INFINITE-DIMENSIONAL QUADRATURE AND APPROXIMATION OF DISTRIBUTIONS

JAKOB CREUTZIG, STEFFEN DEREICH, THOMAS MÜLLER-GRONBACH, AND KLAUS RITTER

Abstract. We study numerical integration of Lipschitz functionals on a Banach space by means of deterministic and randomized (Monte Carlo) algorithms. This quadrature problem is shown to be closely related to the problem of quantization and to the average Kolmogorov widths of the underlying probability measure. In addition to the general setting we analyze in particular integration w.r.t. Gaussian measures and distributions of diffusion processes. We derive lower bounds for the worst case error of every algorithm in terms of its cost, and we present matching upper bounds, up to logarithms, and corresponding almost optimal algorithms. As auxiliary results we determine the asymptotic behaviour of quantization numbers and Kolmogorov widths for diffusion processes.

1. Introduction

Let μ be a Borel probability measure on a Banach space $(\mathfrak{X}, \|\cdot\|)$ such that

$$\int_{\mathfrak{X}} \|x\| \, \mu(dx) < \infty.$$

Moreover, let F denote the class of all Lipschitz continuous functionals $f : \mathfrak{X} \to \mathbb{R}$ with Lipschitz constant at most one, i.e.,

$$|f(x) - f(y)| \le \|x - y\|, \qquad x, y \in \mathfrak{X}.$$

We wish to compute

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

for $f \in F$ by means of deterministic or 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 $\mathfrak{X}$, and in particular we study Gaussian measures μ and distributions μ of diffusion processes. Infinite-dimensional quadrature is applied, e.g., in mathematical finance and quantum physics, and moreover it is used as a computational tool to solve parabolic or elliptic partial differential equations.

We study the information cost of algorithms, which only takes into account the evaluations of the integrand f, and we consider three different cost models. In addition to full space sampling, which permits evaluation of f at any point $x \in \mathfrak{X}$ at cost one and which is commonly studied in the literature, we introduce two variants of subspace sampling. Hereby we restrict the set

Date: February 2008.

1991 *Mathematics Subject Classification.* 60G15, 60H10, 65C30.

Key words and phrases. Quadrature problem, randomized algorithm, minimal error, Wasserstein distance, functional quantization, average Kolmogorov width, multi-level Monte Carlo algorithm.

of admissible sampling points and we increase the cost per evaluation in order to obtain more realistic results for infinite-dimensional quadrature problems.

For fixed subspace sampling, f may only be evaluated at the points from a finite-dimensional subspace $\mathfrak{X}_0 \subset \mathfrak{X}$. For variable subspace sampling we have an increasing sequence of finite-dimensional subspaces $\mathfrak{X}_i \subset \mathfrak{X}$, and evaluation is possible at any point $x \in \bigcup_{i=1}^{\infty} \mathfrak{X}_i$. The cost per evaluation is defined by $\dim(\mathfrak{X}_0)$ in the fixed subspace case and $\inf\{\dim(\mathfrak{X}_i) : x \in \mathfrak{X}_i\}$ in the variable subspace case. The subspace or sequence of subspaces, resp., may be chosen arbitrarily, but it is fixed for a specific algorithm.

To give an example we consider the approximation of $S(f)$ in the diffusion case by means of a weak Itô Taylor scheme together with piecewise linear interpolation. These schemes are most commonly used with a uniform step-size $1/k$, which corresponds to fixed subspace sampling. The dimension of the subspace $\mathfrak{X}_0$, which consists of piecewise linear functions, is proportional to k, and the overall information cost of the algorithm is proportional to the product of the number of replications of the scheme and the number of time steps. The latter parameter would be completely disregarded in the full space sampling model. Heinrich (1998) and Heinrich and Sindambiwe (1999) have introduced multilevel Monte Carlo methods for computation of global solutions of integral equations and for parametric integration, respectively. A multilevel algorithm uses dependent samples in a hierarchy $\mathfrak{X}_1 \subset \ldots \subset \mathfrak{X}_m$ of finite-dimensional subspaces with only a small proportion taken from high-dimensional spaces. In the context of quadrature problems for diffusion processes multilevel Itô Taylor schemes have been introduced by Giles (2006), and here the subspaces consist of piecewise linear functions on a nested sequence of time grids. The analysis in a fixed subspace model corresponding to $\mathfrak{X}_m$ would be inadequate, since it does not capture the fact that a large proportion of samples is taken in low-dimensional spaces.

In this paper we present a worst case analysis of the quadrature problem on the Lipschitz class F, and we optimally relate the error and the cost of algorithms. For each of the three cost models we let $e^{\mathrm{det}}_{N,\mathrm{samp}}$ and $e^{\mathrm{ran}}_{N,\mathrm{samp}}$ with $\mathrm{samp} \in \{\mathrm{fix}, \mathrm{var}, \mathrm{full}\}$ denote the smallest worst case error that can be achieved by any deterministic or randomized algorithm, resp., whose worst case cost is bounded by N. We wish to determine the asymptotic behaviour of the minimal errors $e^{\mathrm{det}}_{N,\mathrm{samp}}$ and $e^{\mathrm{ran}}_{N,\mathrm{samp}}$, and we want to find algorithms with cost close to N and error close to the corresponding minimal error. Hereby we can, for instance, answer the question whether suitable multilevel Monte Carlo algorithms are superior to arbitrary randomized algorithms that sample in a fixed subspace.

In the sequel we present our main results for Gaussian measures and distributions of diffusion processes. We write $a_N \asymp b_N$ for sequences of positive real numbers a_N and b_N if

$$0 < \inf_{N\in\mathbb{N}} a_N/b_N \le \sup_{N\in\mathbb{N}} a_N/b_N < \infty.$$

Moreover, $a_N \overset{\log}{\asymp} b_N$ means that there exists $\gamma > 0$ such that

$$\limsup_{N\to\infty} \frac{a_N}{b_N} (\ln N)^{-\gamma} < \infty \quad \text{and} \quad \limsup_{N\to\infty} \frac{a_N}{b_N} (\ln N)^{\gamma} > 0,$$

which is used to simplify the presentation at this point. Note that $\overset{\log}{\asymp}$ defines a non-symmetric relation, which, roughly speaking, means the following for sequences converging to zero: Up to powers of $\ln N$, a_N converges to zero at least as fast as b_N and subsequences a_{N_k} and b_{N_k} converge to zero at the same rate.

Suppose that μ is a zero mean Gaussian measure on a separable Banach space $\mathfrak{X}$, whose small ball function

$$\varphi(\varepsilon) = -\ln \mu(\{x \in \mathfrak{X} : \|x\| \le \varepsilon\})$$

satisfies

$$\varphi(\varepsilon) \asymp \varepsilon^{-\alpha} \cdot (\ln \varepsilon^{-1})^{\beta}$$

for some constants $\alpha > 0$ and $\beta \in \mathbb{R}$ as ε tends to zero. This asymptotic behaviour typically holds for Gaussian measures on infinite-dimensional spaces, see, e.g., the review article by Li, Shao (2001). Consider, for instance, the distribution μ of the fractional Brownian motion with Hurst parameter $H \in]0,1[$ on the space $\mathfrak{X} = C([0,1])$ or $\mathfrak{X} = L_p([0,1])$ with $p \in [1,\infty[$. Then $\alpha = 1/H$ and $\beta = 0$. A non-zero constant β appears, for example, in case of μ being the distribution of the d-dimensional Brownian sheet on the space $\mathfrak{X} = L_2([0,1]^d)$. Then $\alpha = 2$ and $\beta = 2(d-1)$, see Csáki (1984) and Fill, Torcaso (2004). Given the above small ball asymptotics, the minimal errors satisfy

$$e^{\mathrm{det}}_{N,\mathrm{full}} \asymp (\ln N)^{-1/\alpha} \cdot (\ln \ln N)^{\beta/\alpha}, \tag{1}$$

see Theorem 1 and Proposition 2, and

$$\begin{aligned} e^{\mathrm{ran}}_{N,\mathrm{full}} &\overset{\log}{\asymp} N^{-1/2}, \\ e^{\mathrm{ran}}_{N,\mathrm{var}} &\overset{\log}{\asymp} N^{-\min(1/2,1/\alpha)}, \\ e^{\mathrm{ran}}_{N,\mathrm{fix}} &\overset{\log}{\asymp} N^{-1/(2+\alpha)}, \end{aligned} \tag{2}$$

see Theorems 8–10 for more precise upper and lower bounds involving β, too, via powers of $\ln N$ and $\ln \ln N$.

Suppose that μ is the distribution of an m-dimensional diffusion process on the space $\mathfrak{X} = C([0,1],\mathbb{R}^m)$, equipped with the supremum norm, or on a space $\mathfrak{X} = L_p([0,1],\mathbb{R}^m)$ with $1 \le p < \infty$. Under mild assumptions on the drift and diffusion coefficients the estimates (1) and (2) with $\alpha = 2$ and $\beta = 0$ are valid, too, in the diffusion case, see Theorem 11.

We conclude that randomization helps substantially both in the Gaussian and the diffusion case, as it turns a logarithmic decay for minimal errors even for full space sampling into a polynomial decay already for fixed subspace sampling. Moreover, for randomized algorithms, variable subspace sampling is as powerful as full space sampling if $0 < \alpha \le 2$ and always superior to fixed subspace sampling.

Note that the asymptotic analysis of minimal errors in the sense of $\overset{\log}{\asymp}$ only provides lower bounds that hold for infinitely many integers N. However, lower bounds that hold for every N are available in the Gaussian case with $\alpha \ge 2$, where

$$\lim_{N\to\infty} \ln e^{\mathrm{ran}}_{N,\mathrm{var}} / \ln N^{-1/\alpha} = 1$$

according to Theorem 10, and in the diffusion case on the space $\mathfrak{X} = C([0,1],\mathbb{R}^m)$, where

$$\begin{aligned} \lim_{N\to\infty} \ln e^{\mathrm{ran}}_{N,\mathrm{var}} / \ln N^{-1/2} &= 1, \\ \lim_{N\to\infty} \ln e^{\mathrm{ran}}_{N,\mathrm{fix}} / \ln N^{-1/4} &= 1 \end{aligned}$$

according to Theorem 12.

