The Annals of Applied Probability
2012, Vol. 22, No. 4, 1301–1327
DOI: 10.1214/11-AAP786

HIGH ORDER RECOMBINATION AND AN APPLICATION TO CUBATURE ON WIENER SPACE

BY C. LITTERER[1] AND T. LYONS[2]

Imperial College London and University of Oxford

Particle methods are widely used because they can provide accurate descriptions of evolving measures. Recently it has become clear that by stepping outside the Monte Carlo paradigm these methods can be of higher order with effective and transparent error bounds. A weakness of particle methods (particularly in the higher order case) is the tendency for the number of particles to explode if the process is iterated and accuracy preserved. In this paper we identify a new approach that allows dynamic recombination in such methods and retains the high order accuracy by simplifying the support of the intermediate measures used in the iteration. We describe an algorithm that can be used to simplify the support of a discrete measure and give an application to the cubature on Wiener space method developed by Lyons and Victoir [*Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **460** (2004) 169–198].

1. Introduction. In pricing and hedging financial derivatives, as well as in assessing the risk inherent in complex systems, we often have to find approximations to expectations of functionals of solutions to stochastic differential equations (SDE). We consider a Stratonovich stochastic differential equation

$$d\xi_{t,x} = V_0(\xi_{t,x})\,dt + \sum_{i=1}^{d} V_i(\xi_{t,x}) \circ dB_t^i, \qquad \xi_{0,x} = x,$$

defined by a family of smooth vector fields V_i and driven by Brownian motion. It is well known that computing $P_{T-t}f := E(f(\xi_{T-t,x}))$ corresponds to solving a parabolic partial differential equation (PDE). High dimension and hypo-ellipticity are common obstacles that arise when one calculates these quantities numerically. When facing these obstacles some classical computational methods become unstable and/or intractable.

There are many settings where one is interested in tracking the evolution of a measure over time in an effective numerical fashion. One example is the numerical approximation to the solution of a linear parabolic PDE. In this case, one tracks the evolution of the heat kernel measure associated to the PDE. Another example is the

Received August 2010; revised March 2011.

[1]Supported by the Leverhulme trust Grant F/08772/E. Final version of the paper prepared at TU Berlin spported by ERC Grant 258237 (FP7/2007-2013).

[2]Supported by the EPSRC Grant EP/H000100/1.

MSC2010 subject classifications. 65C30.

Key words and phrases. Cubature, signature, recombination, stochastic differential equations.

filtering problem where one wishes to approximate the unnormalized conditional distribution of the signal, which is governed by a stochastic partial differential equation known as the Zakai equation.

An evolving measure can be viewed as a path in the space of measures. Thus, even if the underlying state space is finite dimensional, we potentially face an infinite-dimensional problem. Particle approximations can, in many cases, provide good descriptions of evolving measures (see, e.g., the survey articles [2, 3]). Higher order methods may allow us to take far fewer time steps than classical methods in the approximations. An example of a higher order particle method may be found in Kusuoka [6]. Although effective in practice (compare Ninomiya [13] and Ninomiya and Victoir [14]), these methods have the drawback that the number of particles can explode exponentially if the process is iterated and accuracy preserved (see, e.g., Lyons and Victoir [12]).

Sometimes the essential properties of a probability measure we care about can accurately be described and captured by the expectations of a finite set of test functions. If we can find such a family of test functions we can replace the original measure with a simpler measure with smaller support that integrates all test functions correctly and hence, still has the right properties, provided, of course, the number of test functions is small compared to the cardinality of the support of the original measure. We will also insist that the reduced measure $\tilde{\mu}$ has $\operatorname{supp}(\tilde{\mu}) \subseteq \operatorname{supp}(\mu)$. This condition ensures that feasibility constraints imposed on the measure μ will also be satisfied by $\tilde{\mu}$. For a finite Borel measure μ on a polish space Ω and a set of integrable functions $\{p_1, \ldots, p_n\}$, we can show that such a reduced measure $\tilde{\mu}$ always exists with $\operatorname{card}(\operatorname{supp}(\tilde{\mu})) \leq n + 1$.

In this paper we present a simple algorithm that can be used to compute reduced measures for discrete measures μ. The runtime is polynomial in the size of the support of the measure μ. The algorithm relies on the observation that if P is the R^n valued random variable $P(x) := (p_1(x), \ldots, p_n(x))$ and μ_P the law of P under the measure μ, then finding a reduced measure $\tilde{\mu}$ is equivalent to finding $\tilde{\mu}_P$ a discrete measure on R^n with $\operatorname{card}(\operatorname{supp}(\tilde{\mu}_P)) = n + 1$ and the same center of mass (CoM) as μ_P.

We describe an application to the Kusuoka–Lyons–Victoir (or KLV cubature on Wiener space) method developed by Lyons and Victoir [12], following Kusuoka [6]. It provides higher order approximations to $E(f(\xi_{T,x}))$ if the test function f is Lipschitz and the vector fields satisfy Kusuoka's UFG condition (see [7]) which is weaker than the usual Hörmander condition. The expectation $E(f(\xi_{T,x}))$ might be viewed as an infinite-dimensional integral against Wiener measure. The authors construct discrete cubature measures $Q_T = \sum_{i=1}^{n} \lambda_i \delta_{\omega_{i,T}}$ supported on continuous paths of bounded variation that approximate Wiener measure in the sense that they integrate iterated integrals up to a fixed degree correctly. The expectation of a Wiener functional $f(\xi_{T,x})$ against the discrete cubature measure may be obtained by computing the endpoints of the solution of the SDE along

the paths in the support of Q_T. Thus the KLV method might be viewed as a discrete Markov kernel taking discrete measures on R^N to discrete measures on R^N. More explicitly we have

$$\mathrm{KLV}(\delta_x, T) = \sum_{i=1}^{n} \lambda_i \delta_{\xi_{T,x}(\omega_{i,T})}$$

and

$$E_{Q_T} f(\xi_{T,x}) = E_{\mathrm{KLV}(\delta_x,T)} f.$$

The bound on the error when replacing the Wiener measure with a cubature measure is given in terms of higher order derivatives of f, so in general will not be small as f is only assumed to be Lipschitz. The results in Kusuoka and Stroock [8] and Kusuoka [7] show that $P_t f$ will be smooth, at least in the direction of the vector fields V_i. This is resolved by applying the method iteratively over a partition of the time interval $[0, T]$. The operator corresponding to the iterated application of the KLV method is Markov and hence, the error of the approximation of $P_T f$ on the global time interval $[0, T]$ is the sum of the error of the approximations over the subintervals of the partition. So considering an uneven partition of the global time interval $[0, T]$ with time steps getting smaller toward the end, we can iteratively apply the cubature method over the subintervals and reduce the error in the approximation to any accuracy. If m is the degree of the cubature formula, we can find a partition such that the error in the weak approximation is uniformly bounded by

$$Ck^{-(m-1)/2} \|\nabla f\|_\infty,$$

where k is the number of time steps in the partition and C a constant independent of k and f.

The iterated KLV method might be viewed as a particle system on R^N where the particles branch in an n-ary tree. Hence, the number of ODEs to solve grows exponentially in the number of iterations. In this paper we add recombination to the KLV method. After each application of the KLV operation we replace the intermediate measures by reduced measures. The property of the KLV measure we are targeting is to integrate $P_t f$, the heat kernel applied to f, correctly. We have identified a finite set of test functions that ensures that the bound on the overall error of the approximation of $P_T f$ is only increased by a constant factor and hence, the modified method has the same convergence properties. Moreover, we can show that under the Hörmander condition for bounded vector fields the number of test functions required grows polynomially in the number of iterations.

We finish the paper with a toy numerical example that illustrates how one blends the methods of this paper together in a concrete example to compute a solution of a one-dimensional PDE to high accuracy when the boundary data is piecewise smooth and the discontinuities are not known to the PDE solver.

We believe that the combination of the two ideas—higher order particle methods to describe the evolution of a measure on the one hand and simplifying the support of the measures used in the description, by characterizing essential properties of a measure using the expectations of a finite set of test functions on the other hand—have more general applications than investigated so far. Applications to the stochastic filtering problem appear to be particularly promising (see Litterer and Lyons [10, 11] for an outline).

2. A reduction algorithm for the support of a discrete measure. Let us start the precise description of the reduction problem. The notation in this section is independent of the notation used in the description of the cubature method in the following sections. Consider a finite set of test functions $P_n = \{p_1, \dots, p_n\}$ on (Ω, μ), a measure space with μ a finite discrete measure

$$\mu = \sum_{i=1}^{\hat{n}} \lambda_i \delta_{z_i}, \qquad \lambda_i > 0, z_i \in \Omega,$$

with large support. By this we mean that $\hat{n}$ is at least of order n^2. In the following we assume that μ is a probability measure, that is, the weights add up to one.

DEFINITION 1. We will call a discrete probability measure $\tilde{\mu}$ a reduced measure with respect to μ and P_n if it satisfies the following three conditions:

(1) $\operatorname{supp}(\tilde{\mu}) \subseteq \operatorname{supp}(\mu)$.
(2) For all $p \in P_n$

$$\int p(x)\tilde{\mu}(dx) = \int p(x)\mu(dx).$$

(3) $\operatorname{card}(\operatorname{supp}(\tilde{\mu})) \leq n + 1$.

