \sqrt{2/h}\Delta\mathbf{W}(h)\mathbf{Y}^T - I_m\right) \otimes I_m.$$

Then we can write

$$
d_m = \sum_{p,\,q=1}^{M} V_{\mathbf{Y}}(\widetilde{\Sigma}(\mathbf{Y})_{pq})
$$

$$
= E_{\mathbf{Y}}\operatorname{tr}\Big((\widetilde{\Sigma}(\mathbf{Y}) - E_{\mathbf{Y}}\widetilde{\Sigma}(\mathbf{Y}))(\widetilde{\Sigma}(\mathbf{Y}) - E_{\mathbf{Y}}\widetilde{\Sigma}(\mathbf{Y}))^{T}\Big)
$$

$$
= E_{\mathbf{Y}}\operatorname{tr}\Big(K_m(I_{m^2} - P_m)Q(I_{m^2} - P_m)K_m^{T}
$$

$$
\times K_m(I_{m^2} - P_m)Q(I_{m^2} - P_m)K_m^{T}\Big)
$$

$$
= E_{\mathbf{Y}}\operatorname{tr}\Big(Q(I_{m^2} - P_m)K_m^{T}K_m(I_{m^2} - P_m)
$$

$$
\times Q(I_{m^2} - P_m)K_m^{T}K_m(I_{m^2} - P_m)\Big)
$$

$$
= E_{\mathbf{Y}}\operatorname{tr}\big(Q(I_{m^2} - P_m)Q(I_{m^2} - P_m)\big)
$$

$$
= E_{\mathbf{Y}}\operatorname{tr}\big(Q^2 - 2Q^2 P_m + QP_m QP_m\big)
$$

$$
= E_{\mathbf{Y}}\operatorname{tr}(Q^2) - 2E_{\mathbf{Y}}\operatorname{tr}\big(Q^2 P_m\big) + E_{\mathbf{Y}}\operatorname{tr}(QP_m QP_m),
$$

where we used (4.3). Now start with $\operatorname{tr}(Q^2)$. We replace $\sqrt{2/h}\Delta\mathbf{W}(h)$ with $\widetilde{\mathbf{W}}$ to simplify the notation:

$$
E_{\mathbf{Y}}\operatorname{tr}(Q^2) = \operatorname{tr}(I_m)E_{\mathbf{Y}}\operatorname{tr}((\mathbf{Y}\mathbf{Y}^T + \widetilde{\mathbf{W}}\mathbf{Y}^T + \mathbf{Y}\widetilde{\mathbf{W}}^T - I_m)^2)
$$

$$
= mE_{\mathbf{Y}}\operatorname{tr}(\mathbf{Y}\mathbf{Y}^T\mathbf{Y}\mathbf{Y}^T + \mathbf{Y}\mathbf{Y}^T\widetilde{\mathbf{W}}\mathbf{Y}^T + \mathbf{Y}\mathbf{Y}^T\mathbf{Y}\widetilde{\mathbf{W}}^T - \mathbf{Y}\mathbf{Y}^T)
$$

$$
+ mE_{\mathbf{Y}}\operatorname{tr}(\widetilde{\mathbf{W}}\mathbf{Y}^T\mathbf{Y}\mathbf{Y}^T + \widetilde{\mathbf{W}}\mathbf{Y}^T\widetilde{\mathbf{W}}\mathbf{Y}^T + \widetilde{\mathbf{W}}\mathbf{Y}^T\mathbf{Y}\widetilde{\mathbf{W}}^T - \widetilde{\mathbf{W}}\mathbf{Y}^T)
$$

$$
+ mE_{\mathbf{Y}}\operatorname{tr}(\mathbf{Y}\widetilde{\mathbf{W}}^T\mathbf{Y}\mathbf{Y}^T + \mathbf{Y}\widetilde{\mathbf{W}}^T\widetilde{\mathbf{W}}\mathbf{Y}^T + \mathbf{Y}\widetilde{\mathbf{W}}^T\mathbf{Y}\widetilde{\mathbf{W}}^T - \mathbf{Y}\widetilde{\mathbf{W}}^T)
$$

$$
+ mE_{\mathbf{Y}}\operatorname{tr}(-\mathbf{Y}\mathbf{Y}^T - \widetilde{\mathbf{W}}\mathbf{Y}^T - \mathbf{Y}\widetilde{\mathbf{W}}^T + I_m)
$$