The upper bounds for $e^{\mathrm{ran}}_{N,\mathrm{fix}}$ are achieved by classical Monte Carlo methods using independent and identically distributed samples. The latter are drawn according to a normal distribution on a suitable finite-dimensional subspace in the Gaussian case and by means of a weak Euler scheme

in the diffusion case. The same kinds of distributions are used as building blocks for multilevel Monte Carlo methods, which yield the upper bounds for $e^{\mathrm{ran}}_{N,\mathrm{var}}$. In view of the lower bounds, these algorithms are almost optimal in the corresponding cost model.

The results in the Gaussian and in the diffusion case are derived from general theorems that relate minimal errors for the quadrature problem to best approximation of the underlying distribution μ with respect to a Wasserstein distance, where two kinds of constraints concerning the support of the approximating measures turn out to be relevant. For the quantization problem these measures are discrete, and the n-th quantization number $q^{(1)}_n$ gives the distance of μ to the class of all probability measures on $\mathfrak{X}$ with support of size at most n. The average Kolmogorov widths $d^{(1)}_k$ is defined as the distance of μ to the class of all probability measures on $\mathfrak{X}$ that are concentrated on finite-dimensional subspaces of dimension at most k. See, e.g., Chen, Fang (2004), Creutzig (2002), Dereich (2003, 2007a, 2007b), Dereich *et al.* (2003), Dereich, Scheutzow (2005), Graf, Luschgy (2000), Luschgy, Pagès (2004, 2006), Mathé (1990), Pagès, Printems (2005), and Ritter (2000) for results and references concerning the quantization problem and Kolmogorov widths.

As a well-known fact,

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

see Theorem 1. In this paper we establish the lower bounds

$$e^{\mathrm{ran}}_{N,\mathrm{full}} \geq \tfrac{1}{8} \cdot N^{1/2} \cdot \sup_{m \geq 4N} (q^{(1)}_{m-1} - q^{(1)}_m),$$

$$e^{\mathrm{ran}}_{N,\mathrm{fix}} \geq \inf_{n \cdot k \leq N} \max(e^{\mathrm{ran}}_{n,\mathrm{full}}, d^{(1)}_k),$$

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

for the minimal errors of randomized algorithms in the different cost models, see Theorems 3, 5, and 7. Upper bounds in terms of quantization numbers and average Kolmogorov widths are available as well, see Theorems 2, 4, and 6.

This paper is organized as follows. In Sections 2 and 3 we introduce the basic concepts and definitions. Full space sampling is studied in Section 4, which contains general results, and Section 5, where we briefly discuss quadrature problems on finite-dimensional spaces. Sections 6 and 7 are devoted to the analysis of randomized algorithms in the fixed subspace and the variable subspace model. In Sections 8 and 9 we study Gaussian measures and diffusion processes, resp., and we apply the general results from Sections 4, 6, and 7. As auxiliary results we determine the asymptotic behaviour of the quantization numbers and the Kolmogorov widths in the diffusion case, see Proposition 3.

2. Algorithms, Error, and Cost

2.1. **Functional Evaluation and Cost Model.** We assume that algorithms for 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

$$c : \mathfrak{X} \to \mathbb{N} \cup \{\infty\}, \tag{3}$$

and we are interested in three particular such models, which will be presented in increasing generality.

For *fixed subspace sampling* evaluations are possible only at the points from a given finite-dimensional subspace

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

and the cost for each oracle call coincides with the dimension of $\mathfrak{X}_0$. Thus,

$$c(x) = \begin{cases} \dim(\mathfrak{X}_0), & \text{if } x \in \mathfrak{X}_0 \\ \infty, & \text{otherwise.} \end{cases} \tag{4}$$

For *variable subspace sampling* we consider a sequence of finite-dimensional subspaces

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

and the cost function is defined by

$$c(x) = \inf\{\dim(\mathfrak{X}_i) : x \in \mathfrak{X}_i\}. \tag{5}$$

Finally, for *full space sampling*, we have

$$c = 1, \tag{6}$$

i.e., evaluation of functionals $f \in F$ is possible at any point $x \in \mathfrak{X}$ at cost one.

Remark 1. If $\dim(\mathfrak{X}) < \infty$ then it is most natural to allow full space sampling, with the constant one possibly being replaced by $\dim(\mathfrak{X})$. This is no longer the case for infinite-dimensional spaces, where full space sampling is mainly used to derive lower bounds for all of the cost models.

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 1. Let us illustrate fixed and variable subspace sampling by means of variants of the weak Euler scheme, which is used here for quadrature with respect to the distribution of a diffusion process with values in $\mathfrak{X} = C([0,1], \mathbb{R})$. See also Remark 8.

With a uniform step-size $1/k$ and piecewise linear interpolation this scheme randomly generates elements x of the $k+1$-dimensional subspace $\mathfrak{X}_0$ of piecewise linear functions with breakpoints at ℓ/k. Hence we have an instance of fixed subspace sampling with an oracle that computes $f(x)$ at a cost $c(x) = k+1$.

Sometimes extrapolation techniques are used to reduce the bias of the Euler scheme. In this case we deal with variable subspace sampling, where $\mathfrak{X}_i$ consists of piecewise linear functions with breakpoints at $\ell/2^{i-1}$, say, so that $\dim(\mathfrak{X}_i) = 2^{i-1} + 1$. See, e.g., Talay, Tubaro (1990) and Bally, Talay (1996). The multilevel Euler scheme, which simultaneously reduces the variance and the bias, is based on variable subspace sampling, too, see Section 7.

2.2. **Deterministic Algorithms.** Any deterministic sequential evaluation of functionals $f \in F$ is formally defined by a point

$$x_1 \in \mathfrak{X}$$

and a sequence of mappings

$$\psi_\ell : \mathbb{R}^{\ell-1} \to \mathfrak{X}, \qquad \ell \geq 2.$$

For every $f \in F$ the evaluation starts at the point x_1, and the mappings ψ_ℓ determine the subsequent evaluation points. More precisely, after n steps the functional values

$$y_1 = f(x_1)$$

and

$$y_\ell = f(\psi_\ell(y_1, \ldots, y_{\ell-1})), \qquad \ell = 2, \ldots, n,$$

are known. A decision to stop or to further evaluate f is made after each step. This is formally described by a sequence of mappings

$$\tau_\ell : \mathbb{R}^\ell \to \{0,1\}, \qquad \ell \geq 1,$$

and the total number $n(f)$ of evaluations is given by

$$n(f) = \min\{\ell \geq 1 : \tau_\ell(y_1, \ldots, y_\ell) = 1\},$$

which is finite for every $f \in F$ by assumption. Finally, an approximation

$$\widehat{S}(f) = \phi_{n(f)}(y_1, \ldots, y_{n(f)}) \tag{7}$$

to $S(f)$ is defined by a sequence of mappings

$$\phi_\ell : \mathbb{R}^\ell \to \mathbb{R}, \qquad \ell \geq 1.$$

The tuple $\widehat{\mathcal{S}} = (x_1, (\psi_\ell), (\tau_\ell), (\phi_\ell))$ will be considered as a *deterministic algorithm*, with algorithm being understood in a broad sense, and the corresponding mapping $\widehat{S} : F \to \mathbb{R}$ given by (7) will be called the mapping induced by $\widehat{\mathcal{S}}$. The class of all deterministic algorithms is denoted by $\mathbb{S}^{\mathrm{det}}$.

In every cost model (3) the cost for applying the algorithm $\widehat{\mathcal{S}}$ to $f \in F$ is defined by

$$\mathrm{cost}_c(\widehat{\mathcal{S}}, f) = c(x_1) + \sum_{\ell=2}^{n(f)} c(\psi_\ell(y_1, \ldots, y_{\ell-1})).$$

Example 2. Note that $\mathbb{S}^{\mathrm{det}}$ contains in particular all *quadrature formulas*. Here we have

$$\widehat{S}(f) = \sum_{\ell=1}^{n} a_\ell \cdot f(x_\ell)$$

with $a_\ell \in \mathbb{R}$ and $x_\ell \in \mathfrak{X}$. Clearly $\widehat{S}$ is induced by $\widehat{\mathcal{S}} = (x_1, (\psi_\ell), (\tau_\ell), (\phi_\ell))$ with constant mappings $\psi_2 = x_2, \ldots, \psi_n = x_n$, $\tau_1 = \ldots = \tau_{n-1} = 0$, and $\tau_n = 1$, i.e., all functionals $f \in F$ are evaluated non-sequentially at the same set of n points. Furthermore, $\phi_n(y_1, \ldots, y_n) = \sum_{\ell=1}^n a_\ell \cdot y_\ell$. Any such algorithm $\widehat{\mathcal{S}}$ will be called a quadrature formula with nodes x_ℓ and weights a_ℓ. We obtain

$$\mathrm{cost}_c(\widehat{\mathcal{S}}, f) = \sum_{\ell=1}^{n} c(x_\ell).$$

For instance, in a fixed subspace model (4) with $\mathfrak{X}_0 = \mathrm{span}\{x_1, \ldots, x_n\}$ we have $\mathrm{cost}_c(\widehat{\mathcal{S}}, f) = n \cdot \dim(\mathfrak{X}_0)$.

2.3. **Randomized Algorithms.** A *randomized (or Monte Carlo) algorithm* based on sequential evaluation is formally defined by a probability space $(\Omega, \mathfrak{A}, P)$ and a mapping

$$\widehat{\mathcal{S}} : \Omega \to \mathbb{S}^{\mathrm{det}},$$

which induces a mapping

$$\widehat{S} : \Omega \times F \to \mathbb{R}.$$

We require that

(i) $\widehat{S}(\cdot, f)$ is measurable for every $f \in F$,

(ii) $\omega \mapsto \mathrm{cost}_c(\widehat{\mathcal{S}}(\omega), f)$ is measurable for every $f \in F$ and every cost function c.

We refer to Nemirovsky, Yudin (1983) and Wasilkowski (1989) for this and an equivalent definition of randomized algorithms. In the sequel the random variable from (i) is denoted by $\widehat{S}(f)$, and we use $\mathbb{S}^{\rm ran}$ to denote all randomized algorithms $\widehat{\mathcal{S}}$ with properties (i) and (ii) on any probability space. Clearly, $\mathbb{S}^{\rm det} \subsetneq \mathbb{S}^{\rm ran}$.

Example 3. Let $a_1, \ldots, a_n \in \mathbb{R}$ and $X_1, \ldots, X_n$ be random variables on a probability space $(\Omega, \mathfrak{A}, P)$ taking values in $\mathfrak{X}$. Consider a randomized algorithm $\widehat{\mathcal{S}} : \Omega \to \mathbb{S}^{\rm det}$ such that $\widehat{\mathcal{S}}(\omega)$ is a quadrature formula with nodes $X_\ell(\omega)$ and weights $a_\ell \in \mathbb{R}$ for every $\omega \in \Omega$, see Example 2. Then $\widehat{\mathcal{S}}$ induces the mapping

$$\widehat{S}(\omega, f) = \sum_{\ell=1}^{n} a_\ell \cdot f(X_\ell(\omega))$$

and we have

$$\mathrm{cost}_c(\widehat{\mathcal{S}}(\omega), f) = \sum_{\ell=1}^{n} c(X_\ell(\omega)),$$

so that (i) and (ii) are satisfied. The algorithm $\widehat{\mathcal{S}} \in \mathbb{S}^{\rm ran}$ will be called a *randomized quadrature formula* with random nodes X_ℓ and deterministic weights a_ℓ.

Taking $a_1 = \ldots = a_n = 1/n$ as well as $X_1, \ldots, X_n$ being independent and distributed according to a Borel probability measure ν on $\mathfrak{X}$, we obtain randomized quadrature formulas such that

$$\widehat{S}_{n,\nu}(f) = 1/n \cdot \sum_{\ell=1}^{n} f(X_\ell).$$

Thus $\mathbb{S}^{\rm ran}$ contains in particular the *classical Monte Carlo method*. By a slight abuse of notation, we use $\widehat{\mathcal{S}}_{n,\nu}$ to denote any such randomized quadrature formula. Clearly, $\mathrm{cost}_c(\widehat{\mathcal{S}}_{n,\nu}, f) = n$ in the full space model. However, $\mathrm{cost}_c(\widehat{\mathcal{S}}_{n,\nu}, f) = \infty$ in every fixed or variable subspace model if $\nu(\mathfrak{X}_0) = 0$ for every finite-dimensional subspace $\mathfrak{X}_0 \subset \mathfrak{X}$, and this excludes the most natural choice of $\nu = \mu$ in many cases.

The weak Euler scheme with uniform step-size $1/k$, see Example 1, leads to a method $\widehat{\mathcal{S}}_{n,\nu}$ with $\nu \neq \mu$ in all non-trivial cases. In the corresponding fixed subspace model we have $\mathrm{cost}_c(\widehat{\mathcal{S}}_{n,\nu}, f) = n \cdot (k+1)$. In other words, the cost is essentially given by the product of the number of replications and the number of time steps.

2.4. **The Worst Case Setting.** The *worst case error* of $\widehat{\mathcal{S}} \in \mathbb{S}^{\rm ran}$ is defined by

$$e(\widehat{\mathcal{S}}) = \sup_{f \in F} \Big(\mathbb{E}|S(f) - \widehat{S}(f)|^2 \Big)^{1/2},$$

which in particular for $\widehat{\mathcal{S}} \in \mathbb{S}^{\rm det}$ reads

$$e(\widehat{\mathcal{S}}) = \sup_{f \in F} |S(f) - \widehat{S}(f)|.$$

The *worst case cost* of $\widehat{\mathcal{S}} \in \mathbb{S}^{\rm ran}$ is defined by

$$\mathrm{cost}_c(\widehat{\mathcal{S}}) = \sup_{f \in F} \mathbb{E}(\mathrm{cost}_c(\widehat{\mathcal{S}}, f)),$$

which in particular for $\widehat{\mathcal{S}} \in \mathbb{S}^{\rm det}$ reads

$$\mathrm{cost}_c(\widehat{\mathcal{S}}) = \sup_{f \in F} \mathrm{cost}_c(\widehat{\mathcal{S}}, f).$$

3. Minimal Errors and Approximation of Distributions

A typical question that will be addressed in this paper is as follows: what is the minimal error $e(\widehat{\mathcal{S}})$ that can be achieved by any randomized algorithm $\widehat{\mathcal{S}}$ using any kind of fixed subspace sampling such that $\mathrm{cost}_c(\widehat{\mathcal{S}})$ is bounded by N?

3.1. **Minimal Errors.** To formally pose this question together with obvious modifications we use the following notation. Let C_{fix} denote the set of all cost functions given by (4) 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 (5) 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 (6). For

$$\mathrm{alg} \in \{\mathrm{det}, \mathrm{ran}\}, \qquad \mathrm{samp} \in \{\mathrm{fix}, \mathrm{var}, \mathrm{full}\},$$

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

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

At the beginning of this section we have therefore asked to determine $e^{\mathrm{ran}}_{N,\mathrm{fix}}$. We add that minimal errors are key quantities in information-based complexity, see, e.g., Traub, Wasilkowski, Woźniakowski (1988), Novak (1988), and Ritter (2000).

Remark 2. Note that

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

3.2. **Quantization and Average Kolmogorov Widths.** We relate minimal errors for the quadrature problem with respect to μ to best approximation of μ by means of

- probability measures with a finite support,
- probability measures supported on finite-dimensional subspaces.

In this way we determine upper and lower bounds for the minimal errors and construct almost optimal algorithms for the quadrature problem.

For both variants of the approximation problem we consider the *Wasserstein distance* $\Delta^{(r)}(\cdot,\cdot)$ of order $r > 0$ on the space of Borel probability measures on $\mathfrak{X}$, i.e.,

$$\Delta^{(r)}(\rho_1, \rho_2) = \inf_{\rho} \left(\int_{\mathfrak{X}\times\mathfrak{X}} \|x_1 - x_2\|^r \, \rho(d(x_1, x_2)) \right)^{1/r},$$

where the infimum is taken over all Borel probability measures ρ on $\mathfrak{X} \times \mathfrak{X}$ with marginals ρ_i. If $\mathfrak{X}$ is separable then $\Delta^{(r)}$ is a metric for $r \ge 1$, and

$$\Delta^{(1)}(\rho_1, \rho_2) = \sup_{f\in F} \left| \int_{\mathfrak{X}} f(x)\, \rho_1(dx) - \int_{\mathfrak{X}} f(x)\, \rho_2(dx) \right|$$

according to the Kantorovich-Rubinstein Theorem. See Rachev (1991).

For the measure μ the *n-th quantization number* $q^{(r)}_n$ of order $r > 0$ is defined as

$$q^{(r)}_n = \inf\{\Delta^{(r)}(\mu, \widetilde{\mu}) : |\operatorname{supp}(\widetilde{\mu})| \le n\},$$

and we have

$$q^{(r)}_n = \inf_{x_1,\dots,x_n\in\mathfrak{X}} q^{(r)}(x_1, \dots, x_n),$$

where

$$q^{(r)}(x_1, \dots, x_n) = \left(\int_{\mathfrak{X}} \min_{\ell=1,\dots,n} \|x - x_\ell\|^r \, \mu(dx) \right)^{1/r}.$$

See, e.g., Graf, Luschgy (2000). In this context a collection of points $x_1, \dots, x_n \in \mathfrak{X}$ is called a codebook for quantization of the probability measure μ. Note that $q_n^{(1)} < \infty$, and furthermore $\lim_{n\to\infty} q_n^{(1)} = 0$ if $\mathfrak{X}$ is separable.

For the measure μ the *k-th average Kolmogorov width* of order $r > 0$ is defined by

$$d_k^{(r)} = \inf\{\Delta^{(r)}(\mu, \widetilde{\mu}) : \dim(\operatorname{span}(\operatorname{supp}(\widetilde{\mu}))) \le k\},$$

and we have

$$d_k^{(r)} = \inf_{\dim(\mathfrak{X}_0)=k} \left(\int_{\mathfrak{X}} \inf_{x_0 \in \mathfrak{X}_0} \|x - x_0\|^r \, \mu(dx) \right)^{1/r}.$$

See, e.g., Ritter (2000, Sec. VII.2.5) and Creutzig (2002) for results and references.

4. General Results for Full Space Sampling

Recall that full space sampling corresponds to the cost model defined by $c = 1$, see (6).

4.1. **Deterministic Algorithms.** The quantization problem and the quadrature problem by means of deterministic algorithms using full space sampling are equivalent in the following sense. Since S is a real-valued linear mapping on a convex and symmetric set F, it follows that

$$e_{N,\mathrm{full}}^{\mathrm{det}} = \inf\{e(\widehat{\mathcal{S}}) : \widehat{\mathcal{S}} \in \mathbb{S}^{\mathrm{det}} \text{ is a quadrature formula, } \mathrm{cost}_1(\widehat{\mathcal{S}}) \le N\}, \tag{8}$$

see Smolyak (1965), Bakhvalov (1971), and also Traub, Wasilkowski, Woźniakowski (1988, Chap. 4.5). Furthermore, for F and S as studied in this paper we have

$$\inf\{e(\widehat{\mathcal{S}}) : \widehat{\mathcal{S}} \in \mathbb{S}^{\mathrm{det}} \text{ is a quadrature formula with nodes } x_1, \dots, x_N\} = q^{(1)}(x_1, \dots, x_N)$$

for every codebook $x_1, \dots, x_N \in \mathfrak{X}$, see Kantorovich, Rubinstein (1958) and Gray, Neuhoff, Shields (1975). The latter infimum is attained by a quadrature formula with induced mapping

$$\widehat{S}(f) = \sum_{\ell=1}^{N} \mu(V_\ell) \cdot f(x_\ell), \tag{9}$$

if $V_1, \dots, V_N$ is a corresponding Voronoi partition of $\mathfrak{X}$. An (almost) optimal codebook therefore yields an (almost) optimal quadrature formula, and the N-th minimal error $e_{N,\mathrm{full}}^{\mathrm{det}}$ coincides with the N-th quantization number of order one.

Theorem 1. *For every $N \in \mathbb{N}$*

$$e_{N,\mathrm{full}}^{\mathrm{det}} = q_N^{(1)}.$$

4.2. **Randomized Algorithms.** We first state an upper bound for the minimal error $e_{N,\mathrm{full}}^{\mathrm{ran}}$ in terms of the quantization number $q_N^{(2)}$, which is a consequence of a well-known variance reduction technique based on quantization, see Pagès, Printems (2005). Note that $\lim_{N\to\infty} q_N^{(2)} = 0$ if $\mathfrak{X}$ is separable and $\int_{\mathfrak{X}} \|x\|^2 \, \mu(dx) < \infty$. Under the latter assumption the classical Monte Carlo method $\widehat{\mathcal{S}}_{N,\mu}$ without variance reduction, see Example 3, only yields errors of order $N^{-1/2}$ in all non-trivial cases.

Theorem 2. *For every $N \in \mathbb{N}$*

$$e_{2N,\mathrm{full}}^{\mathrm{ran}} \le 2 \cdot N^{-1/2} \cdot q_N^{(2)}.$$

Proof. Consider $x_1, \dots, x_N \in \mathfrak{X}$ as well as a corresponding Voronoi partition $V_1, \dots, V_N$ of $\mathfrak{X}$. For $f \in F$ let $J(f)$ denote the interpolation of f at the points x_ℓ that is constant on the corresponding cells V_ℓ, i.e.,

$$J(f) = \sum_{\ell=1}^{N} f(x_\ell) \cdot 1_{V_\ell}.$$

Thus a deterministic algorithm with induced mapping given by (9) approximates $S(f)$ by $S(J(f))$. Consider a randomized algorithm $\widehat{\mathcal{S}} \in \mathbb{S}^{\mathrm{ran}}$ with $\mathrm{cost}_1(\widehat{\mathcal{S}}) \le 2N$ and induced mapping given by

$$\widehat{S}(f) = S(J(f)) + 1/N \cdot \sum_{\ell=1}^{N} (f - J(f))(X_\ell),$$

where $X_1, \dots, X_N$ are independent and distributed according to μ. Hence the non-deterministic part of $\widehat{\mathcal{S}}$ consists of applying the classical Monte Carlo method $\widehat{\mathcal{S}}_{N,\mu}$ to $\widetilde{f} = f - J(f)$. It follows that

$$e(\widehat{\mathcal{S}}) = N^{-1/2} \cdot \sup_{f \in F} \left(\int_{\mathfrak{X}} \left(\widetilde{f}(x) - S(\widetilde{f}) \right)^2 \mu(dx) \right)^{1/2}.$$

Since $|\widetilde{f}(x)| \le \min_{\ell=1,\dots,N} \|x - x_\ell\|$, we obtain

$$\begin{aligned} \left(\int_{\mathfrak{X}} \left(\widetilde{f}(x) - S(\widetilde{f}) \right)^2 \mu(dx) \right)^{1/2} &\le \left(\int_{\mathfrak{X}} \widetilde{f}^2(x)\, \mu(dx) \right)^{1/2} + |S(\widetilde{f})| \\ &\le q^{(2)}(x_1, \dots, x_N) + q^{(1)}(x_1, \dots, x_N) \\ &\le 2 \cdot q^{(2)}(x_1, \dots, x_N), \end{aligned}$$

which completes the proof. □

We now turn to lower bounds for randomized algorithms. In this setting a result analogous to (8) is not available in general, and therefore considerations cannot a priori be restricted to randomized quadrature formulas. We use the following tool, which is due to Bakhvalov (1959) and Novak (1988) and which holds for integration problems in general, see Novak (1988, Sec. 2.2.10).

Proposition 1. *Let $m \ge 4N$, and suppose there are functionals $f_1, \dots, f_m : \mathfrak{X} \to \mathbb{R}$ such that*

$$\{x \in \mathfrak{X} : f_i(x) \neq 0\} \cap \{x \in \mathfrak{X} : f_j(x) \neq 0\} = \emptyset \tag{10}$$

for all $i \neq j$ and

$$\sum_{i=1}^{m} \delta_i \cdot f_i \in F \tag{11}$$

for all $\delta_1, \dots, \delta_m \in \{\pm 1\}$. Then

$$e^{\mathrm{ran}}_{N,\mathrm{full}} \ge \tfrac{1}{4} \cdot N^{1/2} \cdot \min_{i=1,\dots,m} S(f_i).$$

A proper choice of the functionals f_i in Proposition 1 yields a lower bound for the minimal error $e^{\mathrm{ran}}_{N,\mathrm{full}}$ in terms of consecutive differences of quantization numbers of order one.

Theorem 3. *For every $N \in \mathbb{N}$*

$$e^{\mathrm{ran}}_{N,\mathrm{full}} \ge \tfrac{1}{8} \cdot N^{1/2} \cdot \sup_{m \ge 4N} (q^{(1)}_{m-1} - q^{(1)}_m).$$

Proof. For $\varepsilon \in]0,1[$ and $m \geq 4N$ choose $x_1, \ldots, x_m \in \mathfrak{X}$ with

$$q^{(1)}(x_1, \ldots, x_m) \leq \varepsilon \cdot q^{(1)}_{m-1} + (1-\varepsilon) \cdot q^{(1)}_m + \varepsilon, \tag{12}$$

and consider the functionals

$$f_i(x) = \tfrac{1}{2} \cdot \max(0, \min_{j \neq i} \|x - x_j\| - \|x - x_i\|), \qquad i = 1, \ldots, m.$$

Clearly (10) is satisfied and $f_1, \ldots, f_m \in F$. Consequently (11) holds, too.

We claim that

$$S(f_i) \geq \frac{1-\varepsilon}{2} \cdot (q^{(1)}_{m-1} - q^{(1)}_m) - \varepsilon. \tag{13}$$

It suffices to prove the statement for $i = m$. To this end consider a Voronoi partition $V_1, \ldots, V_m$ corresponding to $x_1, \ldots, x_m$, and let $U_1, \ldots, U_{m-1}$ be a Voronoi partition corresponding to $x_1, \ldots, x_{m-1}$. If $j \leq m-1$ and $x \in V_m \cap U_j$ then

$$f_m(x) = \tfrac{1}{2} \cdot (\|x - x_j\| - \|x - x_m\|).$$

Hence

$$\begin{aligned}
\int_{\mathfrak{X}} f_m(x)\, \mu(dx) &= \int_{V_m} f_m(x)\, \mu(dx) \\
&= \tfrac{1}{2} \cdot \sum_{j=1}^{m-1} \int_{V_m \cap U_j} \|x - x_j\|\, \mu(dx) - \tfrac{1}{2} \cdot \int_{V_m} \|x - x_m\|\, \mu(dx) \\
&= \tfrac{1}{2} \cdot \sum_{j=1}^{m-1} \int_{(V_m \cap U_j) \cup V_j} \|x - x_j\|\, \mu(dx) - \tfrac{1}{2} \cdot \sum_{j=1}^{m} \int_{V_j} \|x - x_j\|\, \mu(dx).
\end{aligned}$$

Note that the sets $(V_m \cap U_j) \cup V_j$ with $j \leq m-1$ form a partition of $\mathfrak{X}$ as well, and every $x \in (V_m \cap U_j) \cup V_j$ satisfies

$$\min_{k=1,\ldots,m-1} \|x - x_k\| = \|x - x_j\|.$$

Thus

$$S(f_m) = \tfrac{1}{2} \cdot (q^{(1)}(x_1, \ldots, x_{m-1}) - q^{(1)}(x_1, \ldots, x_m))$$

and (13) follows from (12).

It remains to apply Proposition 1 and to let ε tend to zero. □

The following consequence of Theorem 3 is useful, in particular, for finite-dimensional spaces $\mathfrak{X}$, see Section 5. Recall that a sequence of positive real numbers a_n is regularly varying with index $-\alpha < 0$ if $\lim_{n\to\infty} a_{\lfloor \kappa n \rfloor}/a_n = \kappa^{-\alpha}$ for every $\kappa > 0$.

Corollary 1. *If the sequence $(q^{(1)}_n)_{n\in\mathbb{N}}$ is regularly varying with index $-\alpha < 0$ then*

$$\liminf_{N\to\infty} e^{\mathrm{ran}}_{N,\mathrm{full}} \cdot N^{1/2} / q^{(1)}_N \geq \frac{\alpha}{2^{5+2\alpha}}.$$

Proof. Let $q_n = q^{(1)}_n$ for notational convenience, and put

$$g(m) = \sup_{\ell \geq m} (q_{\ell-1} - q_\ell)$$

for $m \in \mathbb{N} \setminus \{1\}$ and let $\kappa > 1$. Clearly,

$$g(m) \geq \frac{q_m - q_{\lceil \kappa m \rceil}}{\lceil \kappa m \rceil - m} = q_m \cdot \frac{1 - q_{\lceil \kappa m \rceil}/q_m}{\lceil \kappa m \rceil - m}.$$

Since $\lim_{m\to\infty} q_{\lceil \kappa m\rceil}/q_m = \kappa^{-\alpha}$ it follows that

$$\liminf_{m\to\infty}\, g(m)\cdot m/q_m \geq \frac{1-\kappa^{-\alpha}}{\kappa-1}.$$

Letting κ tend to one yields

$$\liminf_{m\to\infty}\, g(m)\cdot m/q_m \geq \alpha.$$

Combining the latter estimate and Theorem 3 completes the proof. □

Corollary 1 is not applicable if the quantization numbers are slowly varying, which often holds true in infinite-dimensional spaces $\mathfrak{X}$, see Sections 8 and 9. Instead, one may use the following result.

Corollary 2. *Let $f : [0,\infty[\to]0,\infty[$ be a convex and differentiable function. If*

$$\limsup_{n\to\infty}\, q_n^{(1)}/f(n) \geq 1$$

and

$$\lim_{n\to\infty} q_n^{(1)} = 0$$

then

$$\limsup_{N\to\infty}\, e^{\mathrm{ran}}_{N,\mathrm{full}}/\left(N^{1/2}\cdot|f'|(4N+3)\right) \geq 1/8.$$

Proof. Fix $\varepsilon \in]0,1[$, and put $q_n = q_n^{(1)}$. By assumption

$$q_{m-1} \geq (1-\varepsilon)\cdot f(m-1) = (1-\varepsilon)\cdot\int_{m-1}^{\infty} -f'(s)\,ds$$

holds for infinitely many integers m. Since $q_{m-1} = \sum_{k=m}^{\infty}(q_{k-1}-q_k)$, we also have

$$q_{m-1}-q_m \geq (1-\varepsilon)\cdot\int_{m-1}^{m} -f'(s)\,ds \geq -(1-\varepsilon)\cdot f'(m)$$

infinitely often. To every such m we associate $N = \lfloor m/4\rfloor$. Then $m \in [4N, 4N+3]$ and Theorem 3 implies

$$e^{\mathrm{ran}}_{N,\mathrm{full}} \geq -(1-\varepsilon)/8\cdot N^{1/2}\cdot f'(4N+3).$$

Letting ε tend to zero finishes the proof. □

5. Application to Finite-Dimensional Spaces $\mathfrak{X}$

There are numerous results on $e^{\mathrm{det}}_{N,\mathrm{full}}$ or $q_N^{(1)}$ for finite-dimensional spaces $\mathfrak{X} = \mathbb{R}^s$, see, e.g., Novak (1988), Graf, Luschgy (2000), Wasilkowski, Woźniakowski (1996, 2001). Assume $r \geq 1$. Then, under rather mild assumptions on μ, and in particular for the uniform distribution on $[0,1]^s$, the quantization numbers satisfy

$$\lim_{N\to\infty} q_N^{(r)}\cdot N^{1/s} = \gamma^{(r)} \tag{14}$$

with some constant $\gamma^{(r)} = \gamma^{(r)}(\mu, s, \|\cdot\|) > 0$, see Graf, Luschgy (2000, Thm. 6.2).

Suppose that (14) is satisfied. Then Corollary 1 is applicable with $\alpha = 1/s$, and we obtain

$$\liminf_{N\to\infty}\, e^{\mathrm{ran}}_{N,\mathrm{full}}\cdot N^{1/2+1/s} \geq \frac{\gamma^{(1)}}{s\cdot 2^{5+2/s}}.$$

A matching upper bound is provided by Theorem 2, so that we end up with the well-known fact

$$e^{\mathrm{ran}}_{N,\mathrm{full}} \asymp N^{-1/2-1/s},$$

see Novak (1988, Sec. 2.2.6) for the case of the uniform distribution μ on $[0,1]^s$.

We conclude that, up to multiplicative constants, neither the upper bound in Theorem 2 nor the lower bound in Theorem 3 can be improved in general.

6. General Results for Fixed Subspace Sampling

Throughout this section we consider randomized algorithms in a cost model defined by (4) with any choice of a fixed subspace. Furthermore, we assume that μ has a finite second moment, i.e.,

$$K = \left(\int_{\mathfrak{X}} \|x\|^2 \, \mu(dx) \right)^{1/2} < \infty. \tag{15}$$

We immediately get an upper bound for the error of the classical Monte Carlo method $\widehat{\mathcal{S}}_{n,\nu}$, see Example 3.

Lemma 1. *For every $n \in \mathbb{N}$,*

$$e(\widehat{\mathcal{S}}_{n,\nu}) \le \left((\Delta^{(2)}(\mu,\nu))^2 \cdot (1+2/n) + 2K^2/n \right)^{1/2}.$$

Proof. Take $k = 1$ in Lemma 2 below. □

Consequently we obtain an upper bound for the minimal errors in terms of the Kolmogorov widths of order two.

Theorem 4. *Let $\gamma > \max(\sqrt{2}\,K, \sqrt{3})$. Then, for every $N \in \mathbb{N}$,*

$$e^{\mathrm{ran}}_{N,\mathrm{fix}} \le \gamma \cdot \inf_{n \cdot k \le N} \left(n^{-1} + \left(d^{(2)}_k\right)^2 \right)^{1/2}.$$

Proof. Suppose that $n, k \in \mathbb{N}$ with $n \cdot k \le N$ and take a Borel probability measure ν on $\mathfrak{X}$ such that

$$\dim(\mathrm{span}(\mathrm{supp}(\nu))) \le k, \qquad \Delta^{(2)}(\nu,\mu) \le \gamma/\sqrt{3} \cdot d^{(2)}_k.$$

Then

$$e(\widehat{\mathcal{S}}_{n,\nu}) \le \gamma \cdot \left(n^{-1} + \left(d^{(2)}_k\right)^2 \right)^{1/2}$$

by Lemma 1, and it remains to observe that $\mathrm{cost}_c(\widehat{\mathcal{S}}_{n,\nu}) \le N$ for the fixed subspace model corresponding to $\mathfrak{X}_0 = \mathrm{span}(\mathrm{supp}(\nu))$. □

The following lower bound for the minimal errors corresponds to the extremal cases, where either the dimension k of the subspace or the number n of evaluations may be arbitrarily large. A corollary to this lower bound involves quantization numbers, too, via Theorem 3.

Theorem 5. *For every $N \in \mathbb{N}$*

$$e^{\mathrm{ran}}_{N,\mathrm{fix}} \ge \inf_{n \cdot k \le N} \max(e^{\mathrm{ran}}_{n,\mathrm{full}}, d^{(1)}_k).$$

Proof. Consider any randomized algorithm $\widehat{\mathcal{S}}$ with $\mathrm{cost}_c(\widehat{\mathcal{S}}) \le N$ in some fixed subspace model. Let $\mathfrak{X}_0$ denote the corresponding subspace, and assume that $\dim(\mathfrak{X}_0) = k$. Define a functional $f_0 \in F$ by $f_0 = \mathrm{dist}(\cdot, \mathfrak{X}_0)$. Since $\widehat{\mathcal{S}}$ evaluates f_0 only at points from $\mathfrak{X}_0$ with probability one, we have

$$\widehat{S}(f_0) = \widehat{S}(-f_0)$$

for the induced mapping, and consequently

$$e(\widehat{\mathcal{S}}) \geq \tfrac{1}{2} \cdot \left(\left(\mathbb{E}|S(f_0) - \widehat{S}(f_0)|^2 \right)^{1/2} + \left(\mathbb{E}|S(-f_0) - \widehat{S}(-f_0)|^2 \right)^{1/2} \right) \geq S(f_0).$$

Hence

$$e(\widehat{\mathcal{S}}) \geq \int_{\mathfrak{X}} \inf_{x_0 \in \mathfrak{X}_0} \|x - x_0\| \, \mu(dx) \geq d_k^{(1)}.$$

On the other hand, put $n = \mathrm{cost}_1(\widehat{\mathcal{S}})$ to obtain

$$e(\widehat{\mathcal{S}}) \geq e_{n,\mathrm{full}}^{\mathrm{ran}}.$$

We conclude that $e(\widehat{\mathcal{S}}) \geq \max(e_{n,\mathrm{full}}^{\mathrm{ran}}, d_k^{(1)})$ for some $k, n \in \mathbb{N}$ such that $k \cdot n \leq N$. □

7. General Results for Variable Subspace Sampling

Throughout this section we consider randomized algorithms in a cost model defined by (5) with any choice of an increasing sequence of subspaces. As in Section 6 we assume that μ has a finite second moment, see (15).

Instead of a classical Monte Carlo method, which is based on independent and identically distributed sampling, we employ a multilevel Monte Carlo algorithm to derive an upper bound for the minimal error $e_{N,\mathrm{var}}^{\mathrm{ran}}$. This algorithm is defined as follows. Consider a sequence

$$\boldsymbol{\rho} = (\rho^{(1)}, \ldots, \rho^{(k)})$$

of probability measures on $\mathfrak{X} \times \mathfrak{X}$, and assume that the marginals $\rho_1^{(j)}$ and $\rho_2^{(j)}$ of the measures $\rho^{(j)}$ satisfy

$$\rho_2^{(j)} = \rho_1^{(j+1)}, \qquad j = 1, \ldots, k-1.$$

Furthermore, let

$$\boldsymbol{n} = (n_1, \ldots, n_k) \in \mathbb{N}^k$$

and consider an independent sequence of $\mathfrak{X} \times \mathfrak{X}$-valued random variables

$$X_\ell^{(j)} = \big(X_{\ell,1}^{(j)}, X_{\ell,2}^{(j)}\big), \qquad \ell = 1, \ldots, n_j, \ j = 1, \ldots, k,$$

with $X_1^{(j)}, \ldots, X_{n_j}^{(j)}$ being distributed according to $\rho^{(j)}$.

A corresponding *multilevel Monte Carlo algorithm* is a randomized quadrature rule with random nodes $X_{\ell,i}^{(j)}$ and deterministic weights $a_{\ell,i}^{(j)} = (-1)^i / n_j$ for $j + i \geq 3$. Slightly abusing notation, we use $\widehat{\mathcal{S}}_{\boldsymbol{n},\boldsymbol{\rho}}$ to denote any such algorithm. Thus, $\widehat{\mathcal{S}}_{\boldsymbol{n},\boldsymbol{\rho}}$ induces the mapping

$$\widehat{S}_{\boldsymbol{n},\boldsymbol{\rho}}(f) = \frac{1}{n_1} \sum_{\ell=1}^{n_1} f\big(X_{\ell,2}^{(1)}\big) + \sum_{j=2}^{k} \frac{1}{n_j} \sum_{\ell=1}^{n_j} \big(f\big(X_{\ell,2}^{(j)}\big) - f\big(X_{\ell,1}^{(j)}\big) \big),$$

and we have

$$\mathbb{E}(\widehat{S}_{\boldsymbol{n},\boldsymbol{\rho}}(f)) = \int_{\mathfrak{X}} f(x) \, \rho_2^{(k)}(dx).$$

For the analysis of $\widehat{\mathcal{S}}_{\boldsymbol{n},\boldsymbol{\rho}}$ we put

$$D(x) = (x, x), \qquad x \in \mathfrak{X},$$

and we use the Wasserstein distance $\Delta_2^{(2)}$ of order two on the space of Borel probability measures on $\mathfrak{X} \times \mathfrak{X}$ equipped with the norm $\|(x_1, x_2)\|_2 = (\|x_1\|^2 + \|x_2\|^2)^{1/2}$. As a generalization of Lemma 1 we have the following estimate.

Lemma 2. *Every multilevel algorithm $\widehat{\mathcal{S}}_{\boldsymbol{n},\boldsymbol{\rho}}$ satisfies*

$$e(\widehat{\mathcal{S}}_{\boldsymbol{n},\boldsymbol{\rho}}) \le \left(\big(\Delta^{(1)}(\mu,\rho_2^{(k)})\big)^2 + \frac{1}{n_1}\cdot\big(\Delta^{(2)}(\mu,\rho_2^{(1)})+K\big)^2 + \sum_{j=2}^{k}\frac{1}{n_j}\big(\Delta_2^{(2)}(D(\mu),\rho^{(j)})\big)^2\right)^{1/2}.$$

Proof. Let $f\in F$. Clearly,

$$|S(f)-\mathbb{E}(\widehat{S}_{\boldsymbol{n},\boldsymbol{\rho}}(f))| = \left|\int_{\mathfrak{X}} f(x)\,\mu(dx) - \int_{\mathfrak{X}} f(x)\,\rho_2^{(k)}(dx)\right| \le \Delta^{(1)}(\mu,\rho_2^{(k)})$$

for the bias of $\widehat{S}_{\boldsymbol{n},\boldsymbol{\rho}}(f)$. Since

$$\left(\int_{\mathfrak{X}\times\mathfrak{X}} \|x_1-x_2\|^2\,\rho^{(j)}(d(x_1,x_2))\right)^{1/2} \le \Delta_2^{(2)}(D(\mu),\rho^{(j)})$$

and

$$\left(\int_{\mathfrak{X}} \|x\|^2\,\rho_2^{(1)}(dx)\right)^{1/2} \le \Delta^{(2)}(\mu,\rho_2^{(1)})+K,$$

we obtain

$$\begin{aligned}\mathbb{V}(\widehat{S}_{\boldsymbol{n},\boldsymbol{\rho}}(f)) &\le \frac{1}{n_1}\int_{\mathfrak{X}}\|x\|^2\,\rho_2^{(1)}(dx) + \sum_{j=2}^{k}\frac{1}{n_j}\int_{\mathfrak{X}\times\mathfrak{X}}\|x_1-x_2\|^2\,\rho^{(j)}(d(x_1,x_2))\\ &\le \frac{1}{n_1}\cdot\big(\Delta^{(2)}(\mu,\rho_2^{(1)})+K\big)^2 + \sum_{j=2}^{k}\frac{1}{n_j}\big(\Delta_2^{(2)}(D(\mu),\rho^{(j)})\big)^2\end{aligned}$$

for the variance of $\widehat{S}_{\boldsymbol{n},\boldsymbol{\rho}}(f)$, which completes the proof. □

We apply Lemma 2 in the following way. Consider a probability space $(\mathfrak{Y},\mathfrak{B},Q)$ as well as measurable mappings $\pi:\mathfrak{Y}\to\mathfrak{X}$ and $\pi_j:\mathfrak{Y}\to\mathfrak{X}$ such that

$$\mu=\pi(Q)$$

and

$$0<\dim(\operatorname{span}(\pi_j(\mathfrak{Y})))\le j, \qquad j\in\mathbb{N}. \tag{16}$$

Let $k\in\mathbb{N}$ and consider a multilevel algorithm

$$\widehat{\mathcal{S}}^{(k)} = \widehat{\mathcal{S}}_{\boldsymbol{n},\boldsymbol{\rho}} \tag{17}$$

by taking

$$\rho^{(j)} = (\pi_{2^{j-1}},\pi_{2^j})(Q), \qquad n_j = \lceil 2^{k-j}/(3k)\rceil$$

for $j=1,\dots,k$.

Lemma 3. *Let $\gamma = 4\sqrt{6}\,\max(K,1)$. Then every multilevel algorithm $\widehat{\mathcal{S}}^{(k)}$ satisfies*

$$e(\widehat{\mathcal{S}}^{(k)}) \le \gamma\cdot k^{1/2}\cdot 2^{-(k+3)/2}\cdot\left(1+\sum_{j=1}^{k}2^j\cdot\int_{\mathfrak{Y}}\|\pi(y)-\pi_{2^j}(y)\|^2\,Q(dy)\right)^{1/2}$$

and

$$\operatorname{cost}_c(\widehat{\mathcal{S}}^{(k)}) \le 2^{k+3}$$

in the variable subspace model corresponding to the increasing sequence of subspaces

$$\mathfrak{X}_j = \operatorname{span}\left(\bigcup_{\ell=1}^{j}\pi_{2^\ell}(\mathfrak{Y})\right).$$

Proof. Put $\theta_j = \int_{\mathfrak{Y}} \|\pi - \pi_{2^j}\|^2 \, dQ$ for notational convenience. We have

$$\Delta^{(1)}(\mu, \rho_2^{(k)}) \le \theta_k^{1/2}, \qquad \Delta^{(2)}(\mu, \rho_2^{(1)}) \le \theta_1^{1/2},$$

and

$$\Delta_2^{(2)}(D(\mu), \rho^{(j)}) \le (\theta_j + \theta_{j-1})^{1/2}.$$

Hence, by Lemma 2,

$$\begin{aligned} e^2(\widehat{\mathcal{S}}^{(k)}) &\le \theta_k + (\theta_1^{1/2} + K)^2 \cdot \frac{3k}{2^{k-1}} + \sum_{j=2}^{k} (\theta_j + \theta_{j-1}) \cdot \frac{3k}{2^{k-j}} \\ &\le 6 \cdot K^2 \cdot \frac{k}{2^{k-1}} + 12 \cdot \frac{k}{2^k} \cdot \sum_{j=1}^{k} 2^j \cdot \theta_j \\ &\le \gamma^2 \cdot \frac{k}{2^{k+3}} \cdot \left(1 + \sum_{j=1}^{k} 2^j \cdot \theta_j\right). \end{aligned}$$

Use $\dim(\mathfrak{X}_j) \le 2^{j+1} - 2$ to obtain

$$\begin{aligned} \mathrm{cost}_c(\widehat{\mathcal{S}}^{(k)}) &\le n_1 \cdot \dim(\mathfrak{X}_1) + \sum_{j=2}^{k} n_j \cdot (\dim(\mathfrak{X}_j) + \dim(\mathfrak{X}_{j-1})) \\ &\le 2^k/(3k) + 2 + \sum_{j=2}^{k} (2^{k-j}/(3k) + 1) \cdot 3 \cdot 2^j \\ &= 2^k \cdot (3k-2)/(3k) + 2 + 3 \cdot (2^{k+1} - 4) \\ &\le 2^{k+3}. \end{aligned}$$

□

Remark 3. Multilevel Monte Carlo methods taking values in infinite-dimensional Banach spaces have been introduced by Heinrich (1998) and Heinrich and Sindambiwe (1999) for computation of global solutions of integral equations and for parametric integration, respectively. Moreover, the authors have shown that suitable multilevel algorithms turned out to be (almost) optimal in both cases. See Heinrich (2001) for further results and references.

In the context of quadrature problems for diffusion processes multilevel algorithms have been studied by Giles (2006), while a two level algorithm has already been considered by Kebaier (2005). Both papers also include numerical examples from computational finance.

A suitable multilevel algorithm yields the following upper bound for the minimal errors in terms of the Kolmogorov widths of order two.

Theorem 6. *Let $\gamma > 2\sqrt{6}\,\max(K, 1)$. Then, for every $N \ge 16$,*

$$e_{N,\mathrm{var}}^{\mathrm{ran}} \le \gamma \cdot (\log_2 N/N)^{1/2} \cdot \left(1 + \sum_{j=1}^{\lfloor \log_2 N \rfloor} 2^j \cdot \left(d_{2^j}^{(2)}\right)^2\right)^{1/2}.$$

Proof. Let $\varepsilon > 0$. Take a sequence of subspaces $\widetilde{\mathfrak{X}}_j \subset \mathfrak{X}$ such that $\dim(\widetilde{\mathfrak{X}}_j) \le j$ and

$$\left(\int_{\mathfrak{X}} \inf_{\widetilde{x} \in \widetilde{\mathfrak{X}}_j} \|x - \widetilde{x}\|^2 \, \mu(dx)\right)^{1/2} \le (1 + \varepsilon) \cdot d_j^{(2)}.$$

Thanks to the Kuratowski and Ryll-Nardzewski selection theorem there exist measurable selections $\pi_j : \mathfrak{X} \to \mathfrak{X}$ for the metric projection onto $\widetilde{\mathfrak{X}}_j$. Hence

$$\left(\int_{\mathfrak{X}} \|x - \pi_j(x)\|^2 \, \mu(dx) \right)^{1/2} \leq (1+\varepsilon) \cdot d_j^{(2)}.$$

Apply Lemma 2 with $\mathfrak{Y} = \mathfrak{X}$, $Q = \mu$, $\pi = \mathrm{id}$, and $k = \lfloor \log_2(N/8) \rfloor$ to complete the proof. □

Similar to Theorem 5 we obtain a lower bound for the minimal errors.

Theorem 7. *For every $N \in \mathbb{N}$*

$$e^{\mathrm{ran}}_{N,\mathrm{var}} \geq \max(e^{\mathrm{ran}}_{N,\mathrm{full}}, \tfrac{1}{2} \cdot d^{(1)}_{2N}).$$

Proof. The lower bound $e^{\mathrm{ran}}_{N,\mathrm{var}} \geq e^{\mathrm{ran}}_{N,\mathrm{full}}$ is already stated in Remark 2. Next, suppose that a randomized algorithm $\widehat{\mathcal{S}}$ is based on an increasing sequence of subspaces $\mathfrak{X}_i$ and satisfies $\mathrm{cost}_c(\widehat{\mathcal{S}}) \leq N$ in the corresponding sampling model. Put

$$i_0 = \max\{i \in \mathbb{N} : \dim(\mathfrak{X}_i) \leq 2N\}$$

and $f_0 = \mathrm{dist}(\cdot, \mathfrak{X}_{i_0})$ as well as

$$A = \{\omega \in \Omega : \mathrm{cost}_c(\widehat{\mathcal{S}}(\omega), f_0) \leq 2N\}.$$

Clearly $P(A) \geq 1/2$, and for $\omega \in A$ the algorithm $\widehat{\mathcal{S}}(\omega)$ evaluates f_0 only at points from $\mathfrak{X}_{i_0}$. Now, proceed as in the proof of Theorem 5 to obtain

$$e(\widehat{\mathcal{S}}) \geq P(A) \cdot S(f_0) \geq \tfrac{1}{2} \cdot d^{(1)}_{2N},$$

which completes the proof. □

8. Application to Gaussian Measures on Infinite-Dimensional Spaces

In this section we consider zero mean Gaussian measures μ on separable Banach spaces $\mathfrak{X}$, and throughout we assume that the corresponding small ball function

$$\varphi(\varepsilon) = -\ln \mu(\{x \in \mathfrak{X} : \|x\| \leq \varepsilon\})$$

satisfies

$$\varphi(\varepsilon) \asymp \varepsilon^{-\alpha} \cdot (\ln \varepsilon^{-1})^{\beta} \tag{18}$$

for some constants $\alpha > 0$ and $\beta \in \mathbb{R}$ as ε tends to zero.

Remark 4. Typically, (18) holds for infinite-dimensional spaces $\mathfrak{X}$, see Li, Shao (2001). For example, if μ is the distribution of a fractional Brownian motion with Hurst parameter $H \in]0,1[$ on $\mathfrak{X} = C([0,1])$ or $\mathfrak{X} = L_p([0,1])$ for some $p \in [1,\infty[$, then $\alpha = 1/H$ and $\beta = 0$. Moreover, $\alpha = 1/(H-\gamma)$ and $\beta = 0$ when $\|\cdot\|$ denotes the γ-Hölder norm. Similar results are known for Sobolev norms, see Kuelbs, Li, Shao (1995) and Li, Shao (1999).

If $\mathfrak{X} = C([0,1]^2)$ and μ is the distribution of the two-dimensional fractional Brownian sheet, then $\alpha = 1/H$ and $\beta = 1 + 1/H$ due to Belinsky, Linde (2002). Moreover, for a d-dimensional Brownian sheet considered in $\mathfrak{X} = L_2([0,1]^d)$ one has $\alpha = 2$ and $\beta = 2(d-1)$, see Csáki (1984) and Fill, Torcaso (2004).

Assumption (18) determines the asymptotic behavior of the quantization numbers and the Kolmogorov widths, see Dereich (2003, Thm. 3.1.2) and Creutzig (2002, Cor. 4.7.2).

Proposition 2. *The quantization numbers* $q_n^{(r)}$ *satisfy*

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

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

$$d_k^{(r)} \asymp k^{-1/\alpha} \cdot (\ln k)^{\beta/\alpha}$$

for every $r > 0$.

Hence, by Theorem 1, quadrature of arbitrary Lipschitz functionals by means of deterministic algorithms is intractable even in the full space sampling model. Now we turn to the analysis of randomized algorithms in the different sampling models. We write $a_N \preceq b_N$ for sequences of positive real numbers a_N and b_N if $\sup_{N\in\mathbb{N}} a_N/b_N < \infty$.

Theorem 8. *For full space sampling the minimal errors satisfy*

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

and

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

Proof. The upper bound immediately follows from Proposition 2 and Theorem 2. To prove the lower bound apply Corollary 2 with f given by $f(t) = c \cdot (\ln t)^{-1/\alpha} \cdot (\ln \ln t)^{\beta/\alpha}$ for t sufficiently large and a suitable constant $c > 0$. □

Theorem 9. *For fixed subspace sampling the minimal errors satisfy*

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

and

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

Proof. Proposition 2 and Theorem 4 yield the upper bound with the following choice

$$n_N = \lfloor N^{2/(2+\alpha)} \cdot (\ln N)^{-2\beta/(2+\alpha)} \rfloor \tag{19}$$

of numbers of replications and

$$k_N = \lfloor N^{\alpha/(2+\alpha)} \cdot (\ln N)^{2\beta/(2+\alpha)} \rfloor \tag{20}$$

of dimensions of subspaces.

For the proof of the lower bound we combine Proposition 2 with Theorems 5 and 8. Due to Theorem 8 there exists a constant $\gamma > 0$ and an increasing sequence of integers $n_\ell \in \mathbb{N}$ such that

$$e_{n_\ell,\mathrm{full}}^{\mathrm{ran}} \geq \gamma \cdot n_\ell^{-1/2} \cdot (\ln n_\ell)^{-1-1/\alpha} \cdot (\ln \ln n_\ell)^{\beta/\alpha}$$

for every $\ell \in \mathbb{N}$. Put

$$N_\ell = \big\lfloor n_\ell^{(2+\alpha)/2} \cdot (\ln n_\ell)^{\alpha+\beta+1} \cdot (\ln \ln n_\ell)^{-\beta} \big\rfloor,$$

and let $n, k \in \mathbb{N}$ with $n \cdot k \leq N_\ell$. If $n > n_\ell$ then $k < N_\ell/n_\ell$, and Proposition 2 implies

$$d_k^{(1)} \geq d_{\lfloor N_\ell/n_\ell \rfloor}^{(1)} \succeq (N_\ell/n_\ell)^{-1/\alpha} \cdot (\ln(N_\ell/n_\ell))^{\beta/\alpha} \asymp n_\ell^{-1/2} \cdot (\ln n_\ell)^{-1-1/\alpha} \cdot (\ln \ln n_\ell)^{\beta/\alpha}. \tag{21}$$

On the other hand, if $n \leq n_\ell$ then $e_{n,\mathrm{full}}^{\mathrm{ran}} \geq e_{n_\ell,\mathrm{full}}^{\mathrm{ran}}$. Consequently, by Theorem 5 and (21)

$$e_{N_\ell,\mathrm{fix}}^{\mathrm{ran}} \succeq n_\ell^{-1/2} \cdot (\ln n_\ell)^{-1-1/\alpha} \cdot (\ln \ln n_\ell)^{\beta/\alpha}.$$

Straightforward computations show

$$n_\ell^{-1/2} \cdot (\ln n_\ell)^{-1-1/\alpha} \cdot (\ln \ln n_\ell)^{\beta/\alpha} \asymp N_\ell^{-1/(2+\alpha)} \cdot (\ln N_\ell)^{-(2+2\alpha-\alpha\beta)/(\alpha(2+\alpha))} \cdot (\ln \ln N_\ell)^{2\beta/(\alpha(2+\alpha))},$$

which completes the proof of the lower bound. □

Remark 5. The upper bound for fixed subspace sampling (in Theorem 4) is established by sampling from distributions ν that are order optimal approximations to μ in the sense of Kolmogorov widths. Often such approximations are not known explicitly and sampling is not feasible.

Therefore, it is quite common to approximately compute the integrals $S(f)$ with respect to Gaussian measures by sampling from a normal distribution on a suitable finite-dimensional subspace of $\mathfrak{X}$. A proper choice of the subspace is suggested by the following general result on average linear widths, which is due to Creutzig (2002, Thm. 4.4.1). There exist points $x_{\ell,k} \in \mathfrak{X}$ and bounded linear functionals $\xi_{\ell,k} \in \mathfrak{X}^*$ such that

$$\Delta^{(2)}(\mu, \pi_k(\mu)) \le \left(\int_{\mathfrak{X}} \|x - \pi_k(x)\|^2 \, \mu(dx) \right)^{1/2} \preceq \ln k \cdot d_k^{(2)} \tag{22}$$

for

$$\pi_k(x) = \sum_{\ell=1}^{k} \xi_{\ell,k}(x) \cdot x_{\ell,k}. \tag{23}$$

Consider the classical Monte Carlo algorithm $\widehat{\mathcal{S}}_{n,\nu_k}$ that is based on the normal distribution

$$\nu_k = \pi_k(\mu)$$

on $\mathfrak{X}_0 = \operatorname{span}\{x_{1,k}, \ldots, x_{k,k}\}$, see Example 3. Choose

$$n_N = \lceil N^{2/(2+\alpha)} \cdot (\ln N)^{-2(\alpha+\beta)/(2+\alpha)} \rceil$$

and

$$k_N = \lceil N^{\alpha/(2+\alpha)} \cdot (\ln N)^{2(\alpha+\beta)/(2+\alpha)} \rceil,$$

and put

$$\widehat{\mathcal{S}}_N = \widehat{\mathcal{S}}_{n_N,\nu_{k_N}}$$

to obtain $\operatorname{cost}_c(\widehat{\mathcal{S}}_N) \preceq N$ in the corresponding fixed subspace cost models, and

$$e(\widehat{\mathcal{S}}_N) \preceq N^{-1/(2+\alpha)} \cdot (\ln N)^{(\alpha+\beta)/(2+\alpha)}$$

by Lemma 1.

A slightly better upper bound is available if the Banach space $\mathfrak{X}$ is B-convex, e.g., if $\mathfrak{X}$ is an L_p-space with $p \in]1, \infty[$. Instead of (22) we then have

$$\left(\int_{\mathfrak{X}} \|x - \pi_k(x)\|^2 \, \mu(dx) \right)^{1/2} \preceq d_k^{(2)}, \tag{24}$$

see Creutzig (2002, Cor. 3.4.2), which yields

$$e(\widehat{\mathcal{S}}_N) \preceq N^{-1/(2+\alpha)} \cdot (\ln N)^{\beta/(2+\alpha)} \tag{25}$$

as in Theorem 9, if the parameters n_N and k_N are chosen according to (19) and (20).

Both of the estimates (22) and (24) are proven non-constructively, so that it remains to actually determine the underlying normal distributions ν_k. For a number of Gaussian measures the Karhunen-Loéve expansion is explicitly known, and hereby we get an approximation π_k that

satisfies (24), if $\mathfrak{X}$ is an L_2-space. In particular for an L_2-space $\mathfrak{X}$ and $\beta = 0$ the upper bound (25) is due to Wasilkowski, Woźniakowski (1996, p. 2076).

Consider the distribution μ of the d-dimensional fractional Brownian sheet with Hurst parameter $H \in]0,1[$ on the space $\mathfrak{X} = C([0,1]^d)$. In this case a direct approach yields

$$\left(\int_{\mathfrak{X}} \|x - \pi_k(x)\|^2 \, \mu(dx) \right)^{1/2} \asymp k^{-H} \cdot (\ln k)^{H(d-1)+d/2}, \tag{26}$$

see Kühn, Linde (2002). See also Ayache, Taqqu (2003) for a wavelet approximation π_k in the case $d = 1$ and Dzhaparidze, van Zanten (2005) for a trigonometric approximation π_k in the case $d \geq 1$, which both satisfy this estimate. From (26) and Lemma 1 we get

$$e(\widehat{\mathcal{S}}_N) \asymp N^{-1/(2+1/H)} \cdot (\ln N)^{d/2-1/(2+1/H)}.$$

Theorem 10. *For variable subspace sampling the minimal errors are bounded as follows. If $\alpha > 2$, then*

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

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

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

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

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

If $0 < \alpha < 2$, then

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

and

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

Proof. For the proof of the upper bounds we study the asymptotic behaviour of

$$g(N) = \sum_{j=1}^{\lfloor \log_2 N \rfloor} 2^j \cdot \left(d^{(2)}_{2^j}\right)^2$$

because of Theorem 6. By Proposition 2

$$2^j \cdot \left(d^{(2)}_{2^j}\right)^2 \asymp 2^{j\cdot(1-2/\alpha)} \cdot j^{2\beta/\alpha}.$$

Thus, if $\alpha > 2$, then

$$g(N) \asymp N^{1-2/\alpha} \cdot (\ln N)^{2\beta/\alpha}.$$

For $\alpha = 2$ we obtain

$$g(N) \asymp (\ln N)^{(\beta+1)_+}$$

if $\beta \neq -1$, while

$$g(N) \asymp \ln \ln N$$

if $\beta = -1$. Finally, if $0 < \alpha < 2$, then

$$g(N) \asymp 1.$$

Apply Theorem 6 to obtain the upper bounds as claimed.

For the proof of the lower bounds we combine Theorem 7 with Proposition 2 in the case $\alpha \geq 2$ and with Theorem 8 in the case $0 < \alpha < 2$. □

Remark 6. In Remark 5 we have discussed classical Monte Carlo algorithms based on normal distributions $\pi_k(\mu)$. Now we discuss their multilevel counterparts based on the normal distributions

$$\rho^{(j)} = (\pi_{2^{j-1}}, \pi_{2^j})(\mu).$$

Choose

$$k_N = \lceil \log_2 N \rceil,$$

and put

$$\widehat{\mathcal{S}}_N = \widehat{\mathcal{S}}^{(k_N)},$$

see (17). By Lemma 3 we have $\mathrm{cost}_c(\widehat{\mathcal{S}}_N) \preceq N$ in the variable subspace cost model corresponding to the sequence of subspaces $\mathfrak{X}_j = \mathrm{span}\{x_{\ell,2^m} : 1 \le \ell \le 2^m, 1 \le m \le j\}$. Furthermore, (22) implies

$$e(\widehat{\mathcal{S}}_N) \preceq u_N(\alpha, \alpha+\beta),$$

where $u_N(\alpha,\beta)$ denotes the upper bound for the minimal error $e^{\mathrm{ran}}_{N,\mathrm{var}}$ in Theorem 10. If the stronger bound (24) holds for the sequence π_j, then we even have

$$e(\widehat{\mathcal{S}}_N) \preceq u_N(\alpha,\beta)$$

as in Theorem 10.

Remark 7. Due to Theorems 8–10 we have sharp upper and lower bounds on the minimal errors in the different cost models, up to logarithmic terms and up to the fact that some of the lower bounds are established only for an infinite sequence of integers N. Furthermore, all of the classical Monte Carlo algorithms $\widehat{\mathcal{S}}_N$ presented in Remark 5 are almost optimal in this sense, i.e., there exist constants $\gamma_i > 0$ such that

$$e(\widehat{\mathcal{S}}_N) \le \gamma_1 \cdot e^{\mathrm{ran}}_{N,\mathrm{fix}} \cdot (\ln N)^{\gamma_2}$$

holds at least for infinitely many integers N. Similarly, all of the multilevel algorithms $\widehat{\mathcal{S}}_N$ presented in Remark 6 satisfy

$$e(\widehat{\mathcal{S}}_N) \le \gamma_1 \cdot e^{\mathrm{ran}}_{N,\mathrm{var}} \cdot (\ln N)^{\gamma_2}$$

for at least infinitely many integers N.

For full space sampling the bounds depend on the specific properties of the Gaussian measure only via logarithmic terms, and the order of the polynomial term is

$$\gamma = 1/2.$$

This is no longer the case for subspace sampling, where the order of the polynomial term is

$$\gamma = \min(1/2, 1/\alpha)$$

for variable subspace sampling and

$$\gamma = 1/(2+\alpha)$$

for fixed subspace sampling. We conclude that variable subspace sampling is as powerful as full space sampling if $0 < \alpha \le 2$ and always superior to fixed subspace sampling. In the latter comparison the extremal case occurs for $\alpha = 2$.

9. Application to Diffusion Processes

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

$$\begin{aligned} dX_t &= a(X_t)\, dt + b(X_t)\, dW_t, \\ X_0 &= u_0 \in \mathbb{R}^m \end{aligned} \tag{27}$$

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

(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 Kolmogorov widths. The corresponding proofs are postponed to Section 9.2.

Proposition 3. *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 $r > 0$.

The asymptotic behavior of the quantization numbers stated in Proposition 3 is partially known. Luschgy and Pagès (2006) study scalar stochastic differential equations under suitable growth and smoothness conditions. In this work the upper bound is established for equations with a strictly positive diffusion coefficient $b : [0,1] \times \mathbb{R} \to \mathbb{R}$, and a matching lower bound is derived if $\inf_{(t,x) \in [0,1] \times \mathbb{R}} b(t,x) > 0$ and $r \ge 1$. More generally, 1-dimensional diffusions are analyzed by Dereich (2007a, 2007b), who determines the exact asymptotic behavior of the quantization numbers for $r \ge 1$ under rather mild smoothness assumptions. The asymptotic behavior of the Kolmogorov widths is determined by Maiorov (1992, 1993) for the Brownian motion.

Observing Theorem 1 we conclude that quadrature of arbitrary Lipschitz functionals is intractable even in the full space sampling model by means of deterministic algorithms. Furthermore, the estimates from Theorems 8–10 with $\alpha = 2$ and $\beta = 0$ are valid, too, in the diffusion case.

Theorem 11. *Let $\mathfrak{X} = C$ or $\mathfrak{X} = L_p$. For full space sampling the minimal errors satisfy*

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

and

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

For fixed subspace sampling the minimal errors satisfy

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

and

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

For variable subspace sampling the minimal errors satisfy

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

Remark 8. Consider the strong Euler scheme with uniform step-size $1/k$, i.e., $X^{(k)}(0) = u_0$ and

$$X^{(k)}((\ell+1)/k) = X^{(k)}(\ell/k) + a(X^{(k)}(\ell/k)) \cdot 1/k + b(X^{(k)}(\ell/k)) \cdot (W((\ell+1)/k) - W(\ell/k)),$$

and let $X^{(k)} = (X^{(k)}(t))_{t\in[0,1]}$ denote the piecewise linear interpolation of $X^{(k)}(0), \dots, X^{(k)}(1)$ at the breakpoints ℓ/k. We then have the well-known error bounds

$$\left(E\|X - X^{(k)}\|^2\right)^{1/2} \preceq k^{-1/2} \tag{28}$$

for $\mathfrak{X} = L_p$ and

$$\left(E\|X - X^{(k)}\|^2\right)^{1/2} \preceq k^{-1/2} \cdot (\ln k)^{1/2} \tag{29}$$

for $\mathfrak{X} = C$. Furthermore, the distribution ν_k of $X^{(k)}$ on the space $\mathfrak{X}$ satisfies

$$\Delta^{(2)}(\mu, \nu_k) \le \left(E\|X - X^{(k)}\|^2\right)^{1/2}.$$

Hence Lemma 1 yields the following upper bounds for the classical Monte Carlo method

$$\widehat{\mathcal{S}}_N = \widehat{\mathcal{S}}_{n_N, \nu_{k_N}},$$

see Example 3, with a suitable choice of $n_N, k_N \in \mathbb{N}$ such that $\mathrm{cost}_c(\widehat{\mathcal{S}}_N) \preceq N$ in the corresponding fixed subspace model. In the L_p-case we take $n_N = k_N = \lfloor N^{1/2} \rfloor$ to obtain

$$e(\widehat{\mathcal{S}}_N) \preceq N^{-1/4}$$

as in Theorem 11. For $\mathfrak{X} = C$ we take $n_N = \lceil N^{1/2} \cdot (\ln N)^{-1/2} \rceil$ and $k_N = \lceil N^{1/2} \cdot (\ln N)^{1/2} \rceil$ to obtain

$$e(\widehat{\mathcal{S}}_N) \preceq N^{-1/4} \cdot (\ln N)^{1/4}.$$

Now we turn to a multilevel construction that is based on the joint distributions $\rho^{(j)}$ of $(X^{(2^{j-2})}, X^{(2^{j-1})})$, with $X^{(2^{-1})} = u_0$, say. Note that $\rho^{(j)}$ can easily be derived from an $m \cdot 2^{j-1}$-dimensional standard normal distribution. Choose $k_N = \lceil \log_2 N \rceil$ and consider the corresponding multilevel algorithm

$$\widehat{\mathcal{S}}_N = \widehat{\mathcal{S}}^{(k_N)},$$

see (17). Apply Lemma 3 to obtain $\mathrm{cost}_c(\widehat{\mathcal{S}}_N) \preceq N$ in the variable subspace cost model corresponding to the sequence of subspaces $\mathfrak{X}_j$ that consist of piecewise linear functions with breakