THE GEOMETRY OF ITERATED STRATONOVICH INTEGRALS

CHRISTIAN BAYER

ABSTRACT. We give a summary on the geometry of iterated Stratonovich integrals. For this exposition, we always have the connection to stochastic Taylor expansion in mind. In particular, we believe that "cubature on Wiener space" is best understood in the setting presented in this text. Besides cubature on Wiener space, we also give a second application regarding the heat kernel on nilpotent free Lie groups.

1. INTRODUCTION

Although iterated Stratonovich integrals of Brownian motion are interesting objects per se, our motivation for studying them stems from the *stochastic Taylor expansion*, where they play a similar rôle as polynomials do in the deterministic Taylor expansion.

Let $B_t = (B_t^1, \dots, B_t^d)$, $t \geq 0$, denote a d-dimensional Brownian motion on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$ satisfying the usual conditions. For notational issues, we introduce a 0th component $B_t^0 = t$. Let $V_0, \dots, V_d : \mathbb{R}^n \to \mathbb{R}^n$ be C^∞-*bounded* vector fields on $\mathbb{R}^n$, i. e. smooth vector fields such that all their derivatives are bounded. We also require the vector fields themselves to be bounded.

Remark 1.1. We identify vector fields with the corresponding first order differential operators, i. e. we identify a vector field $V = (V^1, \dots, V^n)$ as above with the operator $\sum_{i=1}^n V^i(x) \frac{\partial}{\partial x^i}$. Consequently, V^2 denotes the second order differential operator

$$V^2 f(x) = \sum_{i,j=1}^n V^j(x) \frac{\partial V^i}{\partial x^j}(x) \frac{\partial f}{\partial x^i}(x) + \sum_{i,j=1}^n V^j(x) V^i(x) \frac{\partial^2 f}{\partial x^i \partial x^j}(x),$$

where $f \in C^2(\mathbb{R}^n)$.

Furthermore, let Y_t^y be the (unique strong) solution of the SDE (in Stratonovich formulation)

$$\begin{cases} dY_t^y = V_0(Y_t^y)dt + \sum_{i=1}^d V_i(Y_t^y) \circ dB_t^i \\ Y_0^y = y \end{cases}, \tag{1.1}$$

for $y \in \mathbb{R}^n$.

Given a multi-index $I = (i_1, \dots, i_k) \in \{0, 1, \dots, d\}^k$, $k \in \mathbb{N}$, define a degree by $\deg(I) = k + \#\{j \in \{1, \dots, k\} \mid i_j = 0\}$ and $\deg(\emptyset) = 0$, $\emptyset$ denoting the empty multi-index. This definition takes care of the different orders of the $d+1$ components of B in the sense that $\Delta B_t^0 = \Delta t$ and $\Delta B_t^i \approx \sqrt{\Delta t}$, $i \in \{1, \dots, d\}$. The set $\bigcup_{k=0}^\infty \{0, \dots, d\}^k$ of all multi-indices is denoted by $\mathcal{A}$.

Proposition 1.1 (Stochastic Taylor expansion)**.** *Let $f \in C_b^{m+2}(\mathbb{R}^n)$ and fix $m \in \mathbb{N}$. Then*

$$f(Y_t^y) = \sum_{\substack{I=(i_1,\dots,i_k)\in\mathcal{A} \\ \deg(I)\le m,\, k\in\mathbb{N}}} V_{i_1}\cdots V_{i_k} f(y) B_t^I + R_m(t,y,f),$$

where the iterated Stratonovich integral *is defined by*

$$B_t^{(i_1,\dots,i_k)} = \int\limits_{0<t_1<\cdots<t_k<t} \circ dB_{t_1}^{i_1} \circ \cdots \circ dB_{t_k}^{i_k} = \int\limits_0^t \int\limits_0^{t_k} \cdots \int\limits_0^{t_2} \circ dB_{t_1}^{i_1} \circ \cdots \circ dB_{t_k}^{i_k} \tag{1.2}$$

for $\deg(i_1,\dots,i_k) > 0$ *and* $B_t^{\emptyset} = 1$*. The remainder term satisfies*

$$\sup_{y\in\mathbb{R}^n} \sqrt{E\big(R_m(t,y,f)^2\big)} \le C t^{\frac{m+1+\mathbf{1}_{\{t>1\}}}{2}} \sup_{\substack{I=(i_1,\dots,i_k)\in\mathcal{A} \\ m<\deg(I)\le m+2,\, k\in\mathbb{N}}} \left\| V_{i_1}\cdots V_{i_k} f \right\|_\infty, \tag{1.3}$$

i. e. the remainder is of order $\mathcal{O}\big(t^{\frac{m+1}{2}}\big)$ for $t \le 1$.

Just as the Taylor formula in numerical analysis, the stochastic Taylor expansion is the starting point of stochastic numerical analysis, see [3] for details on stochastic Taylor expansions and many methods originating from it.

Ignoring the remainder term, the only stochastic ingredients of the stochastic Taylor expansion are the iterated Stratonovich integrals of Brownian motion and a better understanding of them can give ideas for interesting new methods for numerical treatment of SDEs, as we shall see later. As a first step, we "encode" the iterated Stratonovich integrals of order up to m as a random variable with values in an appropriate algebra.

Before we go back to iterated Stratonovich integrals, we need to study this algebra in some detail, in order to exploit its properties later on.

2. Free Nilpotent Lie Groups

Let $\mathbb{A}_{d,1}$ denote the space of all noncommutative polynomials in $e_0,\dots,e_d$. $\mathbb{A}_{d,1}$ is the free associative algebra with unit generated by $e_0,\dots,e_d$ over the field of real numbers. We define a degree function on $\mathbb{A}_{d,1}$ by $\deg(e_{i_1}\cdots e_{i_k}) = \deg((i_1,\dots,i_k))$, i. e. e_0 hast twice the weight of the other generators. As usual, the degree of a polynomial – i. e. of an element of $\mathbb{A}_{d,1}$ – is defined as the maximum of the degrees of the respective monomials.

Definition 2.1. The *free step-m nilpotent associative real algebra with unit in d generators of degree* 1 *and one generator of degree* 2 is the space of all noncommutative polynomials of degree less or equal m. We denote it by $\mathbb{A}_{d,1}^m$.

Remark 2.1. Algebraically, $\mathbb{A}_{d,1}^m$ is $\mathbb{A}_{d,1}$ factorized by the ideal generated by the monomials of degree greater than m. Thus, it also has the structure of an algebra, $e_{i_1}\cdots e_{i_k} e_{j_1}\cdots e_{j_l} = 0$ for $\deg(i_1,\dots,i_k,j_1,\dots,j_l) > m$. Note that we often simply write "algebra" instead of "associative real algebra with unit".

Remark 2.2. The "1" in the symbols $\mathbb{A}_{d,1}$ and $\mathbb{A}_{d,1}^m$ stands for e_0, which we need for the representation of $B_t^0 = t$. If there is no drift, i. e. if $V_0 = 0$ in (1.1), then we do not need to keep track of the of the "dt" terms, and thus we only need to consider the (step-m nilpotent) algebras generated by $e_1,\dots,e_d$, denoted by $\mathbb{A}_d^m$ and $\mathbb{A}_d$, respectively.

The degree function deg induces a grading on the algebra: if W_k denotes the linear span of all monomials of degree k, then

$$\mathbb{A}_{d,1}^m = \mathbb{R} \oplus W_1 \oplus \cdots \oplus W_m, \tag{2.1}$$

where $\mathbb{R} \simeq W_0 = \langle\{1\}\rangle$, 1 referring to the unit element of the algebra. Here, $\langle A\rangle$ denotes the linear span of a set A. Indeed, $W_k \cdot W_l \subset W_{k+l}$, $k, l \in \mathbb{N}$, keeping in mind that $W_k = \{0\}$ for $k > m$. According to the grading, we will write $x = x_0 + \cdots + x_m$ for $x \in \mathbb{A}_{d,1}^m$, where $x_i \in W_i$ is the projection of x onto W_i, $i = 0, \ldots, m$. For $t \in \mathbb{R}$, we define the *canonical dilatation* $\Delta_t : \mathbb{A}_{d,1}^m \to \mathbb{A}_{d,1}^m$ by

$$\Delta_t(x) = x_0 + tx_1 + t^2 x_2 + \cdots + t^m x_m. \tag{2.2}$$

Note that Δ_t is an algebra-homomorphism.

We define the exponential function on $\mathbb{A}_{d,1}^m$ using the usual power series definition, i. e.

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

Remark 2.3. Note that $\mathbb{A}_{d,1}^m$ is a finite-dimensional vector space, thus the exponential function is well defined – and smooth – using the natural topology on $\mathbb{A}_{d,1}^m$. If we used the same definition in the non-truncated algebra $\mathbb{A}_{d,1}$, as in [7], we would get convergence problems and need to work in the completion of $\mathbb{A}_{d,1}$ with respect to some non-elementary topology. Finite dimensionality of all the spaces involved is one of the benefits of working in the nilpotent setting from the beginning.

The logarithm is defined for $x \in \mathbb{A}_{d,1}^m$ with $x_0 > 0$ by

$$\log(x) = \log(x_0) + \sum_{k=1}^{m} \frac{(-1)^k}{k} \Big(\frac{x - x_0}{x_0}\Big)^k,$$

with $\log(x_0)$ being the logarithm of x_0 interpreted as a real number. Note that we may truncate the power series of the logarithm due to nilpotency: $(x - x_0)^{m+1} = 0$. Thus, the logarithm is a polynomial (for fixed x_0).

Any algebra carries the structure of a Lie algebra with respect to the commutator $[x, y] = xy - yx$. We now consider the Lie algebra generated by $e_0, \ldots, e_d$:

Definition 2.2. The *free step-m nilpotent Lie algebra with d generators of degree* 1 *and one generator of degree* 2 is denoted by $\mathfrak{g}_{d,1}^m$, i. e.

$$\mathfrak{g}_{d,1}^m = \Big\langle \big\{ e_i, [e_i, e_j], [e_i, [e_j, e_k]], \ldots \mid i, j, k \in \{0, \ldots, d\} \big\} \Big\rangle.$$

The *free step-m nilpotent Lie group* $G_{d,1}^m$ is defined as the exponential image of the Lie algebra $\mathfrak{g}_{d,1}^m$.

$G_{d,1}^m$, indeed, is a Lie group and $\mathfrak{g}_{d,1}^m$ is its Lie algebra, as can be seen using the *Baker-Campbell-Hausdorff formula*

$$\exp(y)\exp(z) = \exp\Big(y + z + \frac{1}{2}[y, z] + \frac{1}{12}([y, [y, z]] - [z, [z, y]]) + \cdots\Big), \tag{2.3}$$

for $y, z \in \mathfrak{g}_{d,1}^m$. Note that $\mathfrak{g}_{d,1}^m$ inherits the grading of the algebra via $U_k = \mathfrak{g}_{d,1}^m \cap W_k$, $k = 1, \ldots, m$, hence

$$\mathfrak{g}_{d,1}^m = U_1 \oplus \cdots \oplus U_m.$$

By definition, $z_0 = 0$ for $z \in \mathfrak{g}_{d,1}^m$ and $x_0 = 1$ for $x \in G_{d,1}^m$. Note that $\exp : \mathfrak{g}_{d,1}^m \to G_{d,1}^m$ is smooth and bijective and, hence, the logarithm – being its inverse – is a global chart for the manifold $G_{d,1}^m$!

Example 2.1. The simplest non-trivial example is the case $d = 2$ and $m = 2$ without drift. That is, we consider $\mathbb{A}_2^2$, the space spanned by 1, e_1, e_2 and e_1^2, e_1e_2, e_2e_1, e_2^2. The corresponding Lie algebra $\mathfrak{g}_2^2$ is spanned by e_1, e_2 and $[e_1, e_2]$. The Lie group G_2^2 is called Heisenberg group.

The Heisenberg group has the following representation as a matrix group:

$$G_2^2 = \left\{ \begin{pmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix} \;\middle|\; a, b, c \in \mathbb{R} \right\}$$

with the usual matrix multiplication as group operation. The Lie algebra $\mathfrak{g}_2^2$ – i. e. the tangent space at the unit element $I_3 = \left(\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right)$ – is then given by

$$\mathfrak{g}_2^2 = \left\{ \begin{pmatrix} 0 & x & z \\ 0 & 0 & y \\ 0 & 0 & 0 \end{pmatrix} \;\middle|\; x, y, z \in \mathbb{R} \right\},$$

where the Lie bracket is defined as the usual commutator of matrices. We can think of the Heisenberg group as $\mathbb{R}^3$ with the (noncommutative) multiplication $(x_1, x_2, x_3) \star (y_1, y_2, y_3) = (x_1 + y_1, x_2 + y_2, x_3 + y_3 + x_1y_2)$.

Note that, with $e_1 = \left(\begin{smallmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{smallmatrix}\right)$ and $e_2 = \left(\begin{smallmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{smallmatrix}\right)$ we have $[e_1, e_2] = \left(\begin{smallmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{smallmatrix}\right)$, which verifies that $\mathfrak{g}_2^2$ is spanned by e_1, e_2 and $[e_1, e_2]$. We will come back to this later.

3. Iterated Stratonovich Integrals in $G_{d,1}^m$

Let X_t^x, $x \in \mathbb{A}_{d,1}^m$, $t \geq 0$, denote the stochastic process

$$X_t^x = x \sum_{\substack{I \in \mathcal{A} \\ \deg(I) \leq m}} B_t^I e_I, \tag{3.1}$$

where $e_I = e_{i_1} \cdots e_{i_k}$ for a multi-index $I = (i_1, \ldots, i_k) \in \mathcal{A}$, $k \geq 1$, and $e_\emptyset = 1 \in \mathbb{A}_{d,1}^m$. As long as x is invertible, e. g. for $x = 1$, X_t^x encodes the iterated Stratonovich integrals of degree up to m as a stochastic process in $\mathbb{A}_{d,1}^m$, since the collection of all e_Is as in (3.1) forms a linear basis of $\mathbb{A}_{d,1}^m$. X_t^x is solution to an $\mathbb{A}_{d,1}^m$-valued SDE driven by $(B_t^0, \ldots, B_t^d)$. Indeed,

$$\begin{aligned} dX_t^x &= x \left(\sum_{\substack{I \in \mathcal{A} \\ \deg(I) \leq m-2}} B_t^I \, dt \, e_I \, e_0 + \sum_{i=1}^d \sum_{\substack{I \in \mathcal{A} \\ \deg(I) \leq m-1}} B_t^I \circ dB_t^i \, e_I \, e_i \right) \\ &= x \sum_{\substack{I \in \mathcal{A} \\ \deg(I) \leq m}} B_t^I e_I e_0 \, dt + x \sum_{i=1}^d \sum_{\substack{I \in \mathcal{A} \\ \deg(I) \leq m}} B_t^I e_I e_i \circ dB_t^i, \end{aligned}$$

where we used the nilpotency of the algebra in the second line. Consequently, X_t^x is the unique solution of the SDE

$$\begin{cases} dX_t^x = X_t^x e_0 dt + \sum_{i=1}^d X_t^x e_i \circ dB_t^i \\ X_0^x = x \end{cases}, \tag{3.2}$$

where $x \in \mathbb{A}_{d,1}^m$. We denote its driving vector fields by $D_i(x) = xe_i$, $i = 0, \dots, d$. Depending on the situation, we regard $D_0, \dots, D_d$ as vector fields defined on either $\mathbb{A}_{d,1}^m$ or $G_{d,1}^m$.

Remark 3.1. The well known identity-in-law $B_t =^{\mathcal{L}} \sqrt{t} B_1$ and the corresponding identities-in-law for the iterated Stratonovich integrals can be nicely expressed using the canonical dilatation as $X_t^1 =^{\mathcal{L}} \Delta_{\sqrt{t}} X_1^1$.

Proposition 3.1. *Given an initial value $x \in G_{d,1}^m$, $X_t^x \in G_{d,1}^m$ a. s. for all $t \geq 0$.*

Proof. Since $X_t^x = xX_t^1$, we only need to consider the initial value $x = 1$. By (3.2), X_t^1 solves the Stratonovich-SDE driven by the vector fields $D_i(x) = xe_i$, $i = 0, 1, \dots, d$. Note that these vector fields are tangent to the Lie group $G_{d,1}^m$ and left-invariant. Consequently, they belong to the Lie algebra $\mathfrak{g}_{d,1}^m$ interpreted as the space of all left-invariant vector fields tangential to the Lie group. (This statement is true since $T_1 G_{d,1}^m = \mathfrak{g}_{d,1}^m$ and $T_x G_{d,1}^m = x\mathfrak{g}_{d,1}^m = \{xz \mid z \in \mathfrak{g}_{d,1}^m\}$. Thus, the driving vector fields are tangent to the Lie group, and left-invariance can be checked.) This already implies that $G_{d,1}^m$ is locally invariant with respect to the SDE (3.2), i. e. $\forall x \in G_{d,1}^m$ there is a strictly positive stopping time τ_x such that $X_t^x \in G_{d,1}^m$ $\forall t \leq \tau_x$ a. s.

$G_{d,1}^m$ is a closed subset of $\mathbb{A}_{d,1}^m$, since it is the pre-image of the closed set $\mathfrak{g}_{d,1}^m \subset \mathbb{A}_{d,1}^m$ under the continuous map log. Let τ be the first hitting time of $\mathbb{A}_{d,1}^m \setminus G_{d,1}^m$. Then $X_\tau^x \in G_{d,1}^m$ on $\{\tau < \infty\}$ and, consequently, local invariance implies existence of a positive stopping time $\tilde{\tau}$ such that X_t^x remains in $G_{d,1}^m$ up to time $\tau + \tilde{\tau}$, in contradiction to the construction of τ. □

Hence, X_t^x, $x \in G_{d,1}^m$, really takes its values in the Lie group $G_{d,1}^m$, which is a much smaller dimensional manifold – e. g. $\dim \mathfrak{g}_{2,1}^3 = 8$ whereas $\dim \mathbb{A}_{2,1}^3 = 20$, i. e. the iterated integrals of two Brownian motions up to order 3 evolve on an 8-dimensional submanifold of the 20-dimensional space $\mathbb{A}_{2,1}^3$. On the other hand, $G_{d,1}^m$ is not a linear space and often it is desirable to work in a vector space instead of a manifold. Therefore, we use the global chart and define a stochastic process

$$Z_t = \log(X_t^1), \quad t \geq 0, \tag{3.3}$$

on $\mathfrak{g}_{d,1}^m$, which contains all the information of X_t^1.

Proposition 3.2 (Chen-Strichartz formula)**.** *Let $\mathfrak{S}_k$ denote the symmetric group of k elements. For $\sigma \in \mathfrak{S}_k$ let $e(\sigma) = \#\{j \in \{1, \dots, k-1\} \mid \sigma(j) > \sigma(j+1)\}$ be the number of inversions of σ. Then*

$$Z_t = \log(X_t^1) = \sum_{\substack{I \in \mathcal{A} \setminus \{\emptyset\} \\ \deg(I) \leq m}} \Lambda_t^I e_{[I]},$$

where, for a multi-index $I = (i_1, \dots, i_k) \in \mathcal{A}\backslash\{\emptyset\}$, $e_{[I]} = [e_{i_1}, [e_{i_2}, \dots, [e_{i_{k-1}}, e_{i_k}]\dots]]$ *and*

$$\Lambda_t^{(i_1,\dots,i_k)} = \sum_{\sigma\in\mathfrak{S}_k} \frac{(-1)^{e(\sigma)}}{k^2\binom{k-1}{e(\sigma)}} B_t^{\sigma^{-1}(i_1,\dots,i_k)}$$

with $\sigma(i_1, \dots, i_k) = (i_{\sigma(1)}, \dots, i_{\sigma(k)})$ *for* $\sigma \in \mathfrak{S}_k$.

Proof. The Chen-Strichartz formula is Theorem 1.1 in [1]. We only need to apply the projection $\mathbb{A}_{d,1} \to \mathbb{A}_{d,1}^m$ to the formula as given in [1] and note that the drift, which is included in our setting, does not affect the formula in principle. □

Remark 3.2. The Chen-Strichartz formula roughly says that Z_t is a linear function of X_t^1. Note, however, that the set $\{e_{[I]} \mid I \in \mathcal{A} \setminus \{\emptyset\}\}$ is not a basis of $\mathfrak{g}_{d,1}^m$, e. g. $[e_1, e_2] = -[e_2, e_1]$ and both terms appear. Thus, in a second step one would need to expand the Chen-Strichartz formula with respect to a basis of $\mathfrak{g}_{d,1}^m$. One particular class of basis of $\mathfrak{g}_{d,1}^m$ is the class of *Hall basis*, for which a well-developed theory including efficient construction algorithms exists. We do not delve into details but refer to [10] for a thorough treatment of this topic.

Example 3.1 (Continuation of Example 2.1)**.** In the case of the Heisenberg group, a Hall basis of $\mathfrak{g}_2^2$ is given by $\{e_1, e_2, [e_1, e_2]\}$ and we have $Z_t = B_t^1 e_1 + B_t^2 e_2 + A_t[e_1, e_2]$, where

$$A_t = \frac{1}{2}\int_0^t B_s^1 \circ dB_s^2 - \frac{1}{2}\int_0^t B_s^2 \circ dB_s^1$$

denotes Lévy's area.

The setting $\mathbb{A}_{d,1}^m$ provides a nice way to write the expected value of iterated Stratonovich integrals.

Proposition 3.3. *For* $t \geq 0$ *and* $x \in \mathbb{A}_{d,1}^m$ *we have*

$$E\big(X_t^x\big) = x\exp\Big(te_0 + \frac{t}{2}\sum_{i=1}^d e_i^2\Big).$$

Note that $E(X_t^x) \notin G_{d,1}^m$ *since its logarithm is not in the Lie algebra.*

Proof. Proposition 3.3 is not surprising considering the SDE (3.2) for X_t^x. Indeed, the infinitesimal generator of X_t^x is given by $L = D_0 + \frac{1}{2}\sum_{i=1}^d D_i^2$, again interpreting the vector fields $D_i(x) = xe_i$ as first-order differential operators. Thus, $u(t, x) = E(X_t^x)$ satisfies the Kolmogorov backward equation

$$\frac{\partial}{\partial t}u(t,x) = Lu(t,x)$$

with initial condition $u(0, x) = x$. Formally, this can be expressed by $u(t, x) = e^{tL}(\mathrm{Id})(x)$.

Now let $v(t, x)$ denote the right hand side of the formula in the proposition statement. Then,

$$\frac{\partial v}{\partial t}(t,x) = x\exp\Big(te_0 + \frac{t}{2}\sum_{i=1}^d e_i^2\Big)\Big(e_0 + \frac{1}{2}\sum_{i=1}^d e_i^2\Big) = Lv(t,x)$$

and the result follows by uniqueness of the solution to the Kolmogorov backward equation. □

Of course, Proposition 3.3 is not an interesting result of its own, it is rather a convenient way of writing the easily obtained recursive formulas for the expected value of an iterated Stratonovich integral. Nevertheless, it is the starting point to the applications presented in the next two sections.

4. Cubature on Wiener Space

We start with the classical concept of cubature formulas.

Definition 4.1. Given a positive Borel measure μ on $\mathbb{R}^N$ with support $\operatorname{supp}(\mu)$ and $m \in \mathbb{N}$. The points $x_1, \dots, x_l \in \operatorname{supp}(\mu)$ and the weights $\lambda_1, \dots, \lambda_l > 0$, $l \in \mathbb{N}$, form a *cubature formula of degree m for μ* if for all polynomials p on $\mathbb{R}^N$ with degree less or equal m the following formula holds true:

$$\int_{\mathbb{R}^N} p(x)\mu(dx) = \sum_{i=1}^{l} \lambda_i p(x_i).$$

Cubature formulas of degree m exist for any Borel measure with finite moments of order up to m, a statement known as *Chakalov's theorem*, see [2]. More precisely, Chakalov's theorem implies existence of cubature formulas with size l smaller than or equal to the dimension of the space of polynomials on $\mathbb{R}^N$ of order up to m. Note, however, that even in the finite-dimensional case cubature formulas are difficult to construct, especially in higher dimensions.

Recall that the *Wiener space* is the space $C_0([0,\infty[;\mathbb{R}^d)$ of continuous functions $f : [0,\infty[\to \mathbb{R}^d$ such that $f(0) = 0$, equipped with the Borel σ-field and the Wiener measure P, i. e. the law of d-dimensional Brownian motion. A direct generalization of the notion of a cubature formula to Wiener space does not make sense. Indeed, cubature weights $\lambda_1, \dots, \lambda_l$ and paths $\omega_1, \dots, \omega_l \in C_0([0,\infty[;\mathbb{R}^d)$, $l \in \mathbb{N}$, would have to verify the following: for any collection of continuous linear functionals $\mu_1, \dots, \mu_N$, $N \in \mathbb{N}$, on $C_0([0,\infty[;\mathbb{R}^d)$ and any polynomial p on $\mathbb{R}^N$ of order up to m

$$\int\limits_{C_0([0,\infty[;\mathbb{R}^d)} p\big(\mu_1(\omega), \dots, \mu_N(\omega)\big) P(d\omega) = \sum_{i=1}^{l} \lambda_i p\big(\mu_1(\omega_i), \dots, \mu_N(\omega_i)\big).$$

This cannot work, because there are far too many linearly independent linear functionals.

The previous generalization of cubature formulas to Wiener space does not cover some of the most interesting functionals on Wiener space such as Lévy's area $A_t^{i,j}$, $i, j = 1, \dots, d$. Note that Lévy's area is not linear, but not even continuous with respect to the topology on Wiener space, i. e. uniform topology, even if the path is regular enough such that Lévy's area is well-defined using Riemann-Stieltjes integrals, see [6]. Thus, even though the above condition is far to strong from one point of view, it is too weak from another point of view, since it fails to cover some of the most important functionals.

Instead, we go back to the stochastic Taylor expansion, which suggests the following substitute:

Definition 4.2. Fix $T > 0$ and $m \in \mathbb{N}$. Weights $\lambda_1, \dots, \lambda_l > 0$ and continuous functions $\omega_j : [0,T] \to \mathbb{R}^d$, $j = 1, \dots, l$, of bounded variation define a *cubature formula on Wiener space of degree m* if for all multi-indices $I = (i_1, \dots, i_k) \in \mathcal{A}$

with $\deg(I) \leq m$

$$E(B_T^I) = \sum_{j=1}^{l} \lambda_j \int\limits_{0<t_1<\cdots<t_k<T} d\omega_j^{i_1}(t_1) d\omega_j^{i_2}(t_2) \cdots d\omega_j^{i_k}(t_k)$$

with $\omega_j^0(t) = t$ for $j = 1, \ldots, l$ and $t \in [0, T]$. Note that the iterated integrals on the right hand side are well-defined as Riemann-Stieltjes integrals.

The idea is that – as mentioned regarding the stochastic Taylor expansion – iterated Stratonovich integrals play the rôle of polynomials in the classical case. Note that the integrals on the right hand side of the formula can be interpreted as $B_T^I(\omega)$ evaluated at $\omega = \omega_j$. Definition 4.2 is related to Definition 4.1 applied to Wiener space in the following way: let μ_T^i denote the evaluation functional at time T on Wiener space, i. e. $\mu_T^i(\omega) = \omega^i(T)$, $i = 1, \ldots, d$. Then, for any polynomial p on $\mathbb{R}^d$ of order up to m, Itô's formula implies that

$$\int\limits_{C_0([0,\infty[;\mathbb{R}^d)} p(\mu_T(\omega)) P(d\omega) = E\big(p(B_T)\big) = \sum_{j=1}^{l} \lambda_j p(\omega_j(T)) = \sum_{j=1}^{d} \lambda_j p(\mu_T(\omega_j)),$$

given a cubature formula on Wiener space as in Definition 4.2. This is the required equality for a classical cubature formula in the sense of Definition 4.1 on Wiener space, but only for the d linear functionals $\mu_T^1, \ldots, \mu_T^d$, when the equality should hold for *all* linear functionals on Wiener space. This shows that c. f. w., indeed, is a much weaker concept than the classical concept of a cubature formula generalized to the Wiener space in this respect. On the other hand, cubature on Wiener space does cover Lévy's area and the other iterated Stratonovich integrals.

When we talk about cubature formulas on Wiener space, we always mean cubature formulas in the sense of Definition 4.2. We never mean a direct generalization of classical cubature formulas in the sense of Definition 4.1, neither the one indicated above nor any other possible one.

Remark 4.1. A related concept by Kusuoka [4] and [5] uses *m-moment similar families of random variables*, i. e. families of random variables $Z_I \in \bigcap_{1 \leq p < \infty} L^p(\Omega, \mathcal{F}, P)$, $I \in \mathcal{A}$, such that $Z_\emptyset = 1$ and

$$E\big(Z_{I_1} \cdots Z_{I_k}\big) = E\big(B_1^{I_1} \cdots B_1^{I_k}\big),$$

for all $I_1, \ldots, I_k \in \mathcal{A}$ and $k \in \mathbb{N}$ with $\deg(I_1) + \cdots + \deg(I_k) \leq m$.

In order to compare the concepts of cubature on Wiener space and moment similar random variables, we interpret Definition 4.2 in a slightly different way. Let $\overline{\omega}$ be a random variable taking values in the Wiener space such that $\overline{\omega} = \omega_j$ with probability λ_j, $j = 1, \ldots, l$. Then the moment condition in Definition 4.2 reads

$$E(B_T^I) = \sum_{j=1}^{l} \lambda_j B_T^I(\omega_j) = E_{\overline{\omega}}\big(B_T^I(\overline{\omega})\big),$$

for all $I \in \mathcal{A}$ with $\deg(I) \leq m$. Thus, the moment condition for the method of moment similar random variables is much stronger for the same order m, since the moment condition for cubature on Wiener space only prescribes the first moments. Thus, m-moment similar random variables are more difficult to find and typically more complicated.

Moreover, the method of m-similar random variables neglects the geometry of the problem. For example, the support of the law of the vector $(Z_I : \deg(I) \leq m)$ does not need to lie in the support of the law of the vector of iterated Stratonovich integrals of order up to m. This is an undesirable property, see [12].

Remark 4.2. Using Remark 3.1, it suffices to construct a cubature formula for time $T = 1$, since cubature formulas for any other time $T > 0$ can then be constructed by rescaling.

Before proving existence of cubature formulas on Wiener space, we sketch the intended application, namely weak approximation of the SDE (1.1), see [7] for details. Given a function $f : \mathbb{R}^n \to \mathbb{R}$ smooth enough, stochastic Taylor expansion implies

$$\begin{aligned} E\big(f(Y_T^y)\big) &= \sum_{\substack{I=(i_1,\dots,i_k)\in\mathcal{A}\\ \deg(I)\leq m,\, k\in\mathbb{N}}} V_{i_1}\cdots V_{i_k} f(y) E\big(B_T^I\big) + \mathcal{O}\big(T^{\frac{m+1}{2}}\big) \\ &= \sum_{j=1}^{l} \lambda_j \sum_{\substack{I=(i_1,\dots,i_k)\in\mathcal{A}\\ \deg(I)\leq m,\, k\in\mathbb{N}}} V_{i_1}\cdots V_{i_k} f(y) \omega_j^I(T) + \mathcal{O}\big(T^{\frac{m+1}{2}}\big), \end{aligned} \tag{4.1}$$

where λ_j, ω_j, $j = 1, \dots, l$, stem from a cubature formula of order m at time T and

$$\omega_j^{(i_1,\dots,i_k)}(t) = \int\limits_{0<t_1<\cdots<t_k<t} d\omega_j^{i_1}(t_1)\cdots d\omega_j^{i_k}(t_k)$$

for a multi-index $(i_1, \dots, i_k) \in \mathcal{A}$, $k \in \mathbb{N}$, $j = 1, \dots, l$. By ordinary, deterministic Taylor expansion of the solution $Y_t^y(\omega_j)$ of the ODE

$$dY_t^y(\omega_j) = V_0\big(Y_t^y(\omega_j)\big)dt + \sum_{i=1}^{d} V_i\big(Y_t^y(\omega_j)\big)d\omega_j^i(t), \tag{4.2}$$

with initial condition $Y_t^y(\omega_j) = y$, we get

$$f\big(Y_T^y(\omega_j)\big) = \sum_{\substack{I=(i_1,\dots,i_k)\in\mathcal{A}\\ \deg(I)\leq m,\, k\in\mathbb{N}}} V_{i_1}\cdots V_{i_k} f(y) \omega_j^I(T) + \mathcal{O}\big(T^{m+1}\big). \tag{4.3}$$

Plugging (4.3) into (4.1) yields

$$E\big(f(Y_T^y)\big) = \sum_{j=1}^{l} \lambda_j f\big(Y_T^y(\omega_j)\big) + \mathcal{O}\big(T^{\frac{m+1}{2}}\big), \tag{4.4}$$

i. e. we approximate the solution of the Kolmogorov backward PDE by a convex combination of f evaluated at the solutions of the ODEs (4.2) driven by l paths of bounded variation. This is the one-step cubature on Wiener space scheme.

The multi-step scheme works as follows – for simplicity, we only write down the two-step scheme, the multi-step scheme being a straightforward generalization:

(1) Fix $0 < s < T$. As mentioned above, we get cubature paths for s and $T - s$ by rescaling the cubature paths for T, we use the symbols $\omega_j^{(t)}$, $j = 1, \dots, l$, for the paths of the cubature formula on Wiener space at time t.
(2) Calculate the solutions $Y_s^y(\omega_j^{(s)}) = y_{s,j}$ of (4.2), $j = 1, \dots, l$.
(3) Calculate the solutions $Y_{T-s}^{y_{s,j}}(\omega_i^{(T-s)}) = y_{T,j,i}$ for $i, j = 1, \dots, d$.

(4) Approximate the solution of the PDE by

$$E\big(f(Y_T^y)\big) \approx \sum_{i,j=1}^{l} \lambda_i \lambda_j f(y_{T,i,j}). \tag{4.5}$$

This already clarifies the current weakness of the method, since the effort – the number of ODEs – increases exponentially with the time discretization. In general, subdividing $[0,T]$ using a mesh $0 = s_0 < s_1 < \cdots < s_k = T$ gives an error of order

$$\text{const } T \max_{j=1,\dots,k} (\Delta s_j)^{\frac{m-1}{2}},$$

$\Delta s_j = s_j - s_{j-1}$, $j = 1,\dots,k$, but one needs to solve l^k ODEs. The above error estimate is valid for a function $f \in C_b^{m+2}(\mathbb{R}^n)$, since we need the estimate for the error term (1.3). The techniques of [4] allow, however, to extend the result to uniformly Lipschitz continuous functions f, provided one uses the special grid $s_i^{(\gamma)} = \frac{i^\gamma}{k^\gamma} T$, $i = 0,\dots,k$, $\gamma > m-1$. More precisely,

$$\Big| E\big(f(Y_T)\big) - \sum_{j_1,\dots,j_k=1}^{l} \lambda_{j_1} \cdots \lambda_{j_k} f(y_{T,j_1,\dots,j_l}) \Big| \le \text{const } \frac{\|\nabla f\|_\infty}{k^{(m-1)/2}},$$

where $y_{T,j_1,\dots,j_l}$ is the value of the solution to the ODE (4.2) along the path ω_{j_1} up to time $s_1^{(\gamma)}$, then along ω_{j_2} up to time $s_2^{(\gamma)}$, and so on.

One way out of the dilemma is by identifying two points $y_{s_i,j_1,\dots,j_i}$ and $y_{s_i,j_1',\dots,j_i'}$, $i \in \{1,\dots,l\}$, in the tree if they are "close enough". This needs to be done in a careful, systematic way, since the result is very sensitive with respect to changes in the tree, see [11].

For a different approach note that (4.5) can be interpreted as the expectation of a random variable taking the values $f(y_{T,i,j})$ with probability $\lambda_i \lambda_j$, $i,j = 1,\dots,l$. Instead of calculating all these l^2 – in general l^k – values, we can calculate this expected value by Monte Carlo simulation on the tree, i. e. we only need to calculate through some of the l^2 possibilities. See [9] for a method related to this idea.

Example 4.1. As with classical cubature methods, it is a non-trivial task to construct actual cubature paths for some given d and m, especially if one wants to find optimal ones, i. e. with the minimal number of paths necessary to achieve a cubature formula on Wiener space.

For the case $m = 3$, let $\lambda_1,\dots,\lambda_l$ be the weights and $z_1,\dots,z_l \in \mathbb{R}^d$ the nodes of a classical cubature formula of degree 3 for the standard Gaussian measure on $\mathbb{R}^d$, e. g. $\{z_1,\dots,z_l\} = \{+1,-1\}^d$ with equal weights 2^{-d} – which is far away from being optimal. Then, the paths $\omega_i : [0,1] \to \mathbb{R}^d$, $t \mapsto t z_i$, $i = 1,\dots,l$ together with the same weights as before give a cubature formula on Wiener space.

Example paths for $m > 3$ are more difficult to write down. In [7], piecewise linear cubature paths are constructed for $m = 5$ and $d = 2$ and it is shown how to construct piecewise linear cubature paths for $m = 5$ and arbitrary d. Still, the construction of cubature paths is an interesting area of research.

It is easy to translate the concept of cubature formulas on Wiener space to our algebraic setting: Paths and weights as in Definition 4.2 form a cubature formula

on Wiener space of degree m if

$$E\big(X_T^1\big) = \sum_{j=1}^{l} \lambda_j \Phi(\omega_j)_T^1, \tag{4.6}$$

where $\Phi(\omega)_t^1$ denotes the solution of the ODE (on $\mathbb{A}_{d,1}^m$)

$$dy_t = y_t e_0 dt + \sum_{i=1}^{d} y_t e_i d\omega^i(t) \tag{4.7}$$

with initial value $y_0 = 1$ for a continuous path of bounded variation ω. Note that $\Phi(\omega)_t^1 \in G_{d,1}^m$ for all $t \in [0,T]$.

For the proof of existence of a cubature formula, we proceed in two steps: First we show that the left hand side of (4.6) can be written as a convex combination of points in $G_{d,1}^m$, then we show that any group element is end point of the solution to (4.7) starting at $1 \in G_{d,1}^m$ for some continuous path of bounded variation.

Lemma 4.1. *For any $T > 0$, there are a natural number l, positive weights $\lambda_1, \dots, \lambda_l$ and group elements $x_1, \dots, x_l \in G_{d,1}^m$ such that*

$$E\big(X_T^1\big) = \exp\Big(Te_0 + \frac{T}{2}\sum_{i=1}^{d} e_i^2\Big) = \sum_{j=1}^{l} \lambda_j x_l.$$

Proof. Apply Chakalov's theorem of existence of a (classical) cubature formula of degree 1 to the law of X_T^1, a measure on the finite dimensional vector space $\mathbb{A}_{d,1}^m$ supported by the Lie group $G_{d,1}^m$. Note that Chakalov's theorem in this case is nothing else but the statement that the expected value of any random variable lies in the convex hull of its support, provided it exists, see [2], Theorem 1. □

For the second step we need the following special case of *Chow's theorem* from sub-Riemannian geometry, see [8].

Lemma 4.2. *Given a smooth N-dimensional manifold M. Assume that there are $d < N$ vector fields $D_1, \dots, D_d$ such that for all $x \in M$ the tangent space T_xM of M at x is generated by the vector fields $D_1(x), \dots, D_d(x)$ evaluated at x together with their iterated Lie brackets $[D_i, D_j](x)$, $[D_i, [D_j, D_k]](x)$, ... Then, for any two points $x, y \in M$ and any $T > 0$ there is a continuous path of bounded variation $\omega : [0,T] \to \mathbb{R}^d$ such that the solution of the differential equation*

$$dz_t = \sum_{i=1}^{d} D_i(z_t) d\omega^i(t),$$

with initial condition $z(0) = x$, satisfies $z(T) = y$. In other words: any two points on the manifold can be connected by an absolutely continuous horizontal *path.*

Theorem 4.1. *Cubature formulas on Wiener space exist for any time T and any natural number m.*

Proof. Take the points and weights given by Lemma 4.1. Note that the Lie group $G_{d,1}^m$ satisfies the conditions of Lemma 4.2 using the vector fields D_i defined in the previous section, $i = 0, \dots, d$. For any of the points x_j, $j = 1, \dots, l$, we get a continuous path of bounded variation $\omega_j : [0,T] \to \mathbb{R}^{d+1}$ with $x_j = \Phi(\omega_j)_T^1$. Note that we can choose the paths $\omega_j = (\omega_j^0 \dots, \omega_j^d)$ such that $\omega_j^0(t) = t$, for all $t \in [0,T]$, $j = 1, \dots, l$. □

5. Heat Kernel on G_d^m

In this section, we will mainly work in a setting without a drift, i. e. in the setting indicated in Remark 2.2. In this context, we work with a set $\mathcal{A}$ of multi-indices as before, but without any zeros showing up. Note that deg in this case is nothing else but the length of a multi-index or monomial. As in (3.1), we define a stochastic process on $\mathbb{A}_d^m$ using the above set of multi-indices. All the previous facts remain true with the obvious modifications.

Definition 5.1. The *heat kernels* on the Lie groups $G_{d,1}^m$ and G_d^m are the densities – with respect to the Haar measure on the Lie group – of the respective processes X_t^1. We denote them by $p_t(y)$, $y \in G_{d,1}^m$ or $y \in G_d^m$.

Remark 5.1. Note that $P(X_t^x \in dy) = P(X_t^1 \in x^{-1}dy)$ by definition, thus p_t also contains all the information on the densities of X_t^x. For many situations, it is more convenient to study the heat kernel in the global chart, i. e. to study the density of the process $Z_t = \log(X_t^1)$ on $\mathfrak{g}_{d,1}^m$ or $\mathfrak{g}_d^m$ and denote it by the same symbol $p_t(z)$, $z \in \mathfrak{g}_{d,1}^m$ or $z \in \mathfrak{g}_d^m$, respectively.

Remark 5.2. We call the densities p_t on $G_{d,1}^m$ and G_d^m heat kernels, since they are the fundamental solutions of the heat equations

$$\frac{\partial}{\partial t}u(t,x) = Lu(t,x), \quad x \in G_{d,l}^m, \quad l \in \{0,1\}$$

with respect to the sub-Laplacians $L = D_0 + \frac{1}{2}\sum_{i=1}^d D_i^2$ (case $l = 1$) and $L = \frac{1}{2}\sum_{i=1}^d D_i^2$ (case $l = 0$), respectively.

Proposition 5.1. *(i) In the situation without drift, the heat kernel p_t exists as a Schwartz function.*
(ii) In the situation with drift, the heat kernel only exists in a distributional sense. Let $\langle e_0 \rangle$ be the subspace of $\mathfrak{g}_{d,1}^m$ spanned by e_0. Then the factor process Z_t on $\mathfrak{g}_{d,1}^m/\langle e_0 \rangle$ has a density.

Proof. The first part is a consequence of *Hörmander's theorem*, since X_t^1 is driven by the vector fields $D_i(x)$, $x \in G_d^m$, $i = 1, \dots, d$, which Lie-generate the tangent space $\mathfrak{g}_d^m$.

The heat kernel on $\mathfrak{g}_{d,1}^m$ – and consequently on $G_{d,1}^m$ – cannot exist as a function since the e_0-component of Z_t is deterministically equal to t. Proceed as before using Hörmander's theorem. □

Proposition 5.1 shows that we can use the heat kernel also in the situation with a drift, but we need to modify our problem by factorizing out the direction e_0. This is often possible, but for simplicity, we will restrain ourselves to the unproblematic case without drift in this text.

The existence proof for the heat kernel is unconstructive, so we need some other idea of how to approximate the heat kernel. If we try some kind of polynomial expansion, we need to calculate

$$\int_{\mathfrak{g}_d^m} z^\alpha p_t(z)dz = E\big(Z_t^\alpha\big) \tag{5.1}$$

for multi-indices $\alpha \in \mathbb{N}^N$, $N = \dim \mathfrak{g}_d^m$. But, by the Chen-Strichartz formula, see Proposition 3.2, (5.1) is the expected value of a polynomial in X_t^1, i. e. in iterated

Stratonovich integrals of order up to m. Calculation of (5.1) using Itô's formula is possible, but costly. Proposition 3.3 does not directly help us, because it only gives the expected values of linear functionals on X_t^1. There is, however, a method allowing to use the corresponding formula for an $\widetilde{m} > m$ large enough by factorizing out a particular ideal of $\mathbb{A}_d^{\widetilde{m}}$. Since this method requires quite some algebra, we do not go into details here.

We propose expanding the heat kernel with respect to Hermite polynomials as made precise below. Please note that our work on this subject is still preliminary, so there are several open questions, especially concerning efficiency and numerical practicality. Hermite expansion seems to be preferable since

(1) it harmonizes well with Schwartz functions: given a Schwartz function f on $\mathbb{R}^N$ and denote the Hermite functions – see below – with ϕ_α, $\alpha \in \mathbb{N}^N$, and the corresponding Fourier coefficients with a^α, i. e. $a^\alpha = \langle f, \phi_\alpha \rangle_{L^2(\mathbb{R}^N, dz)}$. Then $\sum_\alpha a^\alpha \phi_\alpha$ converges to f not only in L^2, but even in the topology of the Schwartz space.
(2) Note that p_t is in general not real analytic, see [15]. Thus, the straightforward approach of approximating p_t using a power series must fail in the sense that the series cannot converge uniformly to the function p_t. This makes (1) even more remarkable.
(3) The Gaussian measure is easy to work with. Moreover, the components of the Brownian motion B_t also show up as components of Z_t, thus it seems to be a good starting point to work with a reference measure having the correct marginals in these components.

Let $N = \dim \mathfrak{g}_d^m$ and fix a homogeneous basis $f_1, \dots, f_N$ of $\mathfrak{g}_d^m$, i. e. a basis such that f_i can be written as a linear combination of monomials of the same degree $\deg(f_i)$, $i = 1, \dots, N$. Note that the Hall basis has this property, see [10]. Let Σ_t be an N-dimensional symmetric, positive definite matrix such that $\sigma_t^{ij} = \rho^{ij} t^{(\deg(f_i)+\deg(f_j))/2}$, where σ_t^{ij} denotes the matrix entry corresponding to f_i and f_j and $\rho^{ij} \in \mathbb{R}$, $i, j = 1, \dots, N$. Denote the Gaussian density on $\mathfrak{g}_d^m$ with covariance matrix Σ_t by $r_t(z)$. We choose this reference measure, since it gives the right scaling. The Hermite polynomials $h_\alpha(t, \cdot)$ are the basis of $L^2(\mathfrak{g}_d^m, r_t(z)dz)$ constructed by orthonormalizing the monomials using the Gram-Schmidt method. The Hermite functions $\phi_\alpha(t, \cdot)$, $\alpha \in \mathbb{N}^N$, are given by $\phi_\alpha = \sqrt{r_t} h_\alpha$ and they form an orthonormal basis of $L^2(\mathfrak{g}_d^m, dz)$. Setting

$$a^\alpha = E\big(h_\alpha(t, Z_t)\big), \quad \alpha \in \mathbb{N}^N, \tag{5.2}$$

we get

$$p_t(\cdot) = \sum_{\alpha \in \mathbb{N}^N} a^\alpha \phi_\alpha(t, \cdot) \sqrt{r_t(\cdot)}, \tag{5.3}$$

where the limit in (5.3) is understood in $L^2(\mathfrak{g}_{d,0}^m, dz)$ and in the Schwartz space. Note that a^α is independent of t by the scaling.

Remark 5.3. (5.3) is a consequence of the approximation in $L^2(\mathfrak{g}_d^m, r_t(z)dz)$

$$\frac{p_t(z)}{r_t(z)} = \sum_{\alpha \in \mathbb{N}^N} a^\alpha h_\alpha(t, z).$$

Therefore, we need to choose Σ_t such that the left hand side is an element of $L^2(\mathfrak{g}_d^m, r_t(z)dz)$.

Remark 5.4. (5.2) can be calculated since h_α is a polynomial in z. It depends on the actual application, whether one should directly work with an approximation of the heat kernel as in (5.3). Note that (5.3) is not an expansion with respect to an orthonormal basis of L^2 – neither of $L^2(\mathfrak{g}_d^m, dz)$ nor of $L^2(\mathfrak{g}_d^m, r_t(z)dz)$. Thus, it may not be trivial to get (good) error estimates.

We sketch one possible applications of the expansion (5.3).

Example 5.1 (Milstein Scheme)**.** Let $0 = t_0 < t_1 < \cdots < t_K = T$ be a time discretization of the interval $[0, T]$ and let $\overline{Y}_K$ be the approximation of the solution Y_T to the SDE (1.1) using the Milstein scheme as presented in [3]. For the Milstein scheme, we need the iterated Stratonovich integrals of order up to 2, thus we work in G_d^2. Absence of a drift is not at all necessary, since X_t^1 on $G_{d,1}^2$ differs from X_t^1 on G_d^m only by the deterministic e_0-component. We write the (random) value $\overline{Y}_K$ as a function in the Brownian increments and the increments of Lévy's area, $\overline{Y}_K(Z_{\Delta t_1}, \ldots, Z_{\Delta t_K})$, where $\Delta t_k = t_k - t_{k-1}$, $k = 1, \ldots, K$, and $Z_t = \log(X_t^1) \in \mathfrak{g}_d^2$ is the vector of Brownian motions and Lévy's areas. All other terms of the Milstein scheme can be expressed by Z_t and t and are suppressed in the notation.

For a bounded, measurable function $f : \mathbb{R}^n \to \mathbb{R}$, we have

$$E\big(f(\overline{Y}_K)\big) = \int_{(\mathfrak{g}_d^2)^K} f\big(\overline{Y}_K(z_1, \ldots, z_K)\big) p_{\Delta t_1}(z_1) \cdots p_{\Delta t_K}(z_K) dz_1 \cdots dz_K. \tag{5.4}$$

We cannot directly use (5.4) by replacing the heat kernel with its approximation because we cannot do the – high-dimensional – Lebesgue integration. Instead, we approximate (5.4) using Monte Carlo simulation. We do not want to sample according to the true density p_t – since we want to provide an alternative method and sampling with respect to p_t is costly. Also, sampling with respect to the approximate densities is not possible, since they are no probability densities. We follow an "importance sampling" approach, by – again – using the Gaussian density r_t, i. e. we divide (5.4) by r_t and sample w. r. t. $\mathcal{N}(0, \Sigma_t)$:

$$E\big(f(Y_T)\big) \approx \frac{1}{M} \sum_{l=1}^{M} f\Big(\overline{Y}_K\big(z_1^{(l)}, \ldots, z_K^{(l)}\big)\Big) \frac{\widetilde{p}_{\Delta t_1}(z_1^{(l)})}{r_{\Delta t_1}(z_1^{(l)})} \cdots \frac{\widetilde{p}_{\Delta t_K}(z_K^{(l)})}{r_{\Delta t_K}(z_K^{(l)})}, \tag{5.5}$$

where $z_k^{(l)}$ are independent samples of $\mathcal{N}(0, \Sigma_{\Delta t_k})$, $k = 1, \ldots, K$, $l = 1, \ldots, M$, and $\widetilde{p}$ is an approximation to the heat kernel, i. e.

$$\frac{\widetilde{p}_t(z)}{r_t(z)} = \sum_{|\alpha| \le R} a^\alpha h_\alpha(t, z) \tag{5.6}$$

for some $R > 0$.

We already have preliminary numerical results, which show that some effort still needs to be invested into this method.

6. Outlook

Cubature on Wiener space has an interesting application in mathematical finance. Indeed, it gives a method of constructing non-trivial, highly efficient numerical differentiation schemes for calculation of Greeks, as shown in [13]. We roughly sketch the ideas.

This application is based on the method of calculating Greeks using Malliavin weights. Given an asset modeled by (1.1) – under the risk-less measure, with zero

interest rate – such that the value of an option with (bounded, measurable) pay-off $f : \mathbb{R}^n \to \mathbb{R}$ is given by $E\big(f(Y_T^y)\big)$. For the case of the Delta, the derivative with respect to the initial value of the underlying, the Malliavin weight (for a direction $v \in \mathbb{R}^n$, time T and initial value $y \in \mathbb{R}^n$) is a random variable π such that

$$\frac{\partial}{\partial \epsilon}\bigg|_{\epsilon=0} E\Big(f\big(Y_T^{y+\epsilon v}\big)\Big) = E\Big(f\big(Y_T^y\big)\pi\Big), \tag{6.1}$$

for all bounded, measurable functions f. Such a Malliavin weight can be shown to exist provided that Hörmander's condition is satisfied.

For simplicity, we assume that $V_0 = 0$. Let π_d^m be the Malliavin weight for the process X_T^1 defined in Section 3 at time T in direction $w \in \mathfrak{g}_d^m$. w is linked to $v \in \mathbb{R}^n$ in the following way: by Hörmander's condition, v can be written as a linear combination of the vector fields and their iterated Lie brackets evaluated at y. Now replace all occurrences of V_i by an e_i, $i = 1, \dots, d$, in this linear combination (and omit all references to y). This gives $w \in \mathfrak{g}_d^m$. Then, [13] shows that

$$\frac{\partial}{\partial \epsilon}\bigg|_{\epsilon=0} E\Big(f\big(Y_T^{y+\epsilon v}\big)\Big) = E\Big(f\big(Y_T^y\big)\pi_d^m\Big) + \mathcal{O}\Big(T^{\frac{m+1}{2}}\Big). \tag{6.2}$$

From this point, we can follow both approaches presented in Section 4 and Section 5. Following the cubature-approach yields the formula

$$\frac{\partial}{\partial \epsilon}\bigg|_{\epsilon=0} E\Big(f\big(Y_T^{y+\epsilon v}\big)\Big) = \sum_{j=1}^{l} \mu_j f\big(Y_T^y(\omega_j)\big) + \mathcal{O}\big(T^{\frac{m+1}{2}}\big)$$

for continuous paths $\omega_1, \dots, \omega_l$ of bounded variation satisfying a moment condition similar to the condition for cubature formulas on Wiener space. The weights $\mu_j \neq 0$ can be chosen to satisfy $\mu_1 + \cdots + \mu_l = 0$. (6.2) can be interpreted as some kind of numerical differentiation scheme.

The second approach is to directly use (6.2). For this we need to approximate the Malliavin weight π_d^m. This can be done using the heat kernel on G_d^m, since π_d^m is the logarithmic derivative in direction w of the heat kernel $p_t(y)$ at $y = X_T^1$. Thus, we can approximate π_d^m using the following ideas:

(1) We can define an approximation $\widetilde{p}_t$ of the heat kernel as in (5.6) and differentiate it.
(2) By Proposition 3.3, we can directly calculate the Hermite expansion of the logarithmic derivative of the heat kernel using

$$\frac{\partial}{\partial \epsilon}\bigg|_{\epsilon=0} E\Big(X_t^{1+\epsilon w}\Big) = w \exp\Big(\frac{t}{2}\sum_{i=1}^{d} e_i^2\Big)$$

and the same methods as presented above.

References

[1] Fabrice Baudoin, *An Introduction to the Geometry of Stochastic Flows*, Imperial College Press, 2004.

[2] Christian Bayer and Josef Teichmann, *The Proof of Tchakaloff's Theorem*, Proceedings of the AMS, to appear.

[3] Peter E. Kloeden, Eckhard Platen, *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, 1999.

[4] Shigeo Kusuoka, *Approximation of Expectation of Diffusion Processes and Mathematical Finance*, Advanced Studies in Pure Mathematics 31, Proceedings of the Final Taniguchi Symposium, 147–165, 2001.

[5] Shigeo Kusuoka, *Approximation of Expectation of Diffusion Processes based on Lie Algebra and Malliavin Calculus*, Advances in mathematical economics 6, 69–83, 2004.
[6] Terry Lyons,*Differential Equations Driven by Rough Signals*, Rev. Mat. Iberoam. 14, No.2, 215-310 (1998).
[7] Terry Lyons and Nicolas Victoir, *Cubature on Wiener Space*, Proc. R. Soc. Lond., Ser. A, Math. Phys. Eng. Sci. 460, No.2041, 169-198 (2004).
[8] Richard Montgomery, *A Tour of Subriemannian Geometries, Their Geodesics and Applications*, Mathematical Surveys and Monographs 91, AMS, 2002.
[9] Syoiti Ninomiya and Nicolas Victoir, *Weak Approximation of SDEs and Application to Derivative Pricing*, preprint.
[10] Christophe Reutenauer, *Free Lie Algebras*, London Mathematical Monographs 7, Oxford University Press, 1993.
[11] Christian Schmeiser, Alexander Soreff and Josef Teichmann, *Recombination Schemes for Cubature on Wiener Space*, working paper.
[12] A. H. Stroud, *Approximate Calculation of Multiple Integrals*, Series in Automatic Computation, Prentice Hall, 1971.
[13] Josef Teichmann, *Calculating the Greeks by Cubature Formulas*, Proceedings of the Royal Society London A 462, 647-670, 2006.
[14] Nicolas Victoir, *Asymmetric Cubature Formulae with Few Points in High Dimension for Symmetric Measures*, SIAM J. Numer. Anal. 42, no. 1, 209–227, 2004.
[15] Jacek Zienkiewicz, *Note on Analytic Regularity of Heat Kernels on Nilpotent Lie Groups*, Colloquium Mathematicum 97, No. 2, 163–168, 2003.