The first condition is more important than it looks as it ensures that feasibility constraints imposed on samples drawn from μ will also be satisfied by $\tilde{\mu}$. We wish to construct effective algorithms to compute the reduced measure.

Let P be the R^n-valued random variable $P := (p_1, \dots, p_n)$ defined on (Ω, μ). Then the law μ_P of P is the discrete measure on R^n

$$\mu_P = \sum_{i=1}^{\hat{n}} \lambda_i \delta_{x_i}, \qquad x_i = (p_1(z_i), \dots, p_n(z_i))^T \in R^n. \tag{1}$$

The center of mass (CoM) for the measure μ_P is given by

$$\operatorname{CoM}(\mu_P) = \sum_{i=1}^{\hat{n}} \lambda_i x_i. \tag{2}$$

To find a reduced measure we articulate an equivalent problem in terms of μ_P. The problem becomes finding a subset x_{i_k} of the points x_i and positive weights $\tilde{\lambda}_{i_k}$ to produce a new probability measure $\tilde{\mu}_P$ such that $\mathrm{CoM}(\tilde{\mu}_P) = \mathrm{CoM}(\mu_P)$. A reduced measure $\tilde{\mu}$ is then easily obtained from $\tilde{\mu}_P$ by taking

$$\tilde{\mu} = \sum \tilde{\lambda}_{i_k} \delta_{z_{i_k}}$$

with $z_{i_k} \in \mathrm{supp}(\mu)$ satisfying $P(z_{i_k}) = x_{i_k}$.

Note that given any subset x_{i_k} there exist suitable weights $\tilde{\lambda}_{i_k}$ if and only if $\mathrm{CoM}(\mu_P)$ is contained in the convex hull of these points. Caratheodory's theorem implies that in principle one can always find $\tilde{\mu}_P$ with support having cardinality at most $n+1$ and the algorithm explained below provides a constructive proof to that.

By considering $x_i - \mathrm{CoM}(\mu_P)$ in place of the x_i, we may assume without loss of generality that $\mathrm{CoM}(\mu_P)$ is at the origin. We may also assume that the x_i are all distinct, as we can otherwise eliminate points x_i from the original measure μ by sorting and combining them.

A first algorithm (Algorithm 1), communicated to us by Victoir [15], sequentially eliminates particles from the support of the measure. It is well known and has, for example, been used in constructive proofs of Tchakaloff's theorem (Davis [5]).

Given any $n+2$ points, the system given by

$$\sum_{i=1}^{n+2} u_i x_{k_i} = 0, \tag{3}$$

$$\sum_{i=1}^{n+2} u_i = 0$$

is a linear system with $n+2$ variables, but only $n+1$ constraints. Therefore, it has a nontrivial solution, which may, for example, be determined using Gaussian elimination. Thus we may either add

$$\min_{u_i<0}\left(-\frac{\lambda_i}{u_i}\right)\sum_{j=1}^{n+2} u_j x_{k_j}$$

to (2) or subtract

$$\min_{u_i>0}\left(\frac{\lambda_i}{u_i}\right)\sum_{j=1}^{n+2} u_j x_{k_j}$$

from (2) leaving all weights in the result nonnegative and their overall sum unchanged. In either case, by construction, the coefficient of some x_j vanishes. We now have obtained a new probability measure with the same center of mass and at

least one point less in the support. Applying the procedure iteratively until there are only $n+1$ points left, we obtain a reduced measure. Clearly the method requires no more than $\hat{n}$ iterations of the above procedure.

REMARK 2. If $\tilde{n}$ is the dimension of the lowest-dimensional (affine) subspace of R^n containing the set $\{(p_1(y), \ldots, p_n(y)) | y \in \operatorname{supp}(\mu)\}$, we can continue to apply the elimination procedure described in Algorithm 1 until $\operatorname{card}(\operatorname{supp}(\tilde{\mu})) \leq \tilde{n}+1$.

For improving the order of the overall algorithm we now look at suitable linear combinations instead of points.

To describe the algorithm we define an abstract procedure A that takes a discrete probability measure ν with $2(n+1)$ particles in its support and returns another discrete probability measure $\tilde{\nu}$ with $(n+1)$ particles in its support satisfying $\operatorname{CoM}(\nu) = \operatorname{CoM}(\tilde{\nu})$ and $\operatorname{supp}(\tilde{\nu}) \subseteq \operatorname{supp}(\nu)$. Procedure A may, for example, be realized by $n+1$ applications of the reduction procedure of Algorithm 1.

Main reduction algorithm (*Algorithm* 2): (1) Partition the support of $\mu_P = \sum_{i=1}^{\hat{n}} \lambda_i \delta_{x_i}$ into $2(n+1)$ sets of as near equal size as possible. Let these sets be denoted by I_j, $1 \leq j \leq 2(n+1)$.

(2) Compute the probability measure $\nu = \sum_{i=1}^{2(n+1)} \nu_i \delta_{\tilde{x}_i}$ where

$$\tilde{x}_j = E_{\mu_P}(x | x \in I_j) = \sum_{x_i \in I_j} \frac{\lambda_i x_i}{\nu_j}$$

and $\nu_j = \mu_P(I_j) = \sum_{i\,:\,x_i \in I_j} \lambda_i$.

(3) Apply procedure A to compute a measure $\tilde{\nu} = \sum_{j=1}^{n+1} \tilde{\nu}_{i_j} \delta_{\tilde{x}_{i_j}}$ with $\operatorname{CoM}(\nu) = \operatorname{CoM}(\tilde{\nu})$.

(4) Repeat (1)–(3) with

$$\mu'_P = \sum_{j=1}^{n+1} \sum_{x_k \in I_{i_j}} \tilde{\nu}_{i_j} \frac{\lambda_k}{\nu_{i_j}} \delta_{x_k}$$

for μ_P until $n+1$ particles are left in the support of μ_P.

PROPOSITION 3. *Given μ and P_n, the algorithm described above requires $\lceil \lg(\hat{n}/n) \rceil$ iterations of procedure A to compute a reduced measure.*

PROOF. We might interpret the points $\tilde{x}_j$ as the respective center of masses of the individual subsets I_j.

It is clear that μ'_P has positive weights and support contained in the support of μ_P. Hence, we only need to show that $\operatorname{CoM}(\mu'_P) = \operatorname{CoM}(\mu_P)$.

We have

$$\begin{aligned}\mathrm{CoM}(\mu'_P) &= \sum_{j=1}^{n+1} \tilde{\nu}_{i_j} \sum_{x_k \in I_{i_j}} \frac{\lambda_k x_k}{\nu_j} = \sum_{j=1}^{n+1} \tilde{\nu}_{i_j} \tilde{x}_{i_j} \\ &= \mathrm{CoM}(\tilde{\nu}) = \mathrm{CoM}(\nu) = \sum_{j=1}^{2(n+1)} \nu_j \tilde{x}_j = \sum_{j=1}^{2(n+1)} \nu_j \sum_{x_i \in I_j} \frac{\lambda_i x_i}{\nu_j} \\ &= \mathrm{CoM}(\mu_P).\end{aligned}$$

As $\hat{n} \le n2^{\lceil \lg(\hat{n}/n) \rceil}$, we may assume without loss of generality that $\hat{n} = n2^{\lceil \lg(\hat{n}/n) \rceil}$. It is obvious that each iteration halves the number of particles in the support of μ_P and we require exactly $\lceil \lg(\hat{n}/n) \rceil$ iterations. □

COROLLARY 4. *Using the main reduction algorithm we can compute a reduced measure with respect to μ and P_n in*

$$O\big(n\hat{n} + n \log(\hat{n}/n) C(n+2, n+1)\big)$$

steps where $C(n+2, n+1)$ represents the number of steps required to solve a system of linear equations with $n+2$ variables and $n+1$ constraints.

PROOF. To compute the intermediate measures ν, we need to calculate n-dimensional linear combinations. The number of steps required for these additions is bounded above by the series

$$n \sum_{i=0}^{\infty} \hat{n} 2^{-i} = 2n\hat{n}.$$

The procedure A may be realized by $n+1$ applications of the reduction procedure used in Algorithm 1 described above. □

REMARK 5. Note that the linear systems of equations we need to solve in the algorithm are singular. Hence, for a practical implementation we have used a method based on the singular value decomposition (SVD) to avoid numerical instability.[3]

If the support of the measure μ we wish to target is particularly large or possibly even infinite, we can consider a different approach. If we can find a subset of points that with a reasonably high probability contains the CoM in its convex hull, we may use linear programming to check if a given set of points contains the CoM in

[3] A dll with an implementation of a version of the algorithm and a Visual Studio project with a simple example for its use can currently be found at http://www.maths.ox.ac.uk/~tlyons/Recombination/reduce_dist_01_paper.zip.

its convex hull and reconstruct the weights. The results in Wendel [16] imply, for example, that a collection of k uniform i.i.d. random variables on the unit sphere in R^N contains the origin with probability

$$P_{N,k} = 1 - 2^{-k+1} \sum_{j=0}^{N-1} \binom{k-1}{j}.$$

In particular this yields $P_{N,2N} = 1/2$.

3. Outline of the cubature algorithm. We describe the cubature method developed by Lyons and Victoir [12]. Throughout the paper, C is a constant that may change from line to line; specific constants, however, will be indexed $C_1, C_2, \ldots$. Let $C_b^\infty(R^N, R^N)$ denote the smooth bounded R^N valued functions whose derivatives of any order are bounded. Then $V_i \in C_b^\infty(R^N, R^N)$, $0 \le i \le d$, may be regarded as vector fields on R^N. We define a partial differential operator $L = V_0 + \frac{1}{2}(V_1^2 + \cdots + V_d^2)$ and consider the following parabolic partial differential equation (PDE)

$$\begin{aligned} \frac{\partial u}{\partial t}(t,x) &= -Lu(t,x), \\ u(T,x) &= f(x) \end{aligned} \tag{4}$$

for a given Lipschitz function f. The aim is to find an approximation of $u(0,x)$ for a given x. Consider the probability space $(C_0^0([0,T], R^d), \mathcal{F}, \mathbb{P})$, where $C_0^0([0,T], R^d)$ is the space of R^d valued continuous functions starting at 0, $\mathcal{F}$ its usual Borel σ-field and $\mathbb{P}$ the Wiener measure. Define the coordinate mapping process $B_t^i(\omega) = \omega^i(t)$ for $t \in [0,T]$, $\omega \in \Omega$. Under Wiener measure, $B = (B_t^1, \ldots, B_t^d)$ is a Brownian motion starting at zero. Furthermore, let $B_t^0(t) = t$. Let $\xi_{t,x}$, $t \in [0,T]$, $x \in R^N$ be a version of the solution of the Stratonovich stochastic differential equation (SDE)

$$d\xi_{t,x} = \sum_{i=0}^{d} V_i(\xi_{t,x}) \circ dB_t^i, \qquad \xi_{0,x} = x, \tag{5}$$

that coincides with the pathwise solution on continuous paths of bounded variation. In this case, classical theory tells us that $u(t,x) = E(f(\xi_{T-t,x}))$ is the solution to (4).

We define the Itô functional $\Phi_{T,x} : C_0^0([0,T], R^d) \to R^N$ by

$$\Phi_{T,x}(\omega) = \xi_{T,x}(\omega). \tag{6}$$

Denote by $R_m[X_1, \ldots, X_d]$ the space of polynomials[4] in d variables having degree less or equal to m. Let μ be a positive Borel measure on R^d. A discrete measure $\tilde{\mu}$

$$\tilde{\mu} = \sum_{i=1}^{n} \lambda_i \delta_{x_i}$$

with $x_1, \ldots, x_n$ contained in $\mathrm{supp}(\mu)$ satisfies a cubature formula of degree m if and only if for all polynomials $P \in R_m[X_1, \ldots, X_d]$,

$$\int_{R^d} P(x)\mu(dx) = \int_{R^d} P(x)\tilde{\mu}(dx) = \sum_{i=1}^{n} \lambda_i P(x_i).$$

It is well known that if all moments of μ up to degree m exist we can always find such a measure with

$$\mathrm{card}(\mathrm{supp}(\mu)) \leq \dim(R_m[X_1, \ldots, X_d]) + 1$$

(see, e.g., Bayer and Teichmann [1]). More generally we have the following lemma, which we state without proof.

LEMMA 6. *Let Ω be a polish space, $\mathcal{F}$ its Borel sets and μ a Borel probability measure on $(\Omega, \mathcal{F})$. Let $f_1, \ldots, f_n$ be a finite sequence of real-valued Borel measurable functions on the probability space $(\Omega, \mathcal{F}, \mu)$ with $E(|f_i|) < \infty$ for $1 \leq i \leq n$. Moreover, suppose that D is a Borel set with $\mu(D) = 1$. Then there exist points $w_1, \ldots, w_{n+1} \in D$ and a discrete measure*

$$\tilde{\mu} = \sum_{i=1}^{n+1} \lambda_i \delta_{w_i}$$

such that

$$E_\mu(f_i) = E_{\tilde{\mu}}(f_i)$$

for $1 \leq i \leq n$.

In other words, μ admits a reduced measure $\tilde{\mu}_P$ with respect to any finite set P of integrable functions. In connection with the use of the Taylor formula, a cubature measure provides an effective tool for integration over finite-dimensional spaces.

One can formulate an analogous condition to identify cubature measures on Wiener space. Here the role of polynomials is taken by iterated integrals of the form

$$\int_{0<t_1<\cdots<t_k<T} \circ dB_{t_1}^{i_1} \cdots \circ dB_{t_k}^{i_k}.$$

[4]Any finite-dimensional space of integrable and continuous functions could be used to define cubature. This extension can be helpful.

We identify this iterated integral by the multi-index $(i_1, \dots, i_k)$.

Define the set of all multi-indices A by

$$A = \bigcup_{k=0}^{\infty} \{0, \dots, d\}^k$$

and let $\alpha = (\alpha_1, \dots, \alpha_k) \in A$ be a multi-index. Furthermore, we define a degree on a multi-index α by $\|\alpha\| = k + \operatorname{card}(j : \alpha_j = 0)$ and let

$$A(j) = \{\alpha \in A : \|\alpha\| \le j\}.$$

Moreover, define A_1 by $A_1 = A \setminus \{\varnothing, (0)\}$ and let $A_1(j) = \{\alpha \in A_1 : \|\alpha\| \le j\}$. It follows from the scaling property of Brownian motion that

$$\int_{0<t_1<\cdots<t_k<T} \circ dB_{t_1}^{\alpha_1} \cdots \circ dB_{t_k}^{\alpha_k}$$

equals, in law,

$$T^{\|\alpha\|/2} \int_{0<t_1<\cdots<t_k<1} \circ dB_{t_1}^{\alpha_1} \cdots \circ dB_{t_k}^{\alpha_k}. \tag{7}$$

DEFINITION 7. Fix a finite set of multi-indices $\tilde{A} \subseteq A$. We say that a discrete measure Q_T assigning positive weights $\lambda_1, \dots, \lambda_n$ to paths

$$\omega_1, \dots, \omega_n \in C_{0,bv}^0([0, T], R^d)$$

is a cubature measure if for all $(i_1, \dots, i_k) \in \tilde{A}$,

$$E\left(\int_{0<t_1<\cdots<t_k<T} \circ dB_{t_1}^{i_1} \cdots \circ dB_{t_k}^{i_k}\right) = \sum_{j=1}^{n} \lambda_j \int_{0<t_1<\cdots<t_k<T} d\omega_j^{i_1}(t_1) \cdots d\omega_j^{i_k}(t_k),$$

where the expectation is taken under Wiener measure. If $\tilde{A} = A(m)$ we say that

$$Q_T = \sum_{j=1}^{n} \lambda_j \delta_{\omega_j}$$

is cubature measure of degree m.

In [12], the authors show that one can always find a cubature measure supported on, at most, $\operatorname{card}(\tilde{A})$ continuous paths of bounded variation. More importantly, they give an explicit construction of a degree 5 cubature formula with $O(d^3)$ paths in its support.

Suppose paths $\omega_1, \dots, \omega_n$ and weights λ_i define a cubature measure for $T = 1$. It follows immediately from (7) that the measure supported on paths $\omega_{T,i}$ given by

$$\omega_{T,i}^j = \sqrt{T}\omega_i^j(t/T), \qquad j = 1, \dots, d, \tag{8}$$

and unchanged weights λ_i defines a cubature measure for general T. From now on suppose that the measure $Q := Q_1$ is a cubature measure of degree m.

The following proposition, taken from [12], is the key step in estimating the error E_T when one approximates the expectation of $f(\xi_{T,x})$ under the Wiener measure by its expectation against Q.

PROPOSITION 8.

$$E_T := \sup_{x\in R^n}\left|Ef(\xi_{T,x}) - \sum_{i=1}^{n}\lambda_i f(\Phi_{T,x}(\omega_{T,i}))\right|$$

$$\le C\sum_{j=m+1}^{m+2} T^{j/2} \sup_{(\alpha_1,\dots,\alpha_i)\in A(j)\setminus A(j-1)} \|V_{\alpha_1}\cdots V_{\alpha_i} f\|_\infty,$$

where C is a constant that only depends on d, m and Q_1.

In general, the right-hand side of the inequality in Theorem 8 is not sufficient to directly obtain a good error bound for the approximation of the expectation, in particular if f is only assumed to be Lipschitz, the estimate appears useless. So, instead of approximating

$$P_T f(x) := E(f(\xi_{T,x}))$$

in one step, one considers a partition $\mathcal{D}$ of the interval $[0, T]$

$$t_0 = 0 < t_1 < \cdots < t_k = T$$

with $s_j = t_j - t_{j-1}$ and solves the problem over each of the smaller subintervals by applying the cubature method recursively. If τ and τ' are two path segments, we denote their concatenation by $\tau \otimes \tau'$. For the approximation, we consider all possible concatenations of cubature paths over the subintervals, that is, all paths of the form $\omega_{s_1,i_1} \otimes \cdots \otimes \omega_{s_k,i_k}$. We define a corresponding probability measure ν by

$$\nu = \sum_{i_1,\dots,i_k=1}^{n} \lambda_{i_1}\cdots\lambda_{i_k}\delta_{\omega_{s_1,i_1}\otimes\cdots\otimes\omega_{s_k,i_k}}.$$

The following theorem taken from Lyons and Victoir [12] is the main error estimate for the iterated cubature method, which we in the following also refer to as the Kusuoka–Lyons–Victoir (KLV) method.

