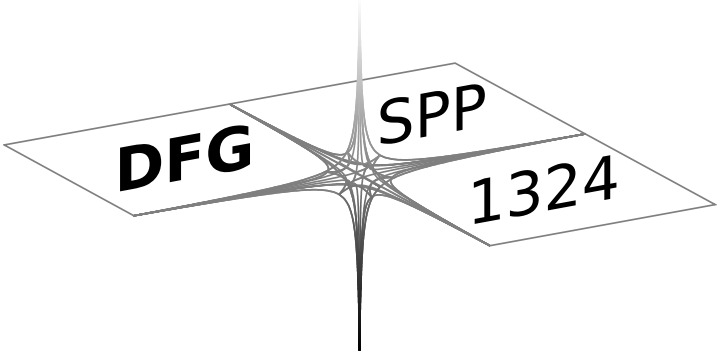
DFG-Schwerpunktprogramm 1324

„Extraktion quantifizierbarer Information aus komplexen Systemen”

Variable Subspace Sampling and Multi-level Algorithms

T. Müller-Gronbach, K. Ritter

Preprint 14



Edited by

AG Numerik/Optimierung
Fachbereich 12 - Mathematik und Informatik
Philipps-Universität Marburg
Hans-Meerwein-Str.
35032 Marburg

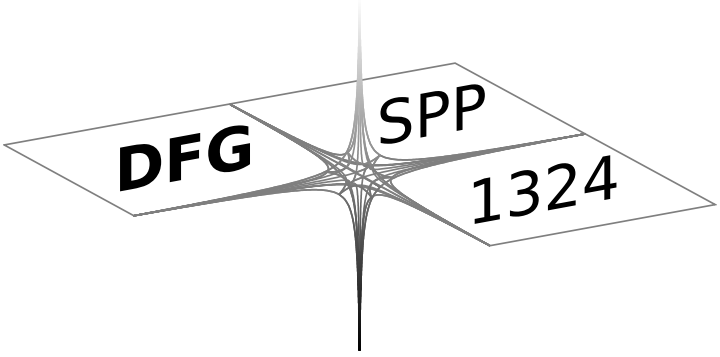
DFG-Schwerpunktprogramm 1324

„Extraktion quantifizierbarer Information aus komplexen Systemen”

Variable Subspace Sampling and Multi-level Algorithms

T. Müller-Gronbach, K. Ritter

Preprint 14



The consecutive numbering of the publications is determined by their chronological order.

The aim of this preprint series is to make new research rapidly available for scientific discussion. Therefore, the responsibility for the contents is solely due to the authors. The publications will be distributed by the authors.

Variable Subspace Sampling and Multi-level Algorithms

Thomas Müller-Gronbach[1] and Klaus Ritter[2]

[1] Fakultät für Informatik und Mathematik
Universität Passau
Innstraße 33
94032 Passau
Germany

[2] Fachbereich Mathematik
Technische Universität Darmstadt
Schloßgartenstraße 7
64289 Darmstadt
Germany
http://www.mathematik.tu-darmstadt.de/~ritter

Summary. We survey recent results on numerical integration with respect to measures μ on infinite-dimensional spaces, e.g., Gaussian measures on function spaces or distributions of diffusion processes on the path space. Emphasis is given to the class of multi-level Monte Carlo algorithms and, more generally, to variable subspace sampling and the associated cost model. In particular we investigate integration of Lipschitz functionals. Here we establish a close relation between quadrature by means of randomized algorithms and Kolmogorov widths and quantization numbers of μ. Suitable multi-level algorithms turn out to be almost optimal in the Gaussian case and in the diffusion case.

1 Introduction

Let μ be a Borel probability measure on a Banach space $(\mathfrak{X}, \|\cdot\|_{\mathfrak{X}})$, and let F denote a class of μ-integrable functionals $f : \mathfrak{X} \to \mathbb{R}$. In the corresponding quadrature problem we wish to compute

$$S(f) = \int_{\mathfrak{X}} f(x)\, \mu(dx)$$

for $f \in F$ by means of randomized (Monte Carlo) algorithms that use the values $f(x)$ of the functional f at a finite number of sequentially (adaptively) chosen points $x \in \mathfrak{X}$.

The classical instance of this quadrature problem is given by $\mathfrak{X} = \mathbb{R}^d$ and μ being the uniform distribution on $[0,1]^d$, say, or the d-dimensional standard normal distribution. In the present paper we are mainly interested in infinite-dimensional spaces $\mathfrak{X}$. The classical instance of infinite-dimensional quadrature is path integration with respect to the Wiener measure μ on $\mathfrak{X} = C([0,1])$ or, more generally, quadrature with respect to a Gaussian measure μ on a function space $\mathfrak{X}$.

Further important instances of quadrature problems arise for stochastic (partial) differential equations, and here the measure μ is usually given only implicitly, since it depends on the solution process of the equation. We have $\dim(\mathfrak{X}) < \infty$ if μ is a marginal distribution of the solution of an SDE and $\dim(\mathfrak{X}) = \infty$ for quadrature on the path space. For SPDEs, both the marginal and the path dependent case lead to infinite-dimensional quadrature problems.

The present paper is motivated by the following developments. On the one hand a new class of algorithms, namely multi-level Monte Carlo algorithms, has been introduced by Heinrich [18] and Giles [14]. On the other hand infinite-dimensional quadrature problems have been studied from a complexity point of view by Wasilkowski and Woźniakowski [37] and Hickernell and Wang [21]. The purpose of this paper is to illustrate the approach and the results from [5], which provides a link between the two developments and which establishes the concept of approximation of distributions as the basis for integration of Lipschitz functionals f on infinite-dimensional spaces $\mathfrak{X}$. Furthermore, we provide a continuation of the survey paper [30] on strong and weak approximation of SDEs with a new focus on multi-level Monte Carlo algorithms.

The content of this paper is organized as follows. In Section 2 we present multi-level Monte Carlo algorithms in general terms together with the particular case of multi-level Euler Monte Carlo algorithms for SDEs, which serve as a basic example in the sequel.

Section 3 is devoted to the presentation of a reasonable cost model for the analysis of infinite-dimensional quadrature problems. We distinguish between full space sampling, variable subspace sampling, and fixed subspace sampling. In the latter case an algorithm may only evaluate the integrands f at the points in a finite-dimensional subspace $\mathfrak{X}_0 \subset \mathfrak{X}$, which may be chosen arbitrarily but which is fixed for a specific algorithm. We add that fixed subspace sampling is frequently used for infinite-dimensional quadrature problems. In contrast, a multi-level algorithm uses dependent samples in a hierarchy of finite-dimensional subspaces $\mathfrak{X}_1 \subset \mathfrak{X}_2 \subset \ldots \subset \mathfrak{X}$ with only a small proportion taken in high-dimensional spaces. For both variants of subspace sampling the cost per evaluation at $x \in \mathfrak{X}$ is given by the dimension of the (minimal) subspace containing x. Full space sampling permits evaluations anywhere in $\mathfrak{X}$ at cost one, which is perfectly reasonable for finite-dimensional quadrature problems; in the infinite-dimensional case its main purpose is to establish lower bounds.

Section 4 contains an analysis of multi-level algorithms. In the particular case of Lipschitz continuous integrands we provide upper error bounds for

these algorithms in terms of average Kolmogorov widths of the underlying measure μ, see Theorem 3.

In Section 5 we introduce the concept of minimal errors, which allows a rigorous comparison of the power of full space sampling, variable subspace sampling, and fixed subspace sampling.

In Section 6 we focus on the Lipschitz case and we present upper and lower bounds for the minimal errors in terms of average Kolmogorov widths and quantization numbers. Since the latter two quantities can equivalently be defined in terms of the Wasserstein distance on the space of Borel probability measures on $\mathfrak{X}$ our error estimates exhibit the tight connection between quadrature by means of randomized algorithms and approximation of the underlying measure μ by means of probability measures with a finite-dimensional or finite support. These results are applied to quadrature with respect to Gaussian measures μ and with respect to the distributions of the solutions of SDEs. Suitable multi-level algorithms turn out to be almost optimal in both cases.

2 Multi-level Algorithms

Multi-level Monte Carlo methods have been introduced by Heinrich [18] and Heinrich and Sindambiwe [20] for computation of global solutions of integral equations and for parametric integration, respectively. Moreover, the authors have shown that suitable multi-level algorithms are (almost) optimal in both cases. See [19] for further results and references. In their work finite-dimensional quadrature problems arise as subproblems and the Monte Carlo methods take values in infinite-dimensional Banach spaces $\mathfrak{Y}$. Here we are interested in the dual situation of infinite-dimensional quadrature with real numbers as outputs of Monte Carlo algorithms, i.e., we have $\mathfrak{Y} = \mathbb{R}$.

In the context of quadrature problems for diffusion processes multilevel algorithms have been introduced by Giles [14], while a two level algorithm has already been considered by Kebaier [22]. Both papers also include numerical examples from computational finance, see also [15, 16].