$$
= m\{(m(m+2) + 0 + 0 - m)
$$

$$
+ (0 + 2|\Delta\mathbf{W}(h)|^2 + 2m|\Delta\mathbf{W}(h)|^2 + 0)/h
$$

$$
+ (0 + 2m|\Delta\mathbf{W}(h)|^2 + 2|\Delta\mathbf{W}(h)|^2 + 0)/h
$$

$$
+ (-m + 0 + 0 + m)\}
$$

$$
= m(m+1)(m + 4|\Delta\mathbf{W}(h)|^2/h).
$$

Continue with $\operatorname{tr}(QP_m QP_m)$. By the definition of P_m,

$$
P_m QP_m = I_m \otimes (\mathbf{Y}\mathbf{Y}^T + \mathbf{Y}\widetilde{\mathbf{W}}^T + \widetilde{\mathbf{W}}^T\mathbf{Y} - I_m),
$$

and thus

$$
\begin{aligned}
E_{\mathbf{Y}}\operatorname{tr}(QP_mQP_m) &= E_{\mathbf{Y}}[\operatorname{tr}\{(\mathbf{Y}\mathbf{Y}^T + \mathbf{Y}\widetilde{\mathbf{W}}^T + \widetilde{\mathbf{W}}^T\mathbf{Y} - I_m) \\
&\qquad \otimes (\mathbf{Y}\mathbf{Y}^T + \mathbf{Y}\widetilde{\mathbf{W}}^T + \widetilde{\mathbf{W}}^T\mathbf{Y} - I_m)\}] \\
&= E_{\mathbf{Y}}[(\operatorname{tr}(\mathbf{Y}\mathbf{Y}^T + \mathbf{Y}\widetilde{\mathbf{W}}^T + \widetilde{\mathbf{W}}^T\mathbf{Y} - I_m))^2] \\
&= E_{\mathbf{Y}}[(\mathbf{Y}^T\mathbf{Y} + 2\widetilde{\mathbf{W}}^T\mathbf{Y} - m)^2] \\
&= E_{\mathbf{Y}}[(\mathbf{Y}^T\mathbf{Y})^2 + 4(\widetilde{\mathbf{W}}^T\mathbf{Y})^2 + m^2 \\
&\qquad + 4\mathbf{Y}^T\mathbf{Y}\widetilde{\mathbf{W}}^T\mathbf{Y} - 2m\mathbf{Y}^T\mathbf{Y} - 4m\widetilde{\mathbf{W}}^T\mathbf{Y}] \\
&= m(m+2) + 8|\Delta\mathbf{W}(h)|^2/h + m^2 + 0 - 2m^2 + 0 \\
&= 2(m + 4|\Delta\mathbf{W}(h)|^2/h).
\end{aligned}
$$

Finally $-2E_{\mathbf{Y}}\operatorname{tr}(Q^2P_m) = -2(m+1)(m+4|\Delta\mathbf{W}(h)|^2)$ is obtained similarly, giving that

$$
d_m = m(m-1)(m + 4|\Delta\mathbf{W}(h)|/h).
$$

This concludes the proof. □

In order to calculate the MSE in the approximation of $\operatorname{vec}(\mathbf{I}(h)^T)$ the following lemma will be useful.

LEMMA 4.1. *If A and B are two symmetric positive definite matrices then*

$$
\|A^{1/2} - B^{1/2}\|_F \le \beta^{-1/2}\|A - B\|_F,
$$

where β is the smallest eigenvalue of B.

PROOF.   First observe that for $a \in \mathbf{R}^+$,

$$
\sqrt{a} = \frac{1}{\pi}\int_0^\infty \frac{a}{a+t}t^{-1/2}\,dt.
$$

By the spectral lemma for normal matrices it then follows that

$$
A^{1/2} = \frac{1}{\pi}\int_0^\infty (A + tI)^{-1}At^{-1/2}\,dt = \frac{1}{\pi}\int_0^\infty A(A + tI)^{-1}t^{-1/2}\,dt
$$

for any symmetric positive definite matrix A. Thus

$$
A^{1/2} - B^{1/2} = \frac{1}{\pi}\int_0^\infty (A + tI)^{-1}(A - B)(B + tI)^{-1}t^{1/2}\,dt,
$$

so that

$$
\|A^{1/2} - B^{1/2}\|_F \le \frac{1}{\pi}\int_0^\infty \|(A + tI)^{-1}(A - B)(B + tI)^{-1}\|_F t^{1/2}\,dt.
$$

Using twice that $\|CD\|_F \le \|C\|\|D\|_F$ for any pair of symmetric matrices C and D, we obtain

$$
\|(A + tI)^{-1}(A - B)(B + tI)^{-1}\|_F \le \|(A + tI)^{-1}\|\|A - B\|_F\|(B + tI)^{-1}\|.
$$

We bound these norms from above by $\|(A + tI)^{-1}\| \leq t^{-1}$ and $\|(B + tI)^{-1}\| \leq 1/(t + \beta)$, yielding

$$\|A^{1/2} - B^{1/2}\|_F \leq \|A - B\|_F \frac{1}{\pi\beta} \int_0^\infty \frac{\beta}{\beta + t} t^{-1/2}\, dt = \beta^{-1/2}\|A - B\|_F. \qquad \square$$

THEOREM 4.3. *The conditional MSE, given* $\Delta\mathbf{W}(h)$, *in the approximation* $(h/2\pi)a_n^{1/2}\sqrt{\Sigma_\infty}G_n$ *of* ε_n *is bounded by*

$$E\left[\left|\varepsilon_n - \frac{h}{2\pi}a_n^{1/2}\sqrt{\Sigma_\infty}G_n\right|^2 \mid \Delta\mathbf{W}(h)\right] \leq \frac{d_m c^2}{6n^2} \quad \text{for } n \geq 1,$$

where $c = h/(2\pi)$ *and* $d_m = m(m - 1)(m + 4|\Delta\mathbf{W}(h)|^2/h)$.

PROOF.

$$E\left[\left|\varepsilon_n - \frac{h}{2\pi}a_n^{1/2}\sqrt{\Sigma_\infty}G_n\right|^2 \mid \Delta\mathbf{W}(h)\right]$$

$$= c^2 a_n E[|\left(\sqrt{\Sigma_n} - \sqrt{\Sigma_\infty}\right)\mathbf{G}_n|^2 \mid \Delta\mathbf{W}(h)]$$

$$= c^2 a_n E_{\mathbf{Y}} E[|\left(\sqrt{\Sigma_n} - \sqrt{\Sigma_\infty}\right)\mathbf{G}_n|^2 \mid \mathbf{Y}, \Delta\mathbf{W}(h)]$$

$$= c^2 a_n E_{\mathbf{Y}} \|\sqrt{\Sigma_n} - \sqrt{\Sigma_\infty}\|_F^2.$$

By Lemma 4.1,

$$E_{\mathbf{Y}}\|\sqrt{\Sigma_n} - \sqrt{\Sigma_\infty}\|_F^2 \leq \frac{1}{\beta_{\Sigma_\infty}} E_{\mathbf{Y}}\|\Sigma_n - \Sigma_\infty\|_F^2 = \frac{1}{2} E_{\mathbf{Y}}\|\Sigma_n - \Sigma_\infty\|_F^2;$$

here $\beta_{\Sigma_\infty} = 2$ follows from the proof of Theorem 4.1. Now use Theorem 4.2 and the bound $a_n \leq 1/n$ to complete the proof. $\square$

We remark that the tail sum associated with the full vector $\text{vec}(A(h)^T)$ of stochastic area integrals is a Gaussian variance mixture as well, and its conditional covariance matrix converges in mean square sense to a nonstochastic limit. However, this limit, as well as the conditional covariance matrix, are singular and indeed have ranks M so that the corresponding Gaussian variance mixture have support on a subspace of dimension M. This invalidates the use of Lemma 4.1 for the asymptotic analysis of the algorithm and is the main reason for working with the set $\widetilde{\mathbf{A}}(h)$ throughout this section.

REFERENCES

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

GOLUB, G. H. and VAN LOAN, C. F. (1996). *Matrix Computations*, 3rd ed. John Hopkins Univ. Press.

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

KLOEDEN, P. E., PLATEN, E. and WRIGHT, W. (1992). The approximation of multiple stochastic integrals. *Stochastic Anal. Appl.* **10** 431–441.

LÉVY, P. (1951). Wiener's random functional and other Laplacian random functionals. *Proc. Second Berkeley Symp. Math. Statist. Probab.* 171–187. Univ. California Press, Berkeley.

LÉVY, P. (1965). *Processus stochastiques et mouvement Brownien*. Gauthier-Villars, Paris.

MATHAI, A. M. and PROVOST, S. B. (1992). *Quadratic Forms in Random Variables: Theory and Applications*. Decker, New York.

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

ROSINSKI, J. (1990). On the representation of infinitely divisible random vectors. *Ann. Probab.* **18** 405–430.

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

RYDÉN, T. and WIKTORSSON, M. (2001). On the simulation of iterated Itô integrals. *Stochastic Process. Appl.* **91** 151–168.

TALACKO, J. (1956). Perks distributions and their role in the theory of Wiener's stochastic variables. *Trabajos de Estadistica* **7** 159–174.

CENTRE FOR MATHEMATICAL SCIENCES
MATHEMATICAL STATISTICS
LUND UNIVERSITY
BOX 118
SE-221 00 LUND
SWEDEN
E-MAIL: magnusw@maths.lth.se