THEOREM 9. *The total error $E_{\mathcal{D}}$ for the approximation*

$$E_{\mathcal{D}} := \sup_{x\in R^N} |P_T f - E_\nu(f(\xi_{T,x}))|$$

$$= \sup_{x\in R^N}\left|P_T f(x) - \sum_{i_1=1}^{n}\cdots\sum_{i_k=1}^{n}\lambda_{i_1}\cdots\lambda_{i_k} f\big(\Phi_{T,x}(\omega_{s_1,i_1}\otimes\cdots\otimes\omega_{s_k,i_k})\big)\right|$$

is bounded by

$$C_1(T)\|\nabla f\|_\infty\left(s_k^{1/2}+\sum_{j=m}^{m+1}\sum_{i=1}^{k-1}\frac{s_i^{(j+1)/2}}{(T-t_i)^{j/2}}\right), \tag{9}$$

where $C_1(T)$ is a constant independent of f and k, the number of time steps in the partition of the time interval $[0,T]$.

To compute the expectation with respect to the measure ν exactly requires one to solve

$$\frac{n^{k+1}-1}{n-1}$$

inhomogeneous ODEs (each corresponding to a path $\omega_{s_1,i_1}\otimes\cdots\otimes\omega_{s_k,i_k}$) where n denotes the number of paths in the support of the cubature measure Q and k the number of subintervals in the partition. Hence, the number of ODEs to solve grows exponentially in the number of iterations.

Following Kusuoka [7], we define for multi-indices $\alpha=(\alpha_1,\dots,\alpha_k),\beta=(\beta_1,\dots,\beta_l)\in A$ a multiplication by

$$\alpha*\beta=(\alpha_1,\dots,\alpha_k,\beta_1,\dots,\beta_l).$$

We inductively define a family of vector fields indexed by A by taking

$$V_{[\varnothing]}=0,\qquad V_{[i]}=V_i,\qquad 0\le i\le d,$$
$$V_{[\alpha*i]}=\bigl[V_{[\alpha]},V_i\bigr],\qquad 0\le i\le d,\alpha\in A.$$

The main ingredients used when obtaining the bound (9) are Proposition 8 and the following regularity result due to Kusuoka and Stroock [8] and Kusuoka [7], which says that even if f is not smooth, P_sf is smooth in the directions of the vector fields V_i. Let f be Lipschitz and $\alpha_1,\dots,\alpha_k\in A_1$, then for all $t\in(0,1]$,

$$\bigl\|V_{[\alpha_1]}\cdots V_{[\alpha_k]}P_tf\bigr\|_\infty\le\frac{Ct^{1/2}}{t^{(\|\alpha_1\|+\cdots+\|\alpha_k\|)/2}}\|\nabla f\|_\infty \tag{10}$$

provided the vector fields satisfy the UFG condition defined below.

Following Kusuoka [7] we introduce a condition on the vector fields.

DEFINITION 10. The family of vector fields V_i, $i=0,\dots,d$, is said to satisfy the condition (UFG) if the Lie algebra generated by it is finitely generated as a C_b^∞ left module, that is, there exists a positive k and $u_{\alpha,\beta}\in C_b^\infty$ satisfying for all $\alpha\in A_1$,

$$V_{[\alpha]}=\sum_{\beta\in A_1(k)}u_{\alpha,\beta}V_{[\beta]}. \tag{11}$$

The bounds for the error of the KLV method derived in Theorem 9 (see Lyons and Victoir [12] for details) assume that the system of vector fields $V_i, i = 0, \dots, d$, satisfies the UFG condition.

DEFINITION 11. We define the (formal) degree of a vector field $V_{[\alpha]}, \alpha \in A$, denoted by d_α to be the minimal integer k such that $V_{[\alpha]}$ may be written as

$$V_{[\alpha]} = \sum_{\beta \in A_1(k)} u_{\alpha,\beta} V_{[\beta]}$$

with $u_{\alpha,\beta} \in C_b^\infty$.

Note that for $\alpha \in A_1$ we always have $d_\alpha \le \|\alpha\|$. It was pointed out in Crisan and Ghazali [4] that the analysis in Lyons and Victoir [12] for the bound in (9) requires V_0 to have formal degree at most 2. If the formal degree of V_0 is greater, the bound in (12) changes and all bounds in the paper will change accordingly. For sake of simplicity we will in the following assume that V_0 has formal degree 2. The bounds can be improved in an obvious way if the degree is 1 or 0. For a generalized error estimate based on Kusuoka's ideas [6] that does not require this additional condition, see Litterer [9].

A trivial generalization of Corollary 18 in Crisan and Ghazali [4] allows us to state a version of the Kusuoka and Stroock estimate in terms of the formal degree of a vector field. Let f be as above and $\alpha_1, \dots, \alpha_k \in A$ then for all $t \in (0, 1]$

$$\| V_{[\alpha_1]} \cdots V_{[\alpha_k]} P_t f \|_\infty \le \frac{C t^{1/2}}{t^{(d_{\alpha_1} + \cdots + d_{\alpha_k})/2}} \| \nabla f \|_\infty. \tag{12}$$

For the remainder of the paper, when we consider recombination, we are going to assume the following uniform Hörmander condition.

DEFINITION 12. We say that a collection of smooth vector fields V_i, $i = 0, \dots, d$, satisfies the uniform Hörmander condition (UH) if there is an integer p such that

$$\inf \left\{ \sum_{\alpha \in A_1(p)} \langle V_{[\alpha]}(x), \xi \rangle^2 ; x, \xi \in R^N, |\xi| = 1 \right\} := M > 0.$$

Note that the uniform Hörmander condition implies the UFG condition. Under this stronger assumption it is straightforward to show that, in addition, $P_t f$ is a smooth function on R^N with explicit bounds on its derivatives. We outline an argument below that follows Kusuoka [7] and gives bounds on the regularity of $P_t f$, which we will use in the following section when we apply recombination to the cubature method.

Following Kusuoka [7], let $F(x) \in C_b^\infty(R^N; R^N \otimes R^N)$ be given by

$$F(x) = \sum_{\alpha \in A_1(p)} V_{[\alpha]}(x) \otimes V_{[\alpha]}(x), \qquad x \in R^N,$$

and $\lambda_0 : R^N \to [0, \infty)$ be the continuous function

$$\lambda_0(x) = \inf\{\langle F(x)y, y\rangle; y \in R^N, |y| = 1\}, \qquad x \in R^N.$$

Note that

$$\langle F(x)y, y\rangle = \sum_{\alpha \in A_1(p)} \langle V_{[\alpha]}(x), y\rangle^2$$

and hence, under the Hörmander condition (UH), we have $\lambda_0(x) \geq M > 0$ for all $x \in R^N$. As in Kusuoka [7], let $e_i = \{\delta_{ij}\}_1^N$ and $a_{\alpha,i} : R^N \to R$, $\alpha \in A_1(p), i = 1, \ldots, N$, be given by

$$a_{\alpha,i}(x) = \langle e_i, F(x)^{-1} V_{[\alpha]}(x)\rangle, \qquad x \in R^N, \tag{13}$$

and observe that

$$\frac{\partial}{\partial x_i} = \sum_{\alpha \in A_1(p)} a_{\alpha,i} V_{[\alpha]}. \tag{14}$$

The following lemma may be found in Kusuoka [7], page 274.

LEMMA 13. *Let $\alpha \in A_1(p), i, i_1, \ldots, i_k \in \{1, \ldots, N\}$. Then $a_{\alpha,i}$ defined as in (13) satisfies*

$$\left| \frac{\partial^k}{\partial x_{i_1} \cdots \partial x_{i_k}} a_{\alpha,i}(x) \right| \leq CN\lambda_0(x)^{-(k+1)} \leq CN \max\bigl(M^{-(k+1)}, 1\bigr) \tag{15}$$

for all x in R^N.

The lemma shows that the functions $a_{\alpha,i}$ are in $C_b^\infty(R^N)$. Together with (14) this immediately implies that the vector fields $\frac{\partial}{\partial x_i}$, $i = 1, \ldots, N$, have finite formal degree no greater than p. Just like identity (12), the following corollary is a trivial generalization of Corollary 18 in [4], the result is also implicit in Kusuoka [7], Proposition 14.

COROLLARY 14. *Suppose the vector fields $(V_i, i = 0, \ldots, d)$ satisfy the uniform Hörmander condition. Then for any $j \geq 1$ there is a constant $C_2 > 0$ independent of f and t such that*

$$\sup_{i_1, \ldots, i_j \in \{1, \ldots, N\}} \left\| \frac{\partial}{\partial x_{i_1}} \cdots \frac{\partial}{\partial x_{i_j}} P_t f \right\|_\infty \leq C_2 t^{-(j-1)p/2} \|\nabla f\|_\infty$$

for all $t \in (0, 1]$, $f \in C_b^\infty(R^N)$.

We point out that the constant C_2 does (via the constant M in the Hörmander condition) depend on the underlying family of vector fields V_i.

4. Application to cubature on Wiener space.

4.1. *The reduction operation.* In the iterated KLV method (Section 3), the total error $E_{\mathcal{D}}$ over the interval of approximation $[0, T]$ is bounded by the sum of the individual errors E_{s_i} over smaller time intervals. The KLV method is sequential. Starting with a unit mass particle at a single point in space time, the measures evolve through time by replacing each particle at time t_i with a family of particles at time t_{i+1}. Together these new particles have the same mass as their parent particle and are carefully positioned to provide a high order approximation to the diffusion of the underlying SDE. The algorithms introduced in Section 2 can be used very effectively to perform a global redistribution of the mass on the particles alive at time t_i so that an essentially minimal number of particles has positive mass. At the same time we do not increase the one step errors E_{s_i} significantly or affect the order of the approximation. In this way we obtain (see Section 4.2) a global error bound over $[0, T]$ for this algorithm that is of the same order (in the number of time steps) as the unmodified KLV method. On the other hand, the blow up in the number of particles is radically reduced.

The property of the intermediate measures we are targeting is to integrate $P_t f$ correctly. To approximate the integral of a smooth function such as $P_t f$ with respect to a discrete measure, we need to find uniform functional approximation schemes that apply to smooth functions on the support of this measure. By definition, smooth functions can always be well approximated on balls by polynomials. However, only after one has set a fixed error bound ε and a degree for the polynomials, the size of the balls on which the approximation holds becomes clear. The main idea will be to localize the intermediate particle measures Q into measures Q_i, where each Q_i has its support in such a good ball. We then replace (using the algorithms of Section 2) the measures Q_i by reduced measures $\tilde{Q}_i$ that integrate polynomial test functions of degree at most r correctly. In that way one knows that for a smooth function g

$$\sum_i \int g \, d\tilde{Q}_i$$

is a good approximation to $\int g \, dQ$. We subsequently prove that we can choose the localization of the measure Q in a way that ensures that we increase the overall bound on the error of the approximation only by a constant factor and examine how well we can cover the support of the intermediate measures Q by balls for the localization.

A main idea for estimating ε is to consider Taylor expansions of the function $P_t f$. We define p to be the minimal integer k such that the vector fields $\{V_\alpha, \alpha \in A_1(k)\}$ uniformly span R^N at each point of $x \in R^N$ (as in the UH condition). For g a smooth function on R^N let $dg : R^N \to \mathrm{Hom}(R^N, R)$ denote the full derivative of g. The second order derivative $d^2 g$ is then mapping

$$R^N \to \mathrm{Hom}(R^N, \mathrm{Hom}(R^N, R)) \cong \mathrm{Hom}(R^N \otimes R^N, R).$$

The higher order derivatives can similarly be regarded as sections of

$$\mathrm{Hom}((R^N)^{\otimes k}, R).$$

We define the rth degree Taylor approximation of g centered at $x_0 \in R^N$ to be

$$Tay_r(g, x_0)(y) = \sum_{i=0}^{r} (d^i g)(x_0) \frac{(y - x_0)^{\otimes i}}{i!} \tag{16}$$

and the remainder $R_r(g, x_0)(y)$ by

$$R_r(g, x_0)(y) = g(y) - Tay_r(g, x_0)(y).$$

It is clear that the rth degree Taylor approximation centered at x_0 is a polynomial of degree at most r. Given $u > 0$ and $y \in R^N$ let $B(y, u)$ denote the Euclidean ball of radius $u > 0$ centered at y. Our estimate for the remainder of the polynomial approximation is the following.

LEMMA 15. *Let $t \in (0, 1]$. The remainder function $R_r(P_t f, x_0)(y)$ is uniformly bounded on $B(x_0, u)$, that is,*

$$\left\| R_r(P_t f, x_0)|_{B(x_0,u)} \right\|_\infty \le C_4 \frac{u^{r+1}}{t^{rp/2}} \|\nabla f\|_\infty,$$

where $C_4 = C_2 C_3$ is a constant independent of f, u and t.

PROOF. By Taylor's theorem we have for $y \in B(x_0, u)$

$$|R_r(P_t f, x_0)(y)| \le \frac{\|d^{r+1} g\|_\infty}{(r+1)!} \|y - x_0\|^{r+1}$$

and we note that

$$\|d^{r+1} g\|_\infty \le C_3(r, N) \sup_{i_1 + \cdots + i_N = r+1} \left\| \frac{\partial^{i_1}}{\partial x_1^{i_1}} \cdots \frac{\partial^{i_N}}{\partial x_N^{i_N}} P_t f(y) \right\|_\infty$$

for some constant C_3 that only depends on r and N. From Corollary 14 we see that

$$\sup_{i_1 + \cdots + i_N = r+1} \left\| \frac{\partial^{i_1}}{\partial x_1^{i_1}} \cdots \frac{\partial^{i_N}}{\partial x_N^{i_N}} P_t f \right\|_\infty \le C_2 t^{-rp/2} \|\nabla f\|_\infty,$$

where C_2 is the constant from Corollary 14 and the claim follows. □

The bound on the remainder of the Taylor expansion of $P_t f$ implies that cubature measures which integrate polynomials up to degree r correctly provide good approximations provided the support of the measure we are targeting is contained in a sufficiently small patch.

PROPOSITION 16. *Suppose the uniform Hörmander condition is satisfied. Let* $t \in (0,1]$ *and* μ *be a positive measure on* R^N *with finite mass* v *satisfying* $\text{supp}(\mu) \subseteq B(x_0,u)$ *for some* $u > 0$, $x_0 \in R^N$. *Suppose a measure* $\tilde{\mu}$ *is a degree* r *cubature measure for* μ (*a reduced measure with respect to* μ *and the polynomials of degree at most* r). *Then*

$$|E_\mu P_t f - E_{\tilde{\mu}} P_t f| \le C_4 v \frac{u^{r+1}}{t^{rp/2}} \|\nabla f\|_\infty,$$

where C_4 *is the constant from Lemma* 15 *and independent of* t, f, x_0 *and* u.

PROOF. We have

$$E_\mu P_t f - E_{\tilde{\mu}} P_t f = (E_\mu - E_{\tilde{\mu}})(Tay_r(P_t f, x_0)) + E_\mu R_r(P_t f, x_0) - E_{\tilde{\mu}} R_r(P_t f, x_0).$$

Since $\tilde{\mu}$ is a cubature measure and integrates polynomials of degree at most r correctly, the first term of the sum vanishes. Lemma 15 gives us the required bounds on the remaining terms. □

Let μ be a discrete probability measure on R^N and $(U_j)_{j=1}^{\ell}$ be a collection of balls of radius u on R^N that covers the support of μ. Then there exists a collection of positive measures μ_j, $1 \le j \le \ell$ such that $\mu_i \perp \mu_j$ for all $i \ne j$ (i.e., the measures have disjoint support),

$$\mu = \sum_{i=1}^{\ell} \mu_i$$

and $\text{supp}(\mu_j) \subseteq U_j \cap \text{supp}(\mu)$. We call such a collection (U_j, μ_j) a localization of μ to the cover $(U_j)_{j=1}^{\ell}$ and say u is the radius of the localization.

DEFINITION 17. We say that a measure $\tilde{\mu}$ is a reduced measure with respect to the localization $(U_j, \mu_j)_{j=1}^{\ell}$ and a finite set of integrable test functions P if there exists a localization $(U_j, \tilde{\mu}_j)_{j=1}^{\ell}$ of $\tilde{\mu}$ such that for $1 \le j \le \ell$ the measures $\tilde{\mu}_j$ are reduced measures (see Definition 1) with respect to μ_j and P.

Note that the localization of the reduced measure $\tilde{\mu}$ is with respect to the same cover as the original measure μ. It is trivial to show that reduced measures $\tilde{\mu}$ exist for any localization $(U_j, \mu_j)_{j=1}^{\ell}$ of a discrete probability measure μ and any finite set of integrable test functions P. Moreover, the number of particles in the support of $\tilde{\mu}$ is bounded above by $(\text{card}(P)+1)\ell$. The following corollary is an immediate consequence of Proposition 16. Let P in the following be a basis for the space of polynomials on R^N with degree at most r.

COROLLARY 18. *Let $t < 1$, μ be a discrete probability measure on R^N and $(U_j, \mu_j)_{j=1}^{\ell}$ a localization of radius u. If $\tilde{\mu}$ is a reduced measure with respect to $(U_j, \mu_j)_{j=1}^{\ell}$ and P, we have*

$$|E_\mu P_t f - E_{\tilde{\mu}} P_t f| \leq C_4 \frac{u^{r+1}}{t^{rp/2}} \|\nabla f\|_\infty,$$

where C_4 is the constant from Lemma 15 and independent of t, f, u and the localization of radius u.

We define the Kusuoka–Lyons–Victoir transition (KLV) over a specified time interval $[0, s]$, based on the cubature on Wiener space approach and already used in the iterative method in Section 3. The transition KLV takes discrete measures on R^N to discrete measure on R^N and may be interpreted as a discrete Markov kernel. Given a measure $\mu = \sum_{i=1}^{l} \mu_i \delta_{x_i}$ on R^N the new measure is obtained by solving differential equations along any path in the support of the cubature measure

$$\sum_{i=1}^{n} \lambda_i \delta_{\omega_i}$$

starting from any particle in the support of μ. We define

$$\mathrm{KLV}(\mu, s) = \sum_{j=1}^{l} \sum_{i=1}^{n} \mu_j \lambda_i \delta_{\Phi_{s,x_j}(\omega_i)}.$$

We are ready to consider recombination for the iterated KLV method. Let $\mathcal{D}$ be a k step partition $t_0 = 0 < t_1 < \cdots < t_k = T$ of $[0, T]$ the global time interval of the approximation and recall that $s_j = t_j - t_{j-1}$. We also let $u = (u_2, \ldots, u_{k-1}) \in R^{k-2}$ where each $u_j > 0$. Let P be a basis for the space of polynomials on R^N with degree at most r. For each time step s_j we first apply the KLV method to move particles forward in time to a measure Q. We then localize the measure Q and use the algorithm of Section 2 to compute a reduced measure with respect to the localized measure and replace Q by this reduced measure. The u_j determine the radius of the balls in the localization of the measure in the jth iteration of the method. The polynomials in P serve as the test function in the reduction.

More precisely, we define two interrelated families $Q^{(i)}_{\mathcal{D},u}(x)$ and $\tilde{Q}^{(i)}_{\mathcal{D},x}(x)$ of measures. As base case we have the measures obtained by applying twice the KLV operation starting from the point mass at x.

$$Q^{(1)}_{\mathcal{D},u}(x) := \mathrm{KLV}(\delta_x, s_1), \qquad Q^{(2)}_{\mathcal{D},u}(x) := \mathrm{KLV}\big(Q^{(1)}_{\mathcal{D},u}(x), s_2\big). \tag{17}$$

For the recursion, the measure $\tilde{Q}^{(i)}_{\mathcal{D},u}(x)$ is defined to be a reduced measure with respect to any fixed localization $(U_j, Q^{(i)}_{\mathcal{D},u}(x)_j)$ of the measure $Q^{(i)}_{\mathcal{D},u}(x)$ with radius u_j and the set of test functions P (polynomials of degree at most r). We define

$Q^{(i+1)}_{\mathcal{D},u}(x)$ by the relation

$$Q^{(i+1)}_{\mathcal{D},u}(x) := \mathrm{KLV}\big(\tilde{Q}^{(i)}_{\mathcal{D},u}(x), s_{i+1}\big) \tag{18}$$

for all $i = 2, \ldots, k-1$. Note that we do not recombine after the first and last application of the KLV operation. The reduced measures $\tilde{Q}^{(i)}_{\mathcal{D},u}(x)$ are not unique even after we fix a localization of $Q^{(i)}_{\mathcal{D},u}(x)$ and a reduced measure may be computed using the reduction algorithms of Section 2.

The main result of the section is the following theorem.

THEOREM 19. *For any choice of localizations $(U_j, Q^{(i)}_{\mathcal{D},u}(x)_j)$ with radius u_i and any reduced measures $\tilde{Q}^{(i)}_{\mathcal{D},u}(x)$ with respect to $(U_j, Q^{(i)}_{\mathcal{D},u}(x)_j)$ and test functions P, $2 \le i \le k-1$, we have*

$$\begin{aligned} E_{\mathcal{D},k} &:= \sup_x \big| P_T f(x) - E_{Q^{(k)}_{\mathcal{D},u}(x)} f \big| \\ &\le \left(C_1(T) \left(s_k^{1/2} + \sum_{i=1}^{k-1} \sum_{j=m}^{m+1} \frac{s_i^{(j+1)/2}}{(T-t_i)^{j/2}} \right) \right. \\ &\qquad \left. + C_5(T) \sum_{i=2}^{k-1} \frac{u_i^{r+1}}{(T-t_i)^{rp/2}} \right) \|\nabla f\|_\infty, \end{aligned} \tag{19}$$

where $C_1(T)$ and $C_5(T)$ are constants independent of f and the choice localizations with radius u_i. The constant $C_5(T)$ can be taken equal to C_4 if $T - t_1 \le 1$.

PROOF. The global error is bounded by

$$\begin{aligned} \big| P_T f(x) - E_{Q^{(k)}_{\mathcal{D},u}(x)} f \big| &\le \big| P_T f(x) - E_{Q^{(1)}_{\mathcal{D},u}(x)} P_{T-t_1} f \big| \\ &\quad + \big| E_{Q^{(1)}_{\mathcal{D},u}(x)} P_{T-t_1} f - E_{Q^{(2)}_{\mathcal{D},u}(x)} P_{T-t_2} f \big| \\ &\quad + \sum_{j=2}^{k-1} \big| E_{Q^{(j)}_{\mathcal{D},u}(x)} P_{T-t_j} f - E_{\tilde{Q}^{(j)}_{\mathcal{D},u}(x)} P_{T-t_j} f \big| \\ &\quad + \sum_{j=2}^{k-1} \big| E_{\tilde{Q}^{(j)}_{\mathcal{D},u}(x)} P_{T-t_j} f - E_{Q^{(j+1)}_{\mathcal{D},u}(x)} P_{T-t_{j+1}} f \big|. \end{aligned}$$

The first two terms and the terms in the second sum are the errors introduced by the KLV operation and can be bounded as in the proof of Theorem 9.

The terms in the first sum may each be bounded by using Corollary 18. □

The bounds for the error derived in this section assume that the function f is Lipschitz. If f has more regularity, it is clear different estimates can be applied

to estimate the derivatives of $P_t f$ giving alternate bounds for $E_{\mathcal{D},k}$. Clearly, a smaller number of balls in the localizations of the measures $Q^{(j)}_{\mathcal{D},u}(x)$ reduces the computational complexity of the method. We have not discussed yet how to choose the localization and the degree r in the reduction to optimize the computational complexity of the method (see Section 4.3).

4.2. *Examples for the rate of convergence of the recombining KLV method.* In this subsection we consider some particular choices of parameters for the recombining KLV method and examine their rate of convergence. We first fix for the remainder of this section (a family of) partitions $\mathcal{D}$ for the time interval $[0, T]$. We recall a family of uneven partitions from Lyons and Victoir [12] which has smaller time steps toward the end and is given by

$$t_j = T\left(1 - \left(1 - \frac{j}{k}\right)^{\gamma}\right). \tag{20}$$

For $\gamma > m - 1$ the results in [12] (see also Kusuoka [6]) show that

$$s_k^{1/2} + \sum_{i=1}^{k-1} \sum_{j=m}^{m+1} \frac{s_i^{(j+1)/2}}{(T - t_i)^{j/2}} \le C_6(m, \gamma) T^{1/2} k^{-(m-1)/2}, \tag{21}$$

while for the case $0 < \gamma < m - 1$ one obtains

$$s_k^{1/2} + \sum_{i=1}^{k-1} \sum_{j=m}^{m+1} \frac{s_i^{(j+1)/2}}{(T - t_i)^{j/2}} \le C_7(m, \gamma) T^{1/2} k^{-\gamma/2}.$$

In the following two examples we work with the partition defined in (20) and the notation of Theorem 19. Using this particular choice of partitions ensures that the bound on the KLV error is of high order in the number of iterations k.

EXAMPLE 20. Let $\gamma > m - 1$, $r = \lceil m/p \rceil$ and $u_j = s_j^{p/2-a}$, where $a := \frac{p-1}{2(\lceil m/p \rceil + 1)} \ge 0$. Then

$$\begin{aligned} &\sup_x \left| P_T f(x) - E_{Q^{(k)}_{\mathcal{D},u}(x)} f \right| \\ &\quad \le \left(C_1(T) \left(s_k^{1/2} + \sum_{i=1}^{k-1} \sum_{j=m}^{m+1} \frac{s_i^{(j+1)/2}}{(T - t_i)^{j/2}} \right) \right. \\ &\qquad \left. + C_5(T) \sum_{i=2}^{k-1} \frac{s_i^{(\lceil m/p \rceil p + 1)/2}}{(T - t_i)^{\lceil m/p \rceil p/2}} \right) \|\nabla f\| \\ &\quad \le C_8 k^{-(m-1)/2} T^{1/2} \|\nabla f\|_\infty, \end{aligned} \tag{22}$$

where $C_8 = C_6(m, \gamma)(C_1(T) + C_5(T))$.

Note that $0 \le p/2 - a \le p/2$ for all positive integers p and m and that for $s_j \le 1$ we have $u_j \ge s_j^{p/2}$. In the next example we choose the radius of the balls in the reduction operation such that at each step in the iteration the bound on the recombination error matches the bound on the KLV error.

EXAMPLE 21. Let $\gamma > m-1, m = r$, that is, the degree of the polynomials used in the reduction operation equals the degree of the cubature in the KLV method. Let u_j, $j = 2, \dots, k-1$ be given by

$$u_j = \left(\frac{s_j^{m+1}}{(T-t_j)^{m-rp}} \right)^{1/(2(r+1))}.$$

Then

$$\begin{aligned} &\sup_x \left| P_T f(x) - E_{Q^{(k)}_{\mathcal{D},u}(x)} f \right| \\ &\quad \le \left(C_1(T) \left(s_k^{1/2} + \sum_{i=1}^{k-1} \sum_{j=m}^{m+1} \frac{s_i^{(j+1)/2}}{(T-t_i)^{j/2}} \right) \right. \\ &\qquad \left. + C_5(T) \sum_{i=2}^{k-1} \frac{s_i^{(m+1)/2}}{(T-t_i)^{m/2}} \right) \|\nabla f\| \\ &\quad \le C_9 k^{-(m-1)/2} T^{1/2} \|\nabla f\|_\infty, \end{aligned} \tag{23}$$

where $C_9 = C_6(m, \gamma)(C_1(T) + C_5(T))$.

As before, if $T - t_1 < 1$, the constants C_8 and C_9 can be taken to be $C_6(m, \gamma)(C_1(1) + C_4)$. The parameters chosen in the above examples guarantee high order convergence, but are not necessarily computationally optimal. In the following section we examine how, for a fixed error ε, the choice of r and u can be varied to be closer to the optimal computational effort in the recombination operation.

4.3. *An optimization.* This paper establishes stable higher order particle approximation methods where the computational effort involved grows polynomially with the number of time steps when the number of steps is large and the underlying system remains compact (see Section 4.4). In concrete examples, an optimization of the different aspects of this algorithm, under the constraint of fixed total error, leads to even more effective approaches; although we expect that different problems would benefit from different distributions of the computational effort. For example, there is a trade-off between the degree of the polynomials that are used as test functions and the size of the balls used to define the localization of the measure for the recombination (smaller patches if we use higher degree polynomials in the test functions and we fix the error of the approximation).

Specifically, suppose we are given a discrete measure μ and the property we care about is the integral of μ against a smooth function g. As in our application to the KLV method we consider a reduced measure $\tilde{\mu}$ (Definition 1) with respect to the polynomials of degree at most r and a localization of μ with radius at most δ. The number of balls of radius δ required to cover the support of μ is at most of order $(\frac{D}{\delta})^N$, where D is the diameter of $\mathrm{supp}(\mu)$. Let ε be the error of the approximation of $\int g\, d\mu$ by $\int g\, d\tilde{\mu}$.

Note that

$$\varepsilon = \frac{\delta^{r+1} c_{r+1}}{(r+1)!}$$

for some $c_{r+1} \leq \sum_{i_1+\cdots+i_N=r+1} \| \frac{\partial^{i_1}}{\partial x_1^{i_1}} \cdots \frac{\partial^{i_N}}{\partial x_N^{i_N}} g \|_\infty$. Fixing the error ε gives a simple relation for δ and r

$$\delta = \left(\frac{\varepsilon (r+1)!}{c_{r+1}} \right)^{1/(r+1)}. \tag{24}$$

Let $\hat{n}$ be the number of particles in the support of μ. The computational complexity of the recombination operation as a function of δ, $\hat{n}$ and r is at most of order

$$\left(\frac{D}{\delta} \right)^N \binom{r+N}{N}^4 \log \hat{n} + \hat{n} \binom{r+N}{N}$$

which may be optimized subject to the constraint (24).

Note that in our application to cubature on Wiener, μ corresponds to $Q^{(j)}_{\mathcal{D},u}(x)$ and the function g is given by $P_{T-t_j} f$. The calculation above also allows us to decide after each step of the iteration if it is of computational benefit to carry out a (full) recombination operation.

4.4. *Simple bounds on the number of test functions*; *covering the support of the particle measures*. In this section we obtain upper bounds for the number of ODEs required to solve in the recombining KLV method with k iterations. For this, it is sufficient to bound the number of balls in the cover of the localizations of the particle measures uniformly for all k iterations. We first find a large ball $B(x, \rho)$ that covers $\mathrm{supp}(Q^{(j)}_{\mathcal{D},u}(x))$, $j = 1, \ldots, k-1$, and then estimate the number of balls that are required to cover $B(x, \rho)$. The balls in the covers of the localizations will have to be sufficiently small to preserve the high order accuracy of the method. We can show that under the assumption that the vector fields V_i are bounded and satisfy the UH condition, we have a high order method and the computational complexity is polynomial in k the number of iterations. Similar results can be obtained if the underlying system remains compact.

The following theorem demonstrates that we can achieve the same rate of convergence in the number of iterations k as in Kusuoka's algorithm and the vanilla

KLV method, but control the complexity of the method by an explicit polynomial in k. This compares to exponential growth in the vanilla KLV method without recombination, which despite its exponential growth leads to numerically highly effective algorithms (see, e.g., Ninomiya and Victoir [14]). The estimates in this section are not designed to be optimal and can be improved. Closer to optimal choices for the radius u_i and degree r in the reduction operation have been discussed in Section 4.3 and may be used to decide if it is computationally efficient to recombine the particle measure at time t_i.

THEOREM 22. *Suppose the uniform Hörmander condition is satisfied and the vector fields V_i are uniformly bounded by some constant $M' > 0$. We can achieve*

$$E_{\mathcal{D},k} = \sup_{x \in R^N} \left| P_T f(x) - E_{Q^{(k)}_{\mathcal{D},u}(x)} f \right| \le C_8 k^{-(m-1)/2} T^{1/2} \|\nabla f\|_\infty, \tag{25}$$

while the number of test functions in the reduction operation, and hence the number of elementary ODEs to solve grows polynomially in k.

PROOF. Let $m > 0$ be the degree of the cubature in the KLV method. Fix the partition $\mathcal{D}$ to (20) for some $\gamma > m - 1$. As in Example 20, let $r = \lceil m/p \rceil$ and $u_j = s_j^{p/2-a}$, $a = \frac{p-1}{2(\lceil m/p \rceil + 1)} \ge 0$ in the reduction operation. We note that the error $E_{\mathcal{D},k}$ satisfies (25) and it remains to show that the number of particles in support of the measures $Q^{(k)}_{\mathcal{D},u}(x)$ grows polynomially in k, which is equivalent to the number of balls in the localizations growing polynomially in k.

Note that if $\omega \in C_0^0([0,1], R^d)$ is a continuous path of bounded variation of length L, we have

$$|x - \Phi_{1,x}(\omega)| \le M' L,$$

where Φ is the Itô functional defined in (6), that is, $\Phi_{1,x}(\omega)$ is the point we obtain by solving the equation (5) along the path ω starting at x. Let L be given by $L = \max_{i=1,\dots,n} \text{length}(\omega_i)$, the maximum of the lengths of the paths in the support of the degree m cubature formula on Wiener space over the unit time interval. Observe that by construction any particle in the support of $Q^{(j)}_{\mathcal{D},u}(x)$ [compare the definition of the measures in (18)] may be written as

$$\Phi_{\sum_{i=1}^j s_i, x}(\omega_{s_1,i_1} \otimes \cdots \otimes \omega_{s_j,i_j})$$

some $i_1, \dots, i_j \in \{1, \dots, n\}$, the $\omega_{s,i}$ are the rescaled paths defined in (8) and $\otimes$ denotes to the concatenation of paths. For k sufficiently large we may assume $s_i < 1$ and we deduce that

$$\text{supp}(Q^{(j)}_{\mathcal{D},u}(x)) \subseteq B\left(x, M' L \sum_{i=1}^j s_i^{1/2}\right) \subseteq B(x, M' L k T^{1/2}).$$

In the reduction operations we consider a basis of the polynomials of degree at most $\lceil m/p \rceil$ and the measure is localized by balls of radius u_j which need to cover $\mathrm{supp}(Q^{(j)}_{\mathcal{D},u}(x))$. For $s_j < 1$, that is, for k sufficiently large, we have $u_j \geq s_j^{p/2}$ and for our uneven family of partitions $\min_{j=2,\ldots,k-1} s_j^{p/2} < s_k^{p/2} = (T/k^\gamma)^{p/2}$. Thus, the number of particles in each of the reduced measures is uniformly bounded above by $\binom{\lceil m/p \rceil + N}{N}$ times the number of balls of radius $(T/k^\gamma)^{p/2}$ required to cover the ball $B(x, M'LkT^{1/2})$ in N-dimensional space, which is a polynomial of degree at most $N(\gamma p/2 + 1)$ in k. $\square$

Similarly, we can derive a result analogous to Theorem 22 if the underlying system remains compact.

APPENDIX: A NUMERICAL TOY EXAMPLE

We consider a linear one-dimensional problem. The boundary data is Lipschitz, piecewise smooth, and the locations of the discontinuities in the derivatives are not known to the program. The answer is required to high accuracy. In our test case we applied the approximation method to the heat equation with boundary data

$$f(x) = \max(1 - e^x, 0),$$

which corresponds to the calculation of a Black–Scholes put option at logarithmic scale. We considered a time horizon of $T = 1$ and various initial conditions $X_0 \in [-4, 4]$. We set our goal to achieve an accuracy of 10^{-10}. This example is particularly suitable as a test example because the solution to the equation is known in closed form in terms of well known special functions which can be used to determine the precise error in the approximation.

We applied a modified form of the KLV method with recombination introduced in this paper. For $\theta < 1$ consider a geometrically converging partition of the unit time interval given by

$$1 - t_j = (1 - \theta)(1 - t_{j-1}), \qquad j = 1, \ldots, k - 1,$$

$t_0 = 0$ and $t_k = 1$. Note that the length of the time steps s_j in the partition is given by $s_j = \theta(1 - t_{j-1})$. In our particular example we chose θ to be 0.4. To achieve the required accuracy we used a 15 point Gaussian quadrature which we had previously computed to high accuracy. For the heat equation, the particles of the cubature approximation are given by the Gaussian quadrature and we do not require to solve ODEs. As described in Section 4.1 we used polynomial test functions of degree m and localized the support of intermediate particle measures in the approximation. We then used a heuristic based on the information provided by the $W^{1,1}$ norm of f to determine, as outlined in Section 4.3, the degree of polynomial approximation that minimizes the computational complexity of the overall reduction process subject to achieving the required accuracy.

TABLE 1
Absolute error and computational effort for the approximation of $u(x, 1)$ for different values of x

X_0		-4	-3	-2	-1
Absolute error ε		3.186E–11	1.01E–11	4.962E–11	1.2014E–10
Evaluations at the boundary		1,410,075	1,416,600	1,426,050	1,432,350
Particles		94,005	94,440	95,070	95,490
X_0	**0**	**1**	**2**	**3**	**4**
ε	3.612E–11	4.173E–12	2.52E–11	5.47E–11	5.62E–12
Evaluations	1,430,775	1,425,600	1,424,700	1,417,725	1,418,175
Particles	95,385	95,040	94,980	94,515	94,545

In addition, we modified the algorithm to make use of the piecewise smooth nature of the boundary data. The algorithm compares for each particle a two step KLV with a one step KLV estimate to the boundary. If both approximates agree to the error tolerance, the algorithm immediately leaps to the boundary. As the required accuracy is close to machine precision, false positives are very unlikely. Recombination is then performed on the remaining particles.

In order to achieve an accuracy of 10^{-10} we chose $m = 8$ and a radius for the localization that was proportional to $\sqrt{1-t}$ and covered the surviving measure with approximately 13 nonempty components in the localization. The runtime of our single threaded C++ code[5] was between 0.5 and 0.6 s. The parameter restricting the maximal depth of the approximation tree was set to 28. Table 1 and Figure 1 summarize the absolute error of the approximation, the number of reduced particles inside the domain and the total number of evaluations of the cubature at the boundary for various values of X_0. Note that the number of particles compares to $\sim 15^{27}$ internal particles for the vanilla cubature algorithm and even if combined with a partial sampling scheme such as the tree based branching algorithm one could not hope to compute an approximation to ten digit accuracy.

Even though the problem we have considered is merely a toy example, computing the solution to high accuracy with a vanilla off the shelf PDE solver appears to be nontrivial. However, a fair comparison must involve at least adaptive methods; we were afraid to do this ourselves as it would not carry much weight because we do not have the computational expertise to get good outcomes from these packages. So we were very grateful that our colleague Kathryn Gillow in Oxford was willing to give it a quick spin on adaptive software she had developed with Endre Suli.

She says: "I've now tried a few approaches to solving your problem but can't get results even close to yours in terms of accuracy achieved in such a small amount of

[5]As measured on a Lenovo Thinkpad x201t notebook computer. We used intel mkl for the lapack support and this might use omp internally.

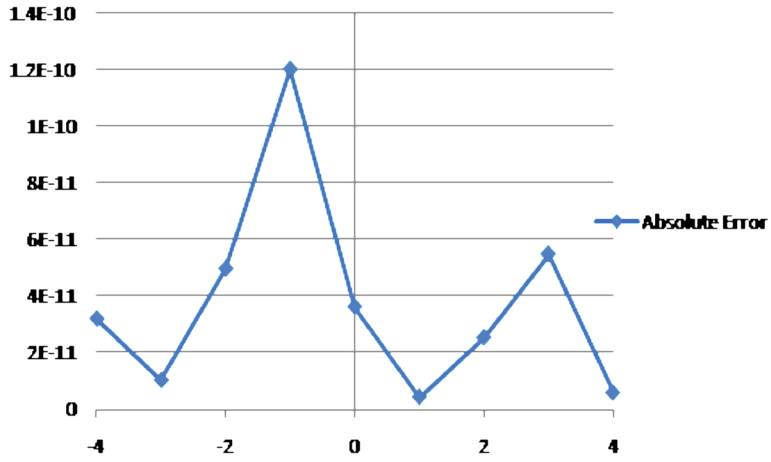

FIG. 1. *Absolute error for the approximation of* $u(x, 1)$ *for different values of* x.

CPU time. In all cases I've solved the heat equation on the spatial interval $-9.9 < x < 10.1$ (so that with a coarse uniform mesh the point $x = 0$ was not a node). Then to look at the error I have computed the solution at time 1 and for x integer between -5 and 5 as you suggest. The first approach I took was to do an adaptive finite element solution with the adaptivity geared toward getting an accurate solution at time $t = 1$. The mesh can change at every timestep which is obviously less than ideal as you then need to keep recomputing the matrices. The code is taking about 30 seconds and giving accuracy of between 10^{-4} and 10^{-7} depending on which integer you look at. It actually turns out to be more efficient to do something a bit more naive, namely, to adapt the mesh to resolve the initial condition well and then use that mesh for the rest of the computation. As expected, this clusters the nodes around $x = 0$ and the mesh is fairly coarse elsewhere. The advantage of this is that you just solve the same matrix problem at every time-step. This speeds things up a lot without degrading the accuracy for this problem. So here I'm getting accuracy of between 10^{-4} and 10^{-6} in about 1 second. Then, finally, I gave Nick Trefethen et al's Matlab package Chebfun a go. In order to solve the heat equation which exploits the fact that the problem is linear so you can write the solution at a given time t as $\exp(t * L)u_0$ where L is the spatial operator (including boundary conditions) and u_0 is the initial condition. It seems that Chebfun struggles when u_0 is not smooth and it actually turns out to be more efficient to compute the solution at time $t = 1$ in two stages, namely, $u(x, dt) = \exp(dt * L)u_0$, $u(x, 1) = \exp((1 - dt) * L) * u(x, dt)$. The best accuracy using this approach is $5 * 10^{-6}$ taking 6.5 seconds. Chebfun does a lot better when you have smooth initial data. Then it can solve the same type of problem in 0.1 s giving errors of 10^{-7}."

No doubt the approach we take tries to do less than that taken by our colleagues, (it only computes the solution at the required points, etc.) and we have tried to pol-

ish the code for our problem but still we find it encouraging evidence that this paper is putting ideas together in a novel way. The linear algebra we do is numerically really heavy, but it seems to pay.

REFERENCES

[1] BAYER, C. and TEICHMANN, J. (2006). The proof of Tchakaloff's theorem. *Proc. Amer. Math. Soc.* **134** 3035–3040 (electronic). MR2231629

[2] CRISAN, D. (2002). Numerical methods for solving the stochastic filtering problem. In *Numerical Methods and Stochastics* (*Toronto*, *ON*, 1999) (T. J. Lyons, T. S. Salisbury, eds.). *Fields Institute Communications* **34** 1–20. Amer. Math. Soc., Providence, RI. MR1944741

[3] CRISAN, D. and DOUCET, A. (2002). A survey of convergence results on particle filtering methods for practitioners. *IEEE Trans. Signal Process.* **50** 736–746. MR1895071

[4] CRISAN, D. and GHAZALI, S. (2007). On the convergence rates of a general class of weak approximations of SDEs. In *Stochastic Differential Equations*: *Theory and Applications*. *Interdisciplinary Mathematical Sciences* **2** 221–248. World Scientific, Hackensack, NJ. MR2393578

[5] DAVIS, P. J. (1967). A construction of nonnegative approximate quadratures. *Math. Comp.* **21** 578–582. MR0222534

[6] KUSUOKA, S. (2001). Approximation of expectation of diffusion process and mathematical finance. In *Taniguchi Conference on Mathematics Nara '98*. *Advanced Studies in Pure Mathematics* **31** 147–165. Math. Soc. Japan, Tokyo. MR1865091

[7] KUSUOKA, S. (2003). Malliavin calculus revisited. *J. Math. Sci. Univ. Tokyo* **10** 261–277. MR1987133

[8] KUSUOKA, S. and STROOCK, D. (1987). Applications of the Malliavin calculus. III. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **34** 391–442. MR0914028

[9] LITTERER, C. (2008). The signature in numerical algorithms. Ph.D. thesis, Univ. Oxford.

[10] LITTERER, C. and LYONS, T. (2011). Cubature on Wiener space continued. In *Stochastic Processes and Application to Mathematical Finance*, *Proceedings of the 6th Ritsumeikan International Symposium*. Oxford Univ. Press, Oxford.

[11] LITTERER, C. and LYONS, T. (2011). Introducing cubature to filtering. In *Oxford Handbook of Non-Linear Filtering* (D. Crisan and B. Rozovsky, eds.) 786–798. Oxford Univ. Press, Oxford.

[12] LYONS, T. and VICTOIR, N. (2004). Cubature on Wiener space. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **460** 169–198. MR2052260

[13] NINOMIYA, S. (2003). A partial sampling method applied to the Kusuoka approximation. *Monte Carlo Methods Appl.* **9** 27–38. MR1987415

[14] NINOMIYA, S. and VICTOIR, N. (2008). Weak approximation of stochastic differential equations and application to derivative pricing. *Appl. Math. Finance* **15** 107–121. MR2409419

[15] VICTOIR, N. Private communication.

[16] WENDEL, J. G. (1962). A problem in geometric probability. *Math. Scand.* **11** 109–111. MR0146858

MATHEMATICAL INSTITUTE
AND
OXFORD-MAN INSTITUTE
OF QUANTITATIVE FINANCE
UNIVERSITY OF OXFORD
24-29 ST GILES'
OXFORD, OX1 3LB
UNITED KINGDOM