In order to describe the multi-level approach in general terms it is convenient to assume that μ is the distribution of an $\mathfrak{X}$-valued random element of the form

$$X = \varphi(\widetilde{X})$$

for some random element $\widetilde{X}$ taking values in a Banach space $(\widetilde{\mathfrak{X}}, \|\cdot\|_{\widetilde{\mathfrak{X}}})$ and some measurable mapping

$$\varphi : \widetilde{\mathfrak{X}} \to \mathfrak{X}.$$

As the key assumption we suppose that we have a sequence of measurable mappings

$$\varphi^{(k)} : \widetilde{\mathfrak{X}} \to \mathfrak{X}$$

at hand, which provide approximations

$$X^{(k)} = \varphi^{(k)}(\widetilde{X})$$

to X. Hence

$$\mathrm{E}(f(X^{(k)})) = \mathrm{E}(f(\varphi^{(k)}(\widetilde{X})))$$

may serve as an approximation to

$$S(f) = \mathrm{E}(f(X)) = \mathrm{E}(f(\varphi(\widetilde{X}))).$$

Example 1. A typical example is provided by an SDE

$$dX(t) = a(t, X(t))\,dt + b(t, X(t))\,d\widetilde{X}(t), \qquad t \in [0,1],$$

with initial value

$$X(0) = \xi \in \mathbb{R}^m,$$

drift coefficient

$$a : [0,1] \times \mathbb{R}^m \to \mathbb{R}^m,$$

diffusion coefficient

$$b : [0,1] \times \mathbb{R}^m \to \mathbb{R}^{m \times d},$$

and with a d-dimensional Brownian motion $\widetilde{X}$. In this case $\widetilde{\mathfrak{X}} = C([0,1], \mathbb{R}^d)$ and $\mathfrak{X} = C([0,1], \mathbb{R}^m)$ are the spaces of continuous functions on $[0,1]$ taking values in $\mathbb{R}^d$ and $\mathbb{R}^m$, respectively, and φ maps the driving Brownian motion $\widetilde{X}$ to the solution process X. The mapping $\varphi^{(k)}$ may correspond to the piecewise linear interpolation of the Euler scheme with step size $\delta^{(k)} = 2^{-(k-1)}$. The time discretization is then given by

$$t_i^{(k)} = i\,\delta^{(k)}, \qquad i = 0, \ldots, 2^{k-1},$$

and we have

$$X_0^{(k)} = \xi$$

and

$$X_{i+1}^{(k)} = X_i^{(k)} + a(t_i^{(k)}, X_i^{(k)})\,\delta^{(k)} + b(t_i^{(k)}, X_i^{(k)})\,\big(\widetilde{X}(t_{i+1}^{(k)}) - \widetilde{X}(t_i^{(k)})\big). \tag{1}$$

Finally, the random element $X^{(k)} = \varphi^{(k)}(\widetilde{X})$ is given by the piecewise linear interpolation of $X_0^{(k)}, \ldots, X_{2^{k-1}}^{(k)}$ at the nodes $t_0^{(k)}, \ldots, t_{2^{k-1}}^{(k)}$.

For a Gaussian measure μ on $\mathfrak{X}$ it is reasonable to take $\widetilde{\mathfrak{X}} = \mathfrak{X}$, $\widetilde{X} = X$, and the identity function φ. Metric projections $\varphi^{(k)}$ onto an increasing sequence of finite-dimensional subspaces of $\mathfrak{X}$ may be used for approximation of X.

The classical Monte Carlo approximation to $\mathrm{E}(f(X^{(k)}))$ is based on independent copies $\widetilde{X}_1, \ldots, \widetilde{X}_n$ of $\widetilde{X}$ and given by the random variable

$$A(f) = \frac{1}{n} \sum_{\ell=1}^{n} f(\varphi^{(k)}(\widetilde{X}_\ell)). \tag{2}$$

For its mean square error we clearly have

$$\mathrm{E}(S(f) - A(f))^2 = \frac{1}{n}\ \mathrm{Var}(f(X^{(k)})) + b_k^2(f) \tag{3}$$

with the bias

$$b_k(f) = \mathrm{E}(f(X^{(k)})) - S(f).$$

The actual computation of a realization of $A(f)$ requires simulation of the distribution of $X^{(k)}$ and evaluation of f at randomly chosen points from the range of $\varphi^{(k)}$.

Note that

$$\mathrm{E}(f(X^{(k)})) = \mathrm{E}(f(X^{(1)})) + \sum_{j=2}^{k} \mathrm{E}(f(X^{(j)}) - f(X^{(j-1)})),$$

where $f(X^{(j)}) = f(\varphi^{(j)}(\widetilde{X}))$ and $f(X^{(j-1)}) = f(\varphi^{(j-1)}(\widetilde{X}))$ are coupled via $\widetilde{X}$. In the multi-level approach each of the expectations on the right-hand side is approximated separately by means of independent, classical Monte Carlo approximations. With $n_1, \ldots, n_k$ denoting the corresponding numbers of replications and with independent copies

$$\widetilde{X}_{j,1}, \ldots, \widetilde{X}_{j,n_j}, \qquad j = 1, \ldots, k,$$

of $\widetilde{X}$ the multi-level approximation is given by the random variable

$$A_k(f) = A^{(1)}(f) + \sum_{j=2}^{k} A^{(j)}(f) \tag{4}$$

where

$$A^{(1)}(f) = \frac{1}{n_1} \sum_{\ell=1}^{n_1} f(\varphi^{(1)}(\widetilde{X}_{1,\ell})) \tag{5}$$

and

$$A^{(j)}(f) = \frac{1}{n_j} \sum_{\ell=1}^{n_j} (f(\varphi^{(j)}(\widetilde{X}_{j,\ell})) - f(\varphi^{(j-1)}(\widetilde{X}_{j,\ell}))) \tag{6}$$

for $j = 2, \ldots, k$. For the mean square error of $A_k(f)$ we get

$$\mathrm{E}(S(f) - A_k(f))^2 = \sum_{j=1}^{k} \frac{v_j(f)}{n_j} + b_k^2(f) \tag{7}$$

where

$$v_1(f) = \mathrm{Var}(f(X^{(1)}))$$

and

$$v_j(f) = \mathrm{Var}(f(X^{(j)}) - f(X^{(j-1)}))$$

for $j = 2, \dots, k$. The actual computation of a realization of $A_k(f)$ requires simulation of the distribution of $X^{(1)}$ and the joint distribution of $X^{(j)}$ and $X^{(j-1)}$ for $j = 2, \dots, k$. Furthermore, evaluation of f at randomly chosen points from the ranges of $\varphi^{(1)}, \dots, \varphi^{(k)}$ is needed.

Typically the variances $v_j(f)$ and the bias $b_k(f)$ are decreasing with increasing values of j and k, respectively, while the computational cost is increasing. One therefore has to properly balance these effects. A comparison of (3) and (7) reveals that the multi-level approach is a variance reduction technique.

Remark 1. The error formula (7) is a consequence of Bienaymé's equality for real-valued random variables, which does not extend to general Banach spaces. Thus, for the analysis of multi-level algorithms taking values in such a space the so-called Rademacher type of this space plays an important role, see [18, 19, 20].

Example 2. Let us present the details for a multi-level Euler algorithm in the case of an SDE, see Example 1. For notational convenience we consider a scalar equation, i.e., $m = d = 1$. We use $\stackrel{\mathrm{d}}{=}$ to denote equality in distribution of two random elements.

The simulation of $\varphi^{(1)}(\widetilde{X}_{1,\ell})$ and $(\varphi^{(j)}(\widetilde{X}_{j,\ell}), \varphi^{(j-1)}(\widetilde{X}_{j,\ell}))$ in (5) and (6) may be based on i.i.d. standard normally distributed random variables $Z_{i,\ell}^{(j)}$ for $j = 1, \dots, k$, $\ell = 1, \dots, n_j$, and $i = 1, \dots, 2^{j-1}$ as follows. We put

$$U_{0,\ell}^{(j)} = \xi$$

as well as

$$U_{i+1,\ell}^{(j)} = U_{i,\ell}^{(j)} + a(t_i^{(j)}, U_{i,\ell}^{(j)})\, \delta^{(j)} + b(t_i^{(j)}, U_{i,\ell}^{(j)})\, \sqrt{\delta^{(j)}}\, Z_{i+1,\ell}^{(j)}$$

for $i = 0, \dots, 2^{j-1} - 1$, cf. (1). Furthermore, if $j > 1$, we put

$$V_{0,\ell}^{(j)} = \xi$$

as well as

$$\begin{aligned} V_{i+1,\ell}^{(j)} = V_{i,\ell}^{(j)} &+ a(t_i^{(j-1)}, V_{i,\ell}^{(j)})\, \delta^{(j-1)} \\ &+ b(t_i^{(j-1)}, V_{i,\ell}^{(j)})\, \sqrt{\delta^{(j-1)}/2}\, \bigl(Z_{2i+1,\ell}^{(j)} + Z_{2i+2,\ell}^{(j)}\bigr) \end{aligned}$$

for $i = 0, \dots, 2^{j-2} - 1$.

We stress that the corresponding piecewise linear interpolations $U_\ell^{(j)}$ and $V_\ell^{(j)}$, respectively, are coupled, since they are based on the same random vector $(Z_{1,\ell}^{(j)}, \dots, Z_{2^{j-1},\ell}^{(j)})$. On the other hand, $U_\ell^{(j-1)}$ and $V_{\ell'}^{(j)}$ are independent with $U_\ell^{(j-1)} \stackrel{\mathrm{d}}{=} V_{\ell'}^{(j)} \stackrel{\mathrm{d}}{=} X^{(j)}$. Altogether we obtain independent random elements

$$U_1^{(1)}, \dots, U_{n_1}^{(1)},$$
$$(U_1^{(2)}, V_1^{(2)}), \dots, (U_{n_2}^{(2)}, V_{n_2}^{(2)}),$$
$$\dots$$
$$(U_1^{(k)}, V_1^{(k)}), \dots, (U_{n_k}^{(k)}, V_{n_k}^{(k)})$$

taking values in $C([0,1])$ or $(C([0,1]))^2$, respectively, whose distributions satisfy

$$U_\ell^{(1)} \stackrel{\mathrm{d}}{=} X^{(1)}$$

and

$$(U_\ell^{(j)}, V_\ell^{(j)}) \stackrel{\mathrm{d}}{=} (X^{(j)}, X^{(j-1)}).$$

Consequently

$$A_k(f) \stackrel{\mathrm{d}}{=} \frac{1}{n_1} \sum_{\ell=1}^{n_1} f(U_\ell^{(1)}) + \sum_{j=2}^{k} \frac{1}{n_j} \sum_{\ell=1}^{n_j} (f(U_\ell^{(j)}) - f(V_\ell^{(j)})).$$

We add that scaling of step sizes for the Euler scheme has already been used in a bias reduction technique by means of extrapolation, see [2, 35].

3 A Cost Model for Variable Subspace Sampling

In this section we present a cost model for the analysis of multi-level algorithms and, more generally, for the complexity analysis of infinite-dimensional quadrature problems. See [5] for details.

We assume that algorithms for the approximation of $S(f)$ have access to the functionals $f \in F$ via an oracle (subroutine) that provides values $f(x)$ for points $x \in \mathfrak{X}$ or a subset thereof. The cost per evaluation (oracle call) is modelled by a measurable function

$$c : \mathfrak{X} \to \mathbb{N} \cup \{\infty\}.$$

We define the cost of a computation as the sum of the cost of all oracle calls that are made during the computation. For a randomized algorithm the cost defines a random variable (under mild measurability assumptions), which may also depend on $f \in F$. This random variable is henceforth denoted by $\mathrm{cost}_c(A, f)$.

Let us look at the particular case of a randomized quadrature formula

$$A(f) = \sum_{\ell=1}^{n} a_\ell \, f(X_\ell) \tag{8}$$

with deterministic weights $a_\ell \in \mathbb{R}$ and random elements X_ℓ taking values in $\mathfrak{X}$. This class of randomized algorithms obviously contains every Monte Carlo

method (2) and every multi-level algorithm (4), where $n = n_1 + 2\sum_{j=2}^{k} n_j$ in the latter case. The cost of a randomized quadrature formula is given by

$$\mathrm{cost}_c(A, f) = \sum_{\ell=1}^{n} c(X_\ell).$$

Now we discuss specific choices of c. In the cost model given by

$$c = 1 \tag{9}$$

evaluation of an integrand f is possible at any point $x \in \mathfrak{X}$ at cost one. In this model, which is called *full space sampling*, $\mathrm{cost}_c(A, f)$ is the number of evaluations of the integrand. For finite-dimensional quadrature, i.e., if $\dim(\mathfrak{X}) < \infty$, full space sampling is the common choice in the literature.

However, if $\dim(\mathfrak{X}) = \infty$, then full space sampling seems to be too generous and therefore of limited practical relevance. It is more reasonable and partially motivated by the multi-level construction to consider *variable subspace sampling* instead. In any such model we consider a sequence of finite-dimensional subspaces

$$\{0\} \subsetneq \mathfrak{X}_1 \subset \mathfrak{X}_2 \subset \ldots \subset \mathfrak{X},$$

and we define the cost function c by

$$c(x) = \inf\{\dim(\mathfrak{X}_j) : x \in \mathfrak{X}_j\}. \tag{10}$$

In particular, in the setting of a multi-level algorithm (4),

$$\mathfrak{X}_j = \mathrm{span}\left(\bigcup_{i=1}^{j} \varphi^{(i)}(\widetilde{\mathfrak{X}})\right) \tag{11}$$

is a natural choice, and the cost of this algorithm then satisfies

$$\mathrm{cost}_c(A_k, f) \le n_1 \, \dim(\mathfrak{X}_1) + \sum_{j=2}^{k} n_j \, (\dim(\mathfrak{X}_j) + \dim(\mathfrak{X}_{j-1})) \tag{12}$$

in the corresponding variable subspace model.

We write $x_k \preceq y_k$ for sequences of positive real numbers x_k and y_k, if $x_k \le \gamma \, y_k$ holds for every $k \in \mathbb{N}$ with a constant $\gamma > 0$. Furthermore, $x_k \asymp y_k$ means $x_k \preceq y_k$ and $y_k \preceq x_k$.

Example 3. Consider the spaces $\mathfrak{X}_j$ according to (11) in the setting from Example 2. Here, $\mathfrak{X}_j = \mathrm{span}(\varphi^{(j)}(\widetilde{\mathfrak{X}}))$ is the space of piecewise linear functions in $\mathfrak{X} = C([0,1])$ with equidistant breakpoints $i\,2^{-(j-1)}$ and we have $\dim(\mathfrak{X}_j) = 2^{j-1} + 1$. It follows that $1/n_j \sum_{\ell=1}^{n_j} f(U_\ell^{(j)})$ can be computed at cost $n_j \, (2^{j-1} + 1)$, and we get

$$\mathrm{cost}_c(A_k, f) \le n_1\, 2 + \sum_{j=2}^{k} n_j\, (2^{j-1} + 2^{j-2} + 2) \asymp \sum_{j=1}^{k} 2^j\, n_j \tag{13}$$

for the multi-level algorithm (4).

Finally we discuss *fixed subspace sampling*. In this case, evaluations are possible only at points in a finite-dimensional subspace

$$\{0\} \subsetneq \mathfrak{X}_0 \subset \mathfrak{X}.$$

For every such evaluation its cost is given by $\dim(\mathfrak{X}_0)$. Thus

$$c(x) = \begin{cases} \dim(\mathfrak{X}_0), & \text{if } x \in \mathfrak{X}_0, \\ \infty, & \text{otherwise.} \end{cases} \tag{14}$$

Clearly fixed subspace sampling constitutes a particular case of variable subspace sampling.

For both kinds of subspace sampling we think of bases associated to the subspaces, so that $c(x)$ is the (minimal) number of real coefficients needed to represent x and this representation is actually submitted to the oracle.

Example 4. Obviously the multi-level Euler algorithm from Example 2 may also be analyzed in the fixed subspace model defined by $\mathfrak{X}_0 = \mathrm{span}(\varphi^{(k)}(\widetilde{\mathfrak{X}}))$, which leads to

$$\mathrm{cost}_c(A_k, f) \le \sum_{j=1}^{k} n_j\, (2^{k-1} + 1) \asymp 2^k \sum_{j=1}^{k} n_j.$$

This analysis, however, would be inadequate, since it does not capture the fact that a large proportion of samples is taken in low-dimensional spaces.

Remark 2. We stress that $\mathrm{cost}_c(A, f)$ is a rough measure of the computational cost for applying the algorithm A to the integrand f, since it only takes into account the information cost, which is caused by the evaluations of f. All further operations needed to compute a realization of $A(f)$ are not considered at all.

In a more detailed analysis it is appropriate to take the real number model of computation as a basis for quadrature problems. See [32, 36] for the definition of this model. Informally, a real number algorithm is like a C-program that carries out exact computations with real numbers. Furthermore, a perfect generator for random numbers from $[0, 1]$ is available and elementary functions like exp, ln, etc. can be evaluated. Finally, algorithms have access to the integrands $f \in F$ via the oracle (subroutine). We think that these assumptions are present at least implicitly in most of the work dealing with quadrature problems.

For simplicity we assume that real number operations as well as calls of the random number generator and evaluations of elementary functions are performed at cost one. Furthermore, in case of μ being the distribution of a diffusion process, function values of its drift and diffusion coefficients are provided at cost one, too. Then the total cost of a computation is given by $\mathrm{cost}_c(A, f)$ plus the total number of real number operations, calls of the random number generator, evaluations of elementary functions, and, eventually, function evaluations of drift and diffusion coefficients.

Example 5. In the analysis according to Remark 2 the right-hand side in (13) still is an upper bound for the cost of the multi-level Euler Monte Carlo algorithm, up to a constant. Indeed, the number of arithmetic operations and calls of the random number generator as well as the number of evaluations of the drift coefficient a and diffusion coefficient b that are needed to compute $1/n_j \sum_{\ell=1}^{n_j} f(U_\ell^{(j)})$ are bounded by the number n_j of replications times the number 2^{j-1} of time steps, up to a constant. Hence $\mathrm{cost}_c(A_k, f)$ properly reflects the computation time in practice.

Example 6. SDEs also give rise to finite-dimensional quadrature problems, where μ is the distribution of the solution X at time $t = 1$, say. Then full space sampling provides the appropriate cost model, and we get

$$\mathrm{cost}_c(A_k, f) \le n_1 + 2 \sum_{j=2}^{k} n_j \asymp \sum_{j=1}^{k} n_j$$

for the multi-level algorithm according to Example 3. In this way, however, we would ignore the impact of the step size on the computational cost of the Euler scheme. Hence an analysis according to Remark 2 is necessary, and then we once more get the right-hand side of (13) as an upper bound for the cost.

4 Analysis of the Multi-level Algorithm

In the sequel we consider a sequence of mappings $\varphi^{(k)}$ with associated bias and variance functions b_k and v_j, respectively, see Section 2. Furthermore, we consider the corresponding variable subspace model with cost function c, see (10) and (11).

4.1 General Results

Suppose that there exist real numbers $M > 1$ and $\gamma, \rho, \tau > 0$ such that

$$|b_k(f)| \le \gamma\, M^{-k\,\rho}, \quad (v_j(f))^{1/2} \le \gamma\, M^{-j\,\tau}, \quad \dim(\mathfrak{X}_j) \le \gamma\, M^j. \tag{15}$$

We use A_k to denote the multi-level approximation given by (4) with the numbers of replications defined by

$$n_j = \begin{cases} \lceil M^{k\,2\rho - j\,(1/2+\tau)} \rceil, & \text{if } \tau \geq 1/2, \\ \lceil M^{k\,(2\rho+1/2-\tau) - j\,(1/2+\tau)} \rceil, & \text{if } \tau < 1/2 \end{cases}$$

for $j = 1, \ldots, k$. By $a_+ = \max(a, 0)$ we denote the positive part of $a \in \mathbb{R}$.

The following result is due to Giles, see [14] for the case $\rho \geq 1/2$.

Theorem 1. *Assume that* (15) *holds and put*

$$\widetilde{\rho} = \min(\rho, 1/2)$$

as well as

$$\Gamma_k = \mathrm{cost}_c(A_k, f).$$

Then there exists a constant $\widetilde{\gamma} > 0$, which may depend on M, γ, ρ, τ, such that the multi-level approximation A_k satisfies

$$\big(\mathrm{E}(S(f) - A_k(f))^2\big)^{1/2} \leq \widetilde{\gamma} \begin{cases} \Gamma_k^{-\widetilde{\rho}}, & \text{if } \tau > 1/2, \\ \Gamma_k^{-\widetilde{\rho}} \log(\Gamma_k) & \text{if } \tau = 1/2 \leq \rho, \\ \Gamma_k^{-\widetilde{\rho}} \left(\log(\Gamma_k)\right)^{1/2} & \text{if } \tau = 1/2 > \rho, \\ \Gamma_k^{-\rho/(1+2\,(\rho-\tau)_+)} & \text{if } \tau < 1/2. \end{cases}$$

Proof. First assume that $\tau \geq 1/2$. Due to (7) and the definition of n_j,

$$\begin{aligned} \mathrm{E}(S(f) - A_k(f))^2 &\preceq \sum_{j=1}^{k} M^{-j\,2\tau}\, M^{-(k\,2\rho - j\,(1/2+\tau))} + M^{-k\,2\rho} \\ &\preceq M^{-k\,2\rho} \sum_{j=1}^{k} M^{-j\,(\tau-1/2)} + M^{-k\,2\rho} \\ &\preceq \begin{cases} M^{-k\,2\rho}, & \text{if } \tau > 1/2, \\ M^{-k\,2\rho}\, k, & \text{if } \tau = 1/2. \end{cases} \end{aligned}$$

By (12),

$$\begin{aligned} \Gamma_k &\leq 2 \sum_{j=1}^{k} n_j\, \dim(\mathfrak{X}_j) \\ &\preceq \sum_{j=1}^{k} \big(1 + M^{k\,2\rho - j\,(1/2+\tau)}\big)\, M^j \\ &\preceq M^k + M^{k\,2\rho} \sum_{j=1}^{k} M^{j\,(1/2-\tau)} \\ &\preceq M^k + \begin{cases} M^{k\,2\rho}, & \text{if } \tau > 1/2, \\ M^{k\,2\rho}\, k, & \text{if } \tau = 1/2. \end{cases} \end{aligned}$$

Furthermore,

$$\Gamma_k \geq n_1 \, \dim(\mathfrak{X}_1) \succeq M^{k\,2\rho},$$

which implies

$$\log(\Gamma_k) \succeq k.$$

Finally, use the relations

$$M^{-k\,2\rho} \asymp (M^k + M^{k\,2\rho})^{-2\widetilde{\rho}}$$

and

$$M^{-k\,2\rho}\, k \asymp \begin{cases} (M^k + M^{k\,2\rho}\, k)^{-2\widetilde{\rho}}\, k^2, & \text{if } \rho \geq 1/2, \\ (M^k + M^{k\,2\rho}\, k)^{-2\widetilde{\rho}}\, k, & \text{if } \rho < 1/2 \end{cases}$$

to finish the proof for $\tau \geq 1/2$.

Next, consider the case $\tau < 1/2$. Then

$$\begin{aligned} \mathrm{E}(S(f) - A_k(f))^2 &\preceq \sum_{j=1}^{k} M^{-j\,2\tau}\, M^{-(k\,(2\rho+(1/2-\tau))-j\,(1/2+\tau))} + M^{-k\,2\rho} \\ &\asymp M^{-k\,2\rho} \end{aligned}$$

and

$$\begin{aligned} \Gamma_k &\preceq \sum_{j=1}^{k} \bigl(1 + M^{k\,(2\rho+(1/2-\tau))-j\,(1/2+\tau)}\bigr)\, M^j \\ &\asymp M^k + M^{k\,(2\rho+1-2\tau)} \\ &\asymp M^{k\,(1+2\,(\rho-\tau)_+)}, \end{aligned}$$

which completes the proof.

Remark 3. For finite-dimensional quadrature a variant of Theorem 1 is applicable, if the underlying definition of the computational cost is chosen according to Remark 2. Instead of the bound on $\dim(\mathfrak{X}_j)$ in (15) one has to assume that the cost for simulation of $(f(X^{(j)}), f(X^{(j-1)}))$ is bounded by $\gamma\, M^j$. Actually, this variant is close to the analysis of the multi-level algorithm in [14, Theorem 3.1].

Next, we discuss the performance of the classical Monte Carlo approximation under the assumption (15). Clearly,

$$\bigl(\mathrm{Var}(f(X^{(k)}))\bigr)^{1/2} \leq \sum_{j=1}^{k} \bigl(v_j(f)\bigr)^{1/2},$$

so that (15) implies

$$|b_k(f)| \leq \gamma\, M^{-k\,\rho}, \quad \mathrm{Var}(f(X^{(k)})) \leq \gamma, \quad \dim(\mathfrak{X}_k) \leq \gamma\, M^k \tag{16}$$

with some constant $\gamma > 0$.

Assume that (16) holds. We use A'_k to denote the classical Monte Carlo approximation (2) with the number of replications defined by

$$n = \lceil M^{k\,2\rho} \rceil$$

and we put

$$\Gamma'_k = \mathrm{cost}_c(A'_k, f),$$

where c is the cost function given by (14) for the appropriate fixed subspace model with $\mathfrak{X}_0 = \mathfrak{X}_k$. Then it is straightforward to check that there exists a constant $\widetilde{\gamma} > 0$, which may depend on M, γ, ρ, such that

$$\left(\mathrm{E}(S(f) - A'_k(f))^2\right)^{1/2} \leq \widetilde{\gamma} \left(\Gamma'_k\right)^{-\rho/(1+2\rho)}. \tag{17}$$

Remark 4. We compare the multi-level algorithm with the classical Monte Carlo approximation on the basis of the upper error bounds provided by Theorem 1 and (17), respectively.

Up to logarithmic factors, the corresponding orders of convergence of these bounds in terms of powers of the cost are given by

$$\theta^*(\rho, \tau) = \begin{cases} \min(1/2, \rho), & \text{if } \tau \geq 1/2, \\ \rho/(1 + 2\,(\rho - \tau)_+), & \text{if } \tau < 1/2 \end{cases}$$

for the multi-level algorithm, and

$$\theta(\rho) = \rho/(1 + 2\rho)$$

for the classical approach. Put $\widetilde{\tau} = \min(1/2, \tau)$. We always have

$$1 < \frac{\theta^*(\rho, \tau)}{\theta(\rho)} \leq \frac{\theta^*(\widetilde{\tau}, \tau)}{\theta(\widetilde{\tau})} = 1 + 2\,\widetilde{\tau} \leq 2$$

and

$$\lim_{\rho \to 0} \frac{\theta^*(\rho, \tau)}{\theta(\rho)} = \lim_{\rho \to \infty} \frac{\theta^*(\rho, \tau)}{\theta(\rho)} = 1.$$

4.2 Lipschitz Continuous Integrands

Now we turn to the particular case of Lipschitz continuous integrands, as we assume that

$$|f(x) - f(y)| \leq \|x - y\|_{\mathfrak{X}}, \qquad x, y \in \mathfrak{X}. \tag{18}$$

Moreover, we put $X^{(0)} = 0$ and

$$\delta_j = \left(\mathrm{E}\,\|X - X^{(j)}\|_{\mathfrak{X}}^2\right)^{1/2}$$

for $j = 0, \dots, k$.

We immediately get

$$b_k(f) \le \delta_k$$

as well as

$$v_j(f) \le \mathrm{E}\,\|X^{(j)} - X^{(j-1)}\|_{\mathfrak{X}}^2 \le (\delta_j + \delta_{j-1})^2$$

for $j = 1, \dots, k$. The analysis for Lipschitz continuous integrands therefore corresponds to the diagonal case $\rho = \tau$ in Theorem 1.

We select mappings $\varphi^{(j)}$ such that

$$\dim \operatorname{span}((\varphi^{(j)}(\widetilde{\mathfrak{X}}))) \le 2^j, \tag{19}$$

and for any integer $N \ge 16$ we define the parameters of the multi-level algorithm A_N by

$$k = \lfloor \log_2(N/8) \rfloor \tag{20}$$

and

$$n_j = \lceil 2^{k-j}/(3k) \rceil \tag{21}$$

for $j = 1, \dots, k$. See [5, Lemma 3] for the following result.

Theorem 2. *Under the assumptions* (18)–(21) *the multi-level algorithm* A_N *satisfies*

$$\mathrm{cost}_c(A_N, f) \le N$$

and

$$\big(\mathrm{E}(S(f) - A_N(f))^2\big)^{1/2} \le 12\sqrt{2}\left(\frac{\log_2 N}{N} \sum_{j=0}^{k} 2^j\, \delta_j^2\right)^{1/2}.$$

Proof. Note that (19) implies $\dim(\mathfrak{X}_j) \le 2^{j+1} - 2$ for $\mathfrak{X}_j$ according to (11), and therefore

$$\mathrm{cost}_c(A_N, f) \le 2\, n_1 + \sum_{j=2}^{k} 3\, 2^j\, n_j \le 2^{k+3} \le N$$

follows from (12), (20), and (21). Moreover

$$\begin{aligned}
\mathrm{E}(S(f) - A_N(f))^2 &\le \sum_{j=1}^{k} \frac{(\delta_j + \delta_{j-1})^2}{n_j} + \delta_k^2 \\
&\le \frac{2}{n_1}\,\delta_0^2 + \sum_{j=1}^{k-1} 2\left(\frac{1}{n_j} + \frac{1}{n_{j+1}}\right)\delta_j^2 + 2\left(\frac{1}{n_k} + \frac{1}{2}\right)\delta_k^2 \\
&\le \frac{18k}{2^k} \sum_{j=0}^{k} 2^j\, \delta_j^2,
\end{aligned}$$

which completes the proof.

Example 7. In the situation of Example 3 the estimate (19) is satisfied. Moreover, it is well known that (under standard smoothness conditions for the drift and diffusion coefficient of the SDE)

$$\delta_j \le c_p \, 2^{-j/2}$$

for all spaces $\mathfrak{X} = L_p([0,1])$ with $1 \le p < \infty$ and

$$\delta_j \le c_\infty \, j \, 2^{-j/2}$$

for $\mathfrak{X} = C([0,1])$. Hence

$$\big(\mathrm{E}(S(f) - A_N(f))^2\big)^{1/2} \le 12\sqrt{2}\, c_p \, \frac{\log_2 N}{N^{1/2}}$$

for $\mathfrak{X} = L_p([0,1])$ and

$$\big(\mathrm{E}(S(f) - A_N(f))^2\big)^{1/2} \le 12\sqrt{2}\, c_\infty \, \frac{(\log_2 N)^2}{N^{1/2}}$$

for $\mathfrak{X} = C([0,1])$. Analogous results are valid for systems of SDEs. See, e.g., [29, 30] for results and references.

Note that Asian as well as look-back options lead to Lipschitz-continuous integrands. We refer to [14, 16] for a corresponding analysis and numerical experiments using multi-level Euler Monte Carlo algorithms, while a multi-level Milstein Monte Carlo algorithm is employed in [15].

Recall that δ_j is based on the choice of the mapping $\varphi^{(j)} : \widetilde{\mathfrak{X}} \to \mathfrak{X}$. Minimizing δ_j subject to a constraint

$$\dim(\mathrm{span}(\varphi^{(j)}(\widetilde{\mathfrak{X}}))) \le \kappa$$

leads to the notion of average Kolmogorov widths of order two, which are defined by

$$d_\kappa^{(r)} = \inf_{\dim(\mathfrak{X}_0) \le \kappa} \Big(\mathrm{E} \inf_{x_0 \in \mathfrak{X}_0} \|X - x_0\|_{\mathfrak{X}}^r \Big)^{1/r} \tag{22}$$

with $r = 2$. Here the infimum with respect to $x_0 \in \mathfrak{X}_0$ corresponds to the best approximation of any realization of X by elements from the subspace $\mathfrak{X}_0$, and $\varphi^{(j)}$ is a metric projection of $\widetilde{\mathfrak{X}}$ onto $\mathfrak{X}_0$. The quality of this subspace is measured by an average distance of X to $\mathfrak{X}_0$, and minimization over all subspaces with dimension at most κ leads to the average Kolmogorov width $d_\kappa^{(r)}$. We add that $\lim_{\kappa\to\infty} d_\kappa^{(r)} = 0$, if $\mathfrak{X}$ is separable and $\mathrm{E}\,\|X\|_{\mathfrak{X}}^r < \infty$. Average Kolmogorov widths and their relation to further scales of approximation quantities for random elements were studied in [4, 27, 28], see also [34].

We now suppose that the sequence of Kolmogorov widths $d_\kappa^{(2)}$ is regularly varying of index $-\rho \in \,]-\infty, 0[$, i.e.,

$$d_\kappa^{(2)} = \kappa^{-\rho}\, L(\kappa) \tag{23}$$

with a slowly varying function $L : [1,\infty[\to]0,\infty[$. This means that $\lim_{x\to\infty} L(r\,x)/L(x) = 1$ for every $r > 0$. By definition L is almost increasing if

$$\inf_{x_0 \le x < y} L(y)/L(x) > 0$$

for some $x_0 > 0$, see [3].

Theorem 3. *Assume that* (23) *holds. If*

(i) $\rho \neq 1/2$ *or*
(ii) $\rho = 1/2$ *and* L *is bounded or almost increasing,*

then there exists a constant $\gamma > 0$ *and a sequence of multi-level algorithms* A_N *such that*

$$\mathrm{cost}_c(A_N, f) \le N$$

and

$$\big(\mathrm{E}(S(f) - A_N(f))^2\big)^{1/2} \le \gamma \begin{cases} N^{-1/2}\,(\log_2 N)^{1/2} & \text{if } \rho > 1/2, \\ \max(N^{-1/2}, d_N^{(2)})\,\log_2 N & \text{if } \rho = 1/2, \\ d_N^{(2)}\,\log_2 N & \text{if } \rho < 1/2. \end{cases}$$

Proof. Consider the multi-level algorithm A_N from Theorem 2, where the mappings $\varphi^{(j)}$ are chosen such that

$$\delta_j \le 2\, d_{2^j}^{(2)}.$$

By assumption,

$$d_{2^j}^{(2)} = 2^{-j\,\rho}\, L(2^j).$$

Since $\mathrm{cost}_c(A_N, f) \le N$ and

$$\mathrm{E}(S(f) - A_N(f))^2 \le 576\, \frac{\log_2 N}{N} \sum_{j=0}^{k} 2^{j(1-2\rho)}\, (L(2^j))^2,$$

it remains to show that

$$\sum_{j=0}^{k} 2^{j(1-2\rho)}\, (L(2^j))^2 \preceq \big(\max\big(1, N^{1-2\rho}\,(L(N))^2\big)\,\log_2 N\big)^{\widetilde{\rho}} \tag{24}$$

with $\widetilde{\rho} = 0$ if $\rho > 1/2$ and $\widetilde{\rho} = 1$ otherwise.

First assume that $\rho > 1/2$. Then $(1-2\rho)/2 < 0$, which implies

$$\lim_{j\to\infty} 2^{j(1-2\rho)/2}\, (L(2^j))^2 = 0,$$

since the function L^2 is slowly varying as well. Consequently,

$$\sum_{j=0}^{k} 2^{j(1-2\rho)}\,(L(2^j))^2 \preceq \sum_{j=0}^{k} 2^{j(1-2\rho)/2} \preceq 1.$$

Next assume that $0 < \rho < 1/2$. Then

$$(L(x))^2/(L(y))^2 \preceq (y/x)^{1-2\rho}$$

for $1 \le x \le y$, and therefore

$$\begin{aligned}\sum_{j=0}^{k} 2^{j(1-2\rho)}\,(L(2^j))^2 &\preceq \sum_{j=0}^{k} 2^{j(1-2\rho)}\,(N/2^j)^{1-2\rho}\,(L(N))^2 \\ &\preceq N^{1-2\rho}\,(L(N))^2\,\log_2 N.\end{aligned}$$

Finally, consider the case $\rho = 1/2$. By assumption L is bounded or almost increasing, and therefore

$$\sum_{j=0}^{k} 2^{j(1-2\rho)}\,(L(2^j))^2 = \sum_{j=0}^{k} (L(2^j))^2 \preceq \max(1, (L(N))^2)\,\log_2 N,$$

which completes the proof.

Remark 5. The error bound in Theorem 3 can be slightly improved in the case $\rho < 1/2$ if the slowly varying function L is almost increasing. Then

$$\big(\mathrm{E}(S(f) - A_N(f))^2\big)^{1/2} \le \gamma\, d_N^{(2)}\,(\log_2 N)^{1/2}$$

for some constant $\gamma > 0$.

Example 8. Consider an SDE, and let $\mathfrak{X} = C([0,1], \mathbb{R}^m)$ or $\mathfrak{X} = L_p([0,1], \mathbb{R}^m)$ with $1 \le p < \infty$. Then (under appropriate smoothness conditions on the coefficients of the SDEs)

$$d_\kappa^{(2)} \asymp \kappa^{-1/2},$$

see [5, Prop. 3]. Hence the estimate from Example 7 can be slightly improved for $\mathfrak{X} = C([0,1], \mathbb{R}^m)$ to an upper bound of order $\log_2 N/N^{1/2}$.

Remark 6. Our proof of Theorem 3 is based on inequality (24), which is equivalent to

$$\sum_{j=0}^{k} (L(2^j))^2 \preceq \max\big(1, (L(2^k))^2\big)\,k \tag{25}$$

in the case $\rho = 1/2$. Note that the latter inequality does not hold without an additional assumption on the slowly varying function L. For example, consider the function

$$L(x) = \exp\big((\log_2 x)^{1/3}\,\cos((\log_2 x)^{1/3})\big)$$

with $x \ge 1$. Then L is slowly varying and we have

$$\limsup_{k\to\infty} \big(\max\big(1, (L(2^k))^2\big)\,k\big)^{-1} \sum_{j=0}^{k} (L(2^j))^2 = \infty.$$

5 Minimal Errors in Different Cost Models

In order to determine the power of variable subspace sampling, and in particular the power of multi-level algorithms, we consider the worst case errors and cost of randomized algorithms A on a class F of integrands $f : \mathfrak{X} \to \mathbb{R}$. These quantities are defined by

$$e(A) = \sup_{f \in F} \big(\mathrm{E}(S(f) - A(f))^2\big)^{1/2}$$

and

$$\mathrm{cost}_c(A) = \sup_{f \in F} \mathrm{E}\, \mathrm{cost}_c(A, f),$$

if the cost per evaluation of $f \in F$ is modelled by $c : \mathfrak{X} \to \mathbb{N} \cup \{\infty\}$.

Actually we have already used the worst case point of view in the previous section. For instance, with $F = \mathrm{Lip}(1)$ denoting the class of all functionals f that satisfy (18), the error bound from Theorem 2 is equivalent to

$$e(A_N) \le 12\sqrt{2} \left(\frac{\log_2 N}{N}\right)^{1/2} \left(\sum_{j=0}^{k} 2^j\, \delta_j^2\right)^{1/2},$$

and obviously

$$\mathrm{cost}_c(A_N) \le N.$$

We extend our analysis beyond the class of multi-level algorithms, as we consider the class $\mathcal{A}^{\mathrm{ran}}$ of all randomized algorithms. See, e.g., [5] for the formal definition. Here we only mention that $\mathcal{A}^{\mathrm{ran}}$ contains in particular all random variables of the form

$$A(f) = \phi(f(X_1), \ldots, f(X_n))$$

with any choice of a joint distribution of $(X_1, \ldots, X_n)$ on $\mathfrak{X}^n$ and any measurable mapping $\phi : \mathbb{R}^n \to \mathbb{R}$. Note that randomized quadrature formulas are a particular instance thereof, see (8).

For comparing the power of different sampling regimes it does not suffice to establish upper bounds for the error and cost of specific algorithms. Instead, one has to study minimal errors and to establish lower bounds.

Let C_{fix} denote the set of all cost functions given by (14) with any finite-dimensional subspace $\{0\} \subsetneq \mathfrak{X}_0 \subset \mathfrak{X}$, let C_{var} denote the set of all cost functions given by (10) with any increasing sequence of finite-dimensional subspaces $\{0\} \subsetneq \mathfrak{X}_i \subset \mathfrak{X}$, and let C_{full} consist of the constant cost function one, see (9). For

$$\mathrm{samp} \in \{\mathrm{fix}, \mathrm{var}, \mathrm{full}\}$$

and $N \in \mathbb{N}$ we introduce the N-th minimal error

$$e^{\mathrm{ran}}_{N,\mathrm{samp}} = \inf\{e(A) : A \in \mathcal{A}^{\mathrm{ran}},\ \exists\, c \in C_{\mathrm{samp}} : \mathrm{cost}_c(A) \le N\}.$$

According to this definition a most favorable cost model $c \in C_{\mathrm{samp}}$ is used for assessing the quality of an algorithm $A \in \mathcal{A}^{\mathrm{ran}}$. We add that minimal errors are key quantities in information-based complexity, see, e.g., [30, 31, 34, 36].

Clearly

$$e^{\mathrm{ran}}_{N,\mathrm{full}} \le e^{\mathrm{ran}}_{N,\mathrm{var}} \le e^{\mathrm{ran}}_{N,\mathrm{fix}},$$

and these quantities allow us to compare the different sampling regimes. For instance, variable subspace sampling is superior to fixed subspace sampling for a class of integrands F and a measure μ iff the minimal errors $e^{\mathrm{ran}}_{N,\mathrm{var}}$ are significantly smaller than the minimal errors $e^{\mathrm{ran}}_{N,\mathrm{fix}}$. Note that a lower bound for $e^{\mathrm{ran}}_{N,\mathrm{fix}}$ and an upper bound for $e^{\mathrm{ran}}_{N,\mathrm{var}}$ are needed to establish this conclusion. Conversely, a lower bound for $e^{\mathrm{ran}}_{N,\mathrm{var}}$ and an upper bound for $e^{\mathrm{ran}}_{N,\mathrm{fix}}$ are needed to prove that variable subspace sampling is not superior to fixed subspace sampling.

6 Optimal Quadrature of Lipschitz Functionals

Throughout this section we assume that

$$F = \mathrm{Lip}(1).$$

In this case the minimal errors for the quadrature problem can be estimated from above and below in terms of average Kolmogorov widths, see (22), and quantization numbers. A partial result was already formulated in Theorem 3.

The quantization numbers of order $r \ge 1$ are defined by

$$q^{(r)}_n = \inf_{|\mathfrak{X}_0| \le n} \Big(\mathrm{E} \min_{x_0 \in \mathfrak{X}_0} \|X - x_0\|^r_{\mathfrak{X}} \Big)^{1/r}.$$

Both, average Kolmogorov widths and quantization numbers correspond to best approximation from optimally chosen subsets $\mathfrak{X}_0 \subseteq \mathfrak{X}$, subject to a constraint on the dimension of the linear subspace $\mathfrak{X}_0$ or the size of the finite set $\mathfrak{X}_0$. Quantization of random elements X that take values in finite-dimensional spaces $\mathfrak{X}$ has been studied since the late 1940's, and we refer to the monograph [17] for an up-to-date account. For random elements X taking values in infinite-dimensional spaces $\mathfrak{X}$, quantization has been studied since about ten years. Results are known for Gaussian processes, see, e.g., [7, 10, 12, 24, 25], and for diffusion processes, see [5, 8, 9, 26].

In the sequel we assume that $q^{(1)}_n < \infty$. Clearly, $\lim_{n\to\infty} q^{(r)}_n = 0$ if $\mathfrak{X}$ is separable and $\mathrm{E}\,\|X\|^r_{\mathfrak{X}} < \infty$.

In the following theorem the upper bounds on $e^{\mathrm{ran}}_{N,\mathrm{full}}$ and $e^{\mathrm{ran}}_{N,\mathrm{fix}}$ as well as all lower bounds are due to [5]. See Theorem 3 for the upper bound on $e^{\mathrm{ran}}_{N,\mathrm{var}}$.

Theorem 4. *For full space sampling*

$$N^{1/2} \sup_{n \ge 4N} (q^{(1)}_{n-1} - q^{(1)}_n) \preceq e^{\mathrm{ran}}_{N,\mathrm{full}} \preceq N^{-1/2}\, q^{(2)}_N.$$

For variable subspace sampling

$$\max(e^{\mathrm{ran}}_{N,\mathrm{full}}, d^{(1)}_{2N}) \preceq e^{\mathrm{ran}}_{N,\mathrm{var}},$$

and, under the assumption of Theorem 3,

$$e^{\mathrm{ran}}_{N,\mathrm{var}} \preceq \max(N^{-1/2}, d^{(2)}_N)\, \log_2 N.$$

For fixed subspace sampling

$$\inf_{\kappa\, n\le N} \max(e^{\mathrm{ran}}_{n,\mathrm{full}}, d^{(1)}_\kappa) \preceq e^{\mathrm{ran}}_{N,\mathrm{fix}} \preceq \inf_{\kappa\, n\le N} (n^{-1/2} + d^{(2)}_\kappa).$$

Remark 7. Clearly $d^{(r)}_\kappa$ and $q^{(r)}_n$ only depend on the distribution μ of X, and they can equivalently be defined in terms of the Wasserstein distance on the space of Borel probability measures on $\mathfrak{X}$. See, e.g., [5] for these facts and for further references. Thus Theorem 4 relates quadrature of Lipschitz functionals by means of randomized algorithms to approximation of μ by distributions with finite support and distributions concentrated on finite-dimensional subspaces, and the latter constraints reflect the restrictions on evaluation of the functionals in the three sampling regimes.

We add that an analogue analysis can be carried out for quadrature of Lipschitz functionals by means of deterministic algorithms only. In the setting of full space sampling it is well known that this quadrature problem is equivalent to the quantization problem in the sense that the corresponding minimal errors satisfy

$$e^{\mathrm{det}}_{N,\mathrm{full}} = q^{(1)}_N. \tag{26}$$

See [5] for details and for further references.

Remark 8. The following algorithms achieve the upper bounds in Theorem 4. For full space sampling we may use quantization for variance reduction, see [5, Thm. 2] for details. For variable subspace sampling we may use the multilevel algorithm according to Theorem 3. For fixed subspace sampling we may choose mappings φ_k such that

$$\dim(\mathrm{span}(\varphi^{(k)}(\widetilde{\mathfrak{X}}))) \le k$$

and

$$\bigl(\mathrm{E}\,\|X - X^{(k)}\|^2_{\mathfrak{X}}\bigr)^{1/2} \le 2\, d^{(2)}_k$$

and employ the classical Monte Carlo algorithm (2), see [5, Thm. 4].

6.1 Gaussian Measures

In this section we study the case of a zero mean Gaussian measure μ on a separable Banach space $\mathfrak{X}$. In order to apply Theorem 4 we have to know the

asymptotic behaviour of the average Kolmogorov widths and the quantization numbers. To this end we consider the small ball function

$$\psi(\varepsilon) = -\ln \mu(\{x \in \mathfrak{X} : \|x\| \le \varepsilon\}), \quad \varepsilon > 0,$$

of μ, and we assume that there exist constants $\alpha > 0$ and $\beta \in \mathbb{R}$ such that

$$\psi(\varepsilon) \asymp \varepsilon^{-\alpha} \, (\ln \varepsilon^{-1})^{\beta} \tag{27}$$

as ε tends to zero. This implies

$$q_n^{(r)} \asymp (\ln n)^{-1/\alpha} \, (\ln \ln n)^{\beta/\alpha},$$

and

$$d_\kappa^{(r)} \asymp \kappa^{-1/\alpha} \, (\ln \kappa)^{\beta/\alpha},$$

see [7, Thm. 3.1.2] and [4, Cor. 4.7.2], respectively.

Typically, (27) holds for infinite-dimensional spaces $\mathfrak{X}$, see, e.g., [23] for results and further references. For example, if μ is the distribution of a d-dimensional Brownian sheet on $\mathfrak{X} = L_2([0,1]^d)$ then $\alpha = 2$ and $\beta = 2(d-1)$, see [6, 13].

Essentially the following results are a consequence of Theorems 2 and 4, see [5, Sec. 8].

Theorem 5. *For variable subspace sampling the minimal errors are bounded as follows.*
If $\alpha > 2$, then

$$N^{-1/\alpha} \, (\ln N)^{\beta/\alpha} \preceq e^{\mathrm{ran}}_{N,\mathrm{var}} \preceq N^{-1/\alpha} \, (\ln N)^{\beta/\alpha + 1/2}.$$

If $\alpha = 2$ and $\beta \neq -1$, then

$$N^{-1/2} \, (\ln N)^{\beta/2} \preceq e^{\mathrm{ran}}_{N,\mathrm{var}} \preceq N^{-1/2} \, (\ln N)^{(\beta/2+1/2)_+ + 1/2}.$$

If $\alpha = 2$ and $\beta = -1$, then

$$N^{-1/2} \, (\ln N)^{-1/2} \preceq e^{\mathrm{ran}}_{N,\mathrm{var}} \preceq N^{-1/2} \, (\ln N)^{1/2} \, (\ln \ln N)^{1/2}.$$

If $0 < \alpha < 2$, then

$$e^{\mathrm{ran}}_{N,\mathrm{var}} \preceq N^{-1/2} \, (\ln N)^{1/2}$$

and

$$\limsup_{N \to \infty} e^{\mathrm{ran}}_{N,\mathrm{var}} \, N^{1/2} \, (\ln N)^{1+1/\alpha} \, (\ln \ln N)^{-\beta/\alpha} > 0.$$

Theorem 5 provides sharp upper and lower bounds on the minimal errors for variable subspace sampling, up to logarithmic factors and up to the fact that one of the lower bounds is established only for an infinite sequence of integers N. The order of the polynomial term $N^{-\gamma_{\mathrm{var}}}$ is

$$\gamma_{\mathrm{var}} = \min(1/2, 1/\alpha).$$

We add that the upper bounds hold for suitable multi-level algorithms, which thus turn out to be almost optimal for variable subspace sampling, see [5].

Theorem 6. *For full space sampling the minimal errors satisfy*

$$e_{N,\mathrm{full}}^{\mathrm{ran}} \preceq N^{-1/2} \, (\ln N)^{-1/\alpha} \, (\ln \ln N)^{\beta/\alpha}$$

and

$$\limsup_{N\to\infty} e_{N,\mathrm{full}}^{\mathrm{ran}} \, N^{1/2} \, (\ln N)^{1+1/\alpha} \, (\ln \ln N)^{-\beta/\alpha} > 0.$$

Roughly speaking, Theorem 6 determines the asymptotic behaviour of the minimal errors for full space sampling, and the order of the polynomial term $N^{-\gamma_{\mathrm{full}}}$ is

$$\gamma_{\mathrm{full}} = 1/2.$$

We conclude that variable subspace sampling is as powerful as full subspace sampling iff $\alpha \leq 2$ and, consequently, suitable multi-level algorithms are almost optimal even in a much stronger sense in this case. As a specific example we mention any fractional Brownian motion with Hurst parameter $H \in \,]0,1[$ either on $\mathfrak{X} = C([0,1])$ or on $\mathfrak{X} = L_p([0,1])$ with $1 \leq p < \infty$. In all cases we have $\alpha = 1/H$ and therefore $\gamma_{\mathrm{full}} = \gamma_{\mathrm{var}}$ iff $H \geq 1/2$.

Theorem 7. *For fixed subspace sampling the minimal errors satisfy*

$$e_{N,\mathrm{fix}}^{\mathrm{ran}} \preceq N^{-1/(2+\alpha)} \, (\ln N)^{\beta/(2+\alpha)}$$

and

$$\limsup_{N\to\infty} e_{N,\mathrm{fix}}^{\mathrm{ran}} \, N^{1/(2+\alpha)} \, (\ln N)^{(2+2\alpha-\alpha\beta)/(\alpha(2+\alpha))} \, (\ln \ln N)^{-2\beta/(\alpha(2+\alpha))} > 0.$$

Ignoring again logarithmic factors as well as the shortcoming of the lower bound result, Theorem 7 states that the minimal errors for fixed subspace sampling behave like $N^{-\gamma_{\mathrm{fix}}}$ with order

$$\gamma_{\mathrm{fix}} = 1/(2+\alpha).$$

Clearly, $\gamma_{\mathrm{fix}} < \gamma_{\mathrm{var}}$ for all $\alpha > 0$ so that variable subspace sampling is always superior to fixed subspace sampling, and this superiority is maximal for $\alpha = 2$ when $\gamma_{\mathrm{var}} = 1/2 = 2\,\gamma_{\mathrm{fix}}$. The dependence of the orders γ_{var}, γ_{full}, and γ_{fix} on the parameter α of the small ball function (27) is illustrated in Figure 6.1, which summarizes the essential content of Theorems 5 to 7.

6.2 Diffusion Processes

In this section we consider the distribution μ of an m-dimensional diffusion process X on the space $\mathfrak{X} = C = C([0,1], \mathbb{R}^m)$ or on a space $\mathfrak{X} = L_p = L_p([0,1], \mathbb{R}^m)$ with $1 \leq p < \infty$. More precisely, X is given by

$$\begin{aligned} dX_t &= a(X_t)\, dt + b(X_t)\, dW_t, \\ X_0 &= u_0 \in \mathbb{R}^m \end{aligned} \tag{28}$$

for $t \in [0,1]$ with an m-dimensional Brownian motion W, and we assume that the following conditions are satisfied:

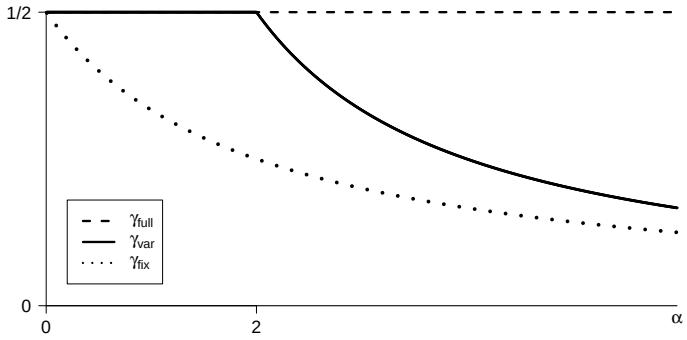


Fig. 1. Dependence of γ_{var}, γ_{full}, γ_{fix} on α

(i) $a : \mathbb{R}^m \to \mathbb{R}^m$ is Lipschitz continuous
(ii) $b : \mathbb{R}^m \to \mathbb{R}^{m\times m}$ has bounded first and second order partial derivatives and is of class C^∞ in some neighborhood of u_0
(iii) $\det b(u_0) \neq 0$

We first present bounds for the quantization numbers and the average Kolmogorov widths. Let $\mathfrak{X} = C$ or $\mathfrak{X} = L_p$. The quantization numbers $q_n^{(r)}$ satisfy

$$q_n^{(r)} \asymp (\ln n)^{-1/2}$$

for every $r > 0$. The average Kolmogorov widths $d_k^{(r)}$ satisfy

$$d_k^{(r)} \asymp k^{-1/2}$$

for every $r > 0$. See [5, Prop. 3].

The estimates from Theorems 5–7 with $\alpha = 2$ and $\beta = 0$ are valid, too, in the diffusion case, see [5, Sec. 9].

Theorem 8. *Let $\mathfrak{X} = C$ or $\mathfrak{X} = L_p$. For full space sampling the minimal errors satisfy*

$$e^{\mathrm{ran}}_{N,\mathrm{full}} \preceq N^{-1/2} (\ln N)^{-1/2}$$

and

$$\limsup_{N\to\infty} e^{\mathrm{ran}}_{N,\mathrm{full}} N^{1/2} (\ln N)^{3/2} > 0.$$

For fixed subspace sampling the minimal errors satisfy

$$e^{\mathrm{ran}}_{N,\mathrm{fix}} \preceq N^{-1/4}$$

and

$$\limsup_{N\to\infty} e^{\mathrm{ran}}_{N,\mathrm{fix}}\, N^{1/4}\,(\ln N)^{3/4} > 0.$$

For variable subspace sampling the minimal errors satisfy

$$N^{-1/2} \preceq e^{\mathrm{ran}}_{N,\mathrm{var}} \preceq N^{-1/2}\,\ln N.$$

For full space and fixed subspace sampling the lower bounds from Theorem 8 can be improved in the case $\mathfrak{X} = C$, see [5, Thm. 12].

Theorem 9. *Let $\mathfrak{X} = C$. For full space sampling the minimal errors satisfy*

$$e^{\mathrm{ran}}_{N,\mathrm{full}} \succeq N^{-1/2}\,(\ln N)^{-3/2}.$$

For fixed subspace sampling the minimal errors satisfy

$$e^{\mathrm{ran}}_{N,\mathrm{fix}} \succeq N^{-1/4}\,(\ln N)^{-3/4}.$$

Remark 9. For a Gaussian measure μ on an infinite-dimensional space, as studied in Section 6.1, as well as for μ being the distribution of the solution of an SDE on the path space, the corresponding quantization numbers $q_N^{(1)}$ essentially behave like powers of $\ln N$, asymptotically. Observing (26) we conclude that in both cases quadrature of arbitrary Lipschitz functionals is intractable by means of deterministic algorithms.

7 Concluding Remarks

The majority of results presented in this survey is concerned with Lipschitz continuous integrands f. The multi-level approach, however, is not at all linked to any kind of smoothness assumption on f. Instead, only bias and variance estimates are needed, see Theorem 1, and there are good reasons to consider classes F of integrands that either contain non-Lipschitz functionals or are substantially smaller than Lip(1).

Motivated by applications from computational finance non-continuous integrands are considered in [1] and [16]. These authors establish new results on strong approximation of SDEs, which in turn are used in the multi-level approach. In particular the computation of the expected payoff for digital and barrier options is covered by this work.

For finite-dimensional spaces $\mathfrak{X}$ much smaller classes F of integrands than Lip(1) are studied since long. With a view towards infinite-dimensional integration as a limiting case, tractability results for d-dimensional integration are most interesting, since they provide bounds on the minimal errors with an explicit dependence on the dimension d. We refer to the recent monograph [33]. Here weighted Hilbert spaces with a reproducing kernel play an important role, and in this setting full space sampling for infinite-dimensional quadrature case has already been analyzed in [21].

As for the class F of functionals, the multi-level approach also does not rely on specific properties of the measure μ. Actually, only suitable subspaces have to be identified and the simulation of two-level couplings of corresponding distributions must be feasible. So far, most of the work on multi-level algorithms is dealing with SDEs that are driven by a Brownian motion, and results for Gaussian measures μ are available as well. Recent progress in a different direction is made in [11], which provides the construction and analysis of a multi-level algorithm for Lévy-driven SDEs.

Acknowledgments

We are grateful to the editor, the referees, as well as to Rainer Avikainen and Mike Giles for their valuable remarks. This work was supported by the Deutsche Forschungsgemeinschaft (DFG) within the Priority Program 1324.

References

1. R. Avikainen, On irregular functionals of SDEs and the Euler scheme, to appear in Finance Stoch. (2009).
2. V. Bally, D. Talay, The law of the Euler scheme for stochastic differential equations. I: Convergence rate of the distribution function, Probab. Theory Relat. Fields **104**, 43–60 (1996).
3. N. H. Bingham, C. M. Goldie, J. L. Teugels, Regular Variation, Cambridge Univ. Press, Cambridge (1987).
4. J. Creutzig, Approximation of Gaussian Random Vectors in Banach Spaces, Ph.D. Dissertation, Fakultät für Mathematik und Inf., Friedrich-Schiller Universität Jena (2002).
5. J. Creutzig, S. Dereich, T. Müller-Gronbach, K. Ritter, Infinite-dimensional quadrature and approximation of distributions, to appear in Found. Comput. Math. (2009).
6. E. Csáki, On small values of the square integral of a multiparameter Wiener process, in: Statistics and Probability, (J. Mogyorodi, I. Vincze, W. Wertz, eds.), pp. 19–26, Reidel, Dordrecht (1984).
7. S. Dereich, High Resolution Coding of Stochastic Processes and Small Ball Probabilities, Ph.D. Dissertation, Institut für Mathematik, TU Berlin (2003).
8. S. Dereich, The coding complexity of diffusion processes under supremum norm distortion, Stochastic Processes Appl. **118**, 917–937 (2008).
9. S. Dereich, The coding complexity of diffusion processes under $L^p[0,1]$-norm distortion, Stochastic Processes Appl. **118**, 938–951 (2008).
10. S. Dereich, F. Fehringer, A. Matoussi, M. Scheutzow, On the link between small ball probabilities and the quantization problem, J. Theoret. Probab. **16**, 249–265 (2003).
11. S. Dereich, F. Heidenreich, A multilevel Monte Carlo algorithm for Lévy driven stochastic differential equations, work in progress.
12. S. Dereich, M. Scheutzow, High-resolution quantization and entropy coding for fractional Brownian motion, Electron. J. Probab. **11**, 700–722 (2006).

13. J. A. Fill, F. Torcaso, Asymptotic analysis via Mellin transforms for small deviations in L^2-norm of integrated Brownian sheets, Probab. Theory Relat. Fields **130**, 259–288 (2004).
14. M. B. Giles, Multilevel Monte Carlo path simulation, Oper. Res. **56**, 607–617 (2008).
15. M. B. Giles, Improved multilevel Monte Carlo convergence using the Milstein scheme, in: Monte Carlo and Quasi-Monte Carlo Methods 2006 (A. Keller, S. Heinrich, H. Niederreiter, eds.), pp. 343–358, Springer-Verlag, Berlin (2008).
16. M. B. Giles, D. J. Higham, X. Mao, Analysing multilevel Monte Carlo for options with non-globally Lipschitz payoff, to appear in Finance Stoch. (2009).
17. S. Graf, H. Luschgy, Foundations of Quantization for Probability Distributions, Lect. Notes in Math. **1730**, Springer-Verlag, Berlin (2000).
18. S. Heinrich, Monte Carlo complexity of global solution of integral equations, J. Complexity **14**, 151–175 (1998).
19. S. Heinrich, Multilevel Monte Carlo methods, in: Large Scale Scientific Computing, Lect. Notes in Comp. Sci. **2179** (S. Margenov, J. Wasniewski, P. Yalamov, eds.), pp. 58–67, Springer-Verlag, Berlin (2001).
20. S. Heinrich, E. Sindambiwe, Monte Carlo complexity of parametric integration, J. Complexity **15**, 317–341 (1999).
21. F. J. Hickernell, X. Wang, The error bounds and tractability of quasi-Monte Carlo algorithms in infinite dimension, Math. Comp. **71**, 1641–1661 (2001).
22. A. Kebaier, Statistical Romberg extrapolation: a new variance reduction method and applications to option pricing, Ann. Appl. Prob. **15**, 2681–2705 (2005).
23. W. V. Li, Q.-M. Shao, Gaussian processes: inequalities, small ball probabilities and applications, in: Stochastic Processes: Theory and Methods, Handbook of Statist., Vol. 19, (D. N. Shanbhag, C. R. Rao, eds.), pp. 533–597, North-Holland, Amsterdam (2001).
24. H. Luschgy, G. Pagès, Functional quantization of Gaussian processes, J. Funct. Anal. **196**, 486–531 (2002).
25. H. Luschgy, G. Pagès, Sharp asymptotics of the functional quantization problem for Gaussian processes, Ann. Appl. Prob. **32**, 1574–1599 (2004).
26. H. Luschgy, G. Pagès, Functional quantization of a class of Brownian diffusions: a constructive approach, Stochastic Processes Appl. **116**, 310–336 (2006).
27. V. E. Maiorov, Widths and distribution of values of the approximation functional on the Sobolev space with measure, Constr. Approx. **12**, 443–462 (1996).
28. P. Mathé, s-numbers in information-based complexity, J. Complexity **6** 41–66 (1990).
29. T. Müller-Gronbach, Optimal uniform approximation of systems of stochastic differential equations, Ann. Appl. Prob. **12**, 664–690 (2002).
30. T. Müller-Gronbach, K. Ritter, Minimal errors for strong and weak approximation of stochastic differential equations, in: Monte Carlo and Quasi-Monte Carlo Methods 2006 (A. Keller, S. Heinrich, H. Niederreiter, eds.), pp. 53–82, Springer-Verlag, Berlin (2008).
31. E. Novak, Deterministic and Stochastic Error Bounds in Numerical Analysis, Lect. Notes in Math. **1349**, Springer-Verlag, Berlin (1988).
32. E. Novak, The real number model in numerical analysis, J. Complexity **11**, 57–73 (1995).
33. E. Novak, H. Woźniakowski, Tractability of Multivariate Problems, Vol. I: Linear Information, European Mathematical Society, Zürich (2008).

34. K. Ritter, Average-Case Analysis of Numerical Problems, Lect. Notes in Math. **1733**, Springer-Verlag, Berlin (2000).
35. D. Talay, L. Tubaro, Expansion of the global error for numerical schemes solving stochastic differential equations, Stochastic Anal. Appl. **8**, 483–509 (1990).
36. J. F. Traub, G. W. Wasilkowski, H. Woźniakowski, Information-Based Complexity, Academic Press, New York (1988).
37. G. W. Wasilkowski, H. Woźniakowski, On tractability of path integration, J. Math. Phys. **37**, 2071–2088 (1996).

Preprint Series DFG-SPP 1324

http://www.dfg-spp1324.de

Reports

[1] R. Ramlau, G. Teschke, and M. Zhariy. A Compressive Landweber Iteration for Solving Ill-Posed Inverse Problems. Preprint 1, DFG-SPP 1324, September 2008.

[2] G. Plonka. The Easy Path Wavelet Transform: A New Adaptive Wavelet Transform for Sparse Representation of Two-dimensional Data. Preprint 2, DFG-SPP 1324, September 2008.

[3] E. Novak and H. Woźniakowski. Optimal Order of Convergence and (In-) Tractability of Multivariate Approximation of Smooth Functions. Preprint 3, DFG-SPP 1324, October 2008.

[4] M. Espig, L. Grasedyck, and W. Hackbusch. Black Box Low Tensor Rank Approximation Using Fibre-Crosses. Preprint 4, DFG-SPP 1324, October 2008.

[5] T. Bonesky, S. Dahlke, P. Maass, and T. Raasch. Adaptive Wavelet Methods and Sparsity Reconstruction for Inverse Heat Conduction Problems. Preprint 5, DFG-SPP 1324, January 2009.

[6] E. Novak and H. Woźniakowski. Approximation of Infinitely Differentiable Multivariate Functions Is Intractable. Preprint 6, DFG-SPP 1324, January 2009.

[7] J. Ma and G. Plonka. A Review of Curvelets and Recent Applications. Preprint 7, DFG-SPP 1324, February 2009.

[8] L. Denis, D. A. Lorenz, and D. Trede. Greedy Solution of Ill-Posed Problems: Error Bounds and Exact Inversion. Preprint 8, DFG-SPP 1324, April 2009.

[9] U. Friedrich. A Two Parameter Generalization of Lions' Nonoverlapping Domain Decomposition Method for Linear Elliptic PDEs. Preprint 9, DFG-SPP 1324, April 2009.

[10] K. Bredies and D. A. Lorenz. Minimization of Non-smooth, Non-convex Functionals by Iterative Thresholding. Preprint 10, DFG-SPP 1324, April 2009.

[11] K. Bredies and D. A. Lorenz. Regularization with Non-convex Separable Constraints. Preprint 11, DFG-SPP 1324, April 2009.

[12] M. Döhler, S. Kunis, and D. Potts. Nonequispaced Hyperbolic Cross Fast Fourier Transform. Preprint 12, DFG-SPP 1324, April 2009.

[13] C. Bender. Dual Pricing of Multi-Exercise Options under Volume Constraints. Preprint 13, DFG-SPP 1324, April 2009.

[14] T. Müller-Gronbach and K. Ritter. Variable Subspace Sampling and Multi-level Algorithms. Preprint 14, DFG-SPP 1324, May 2009.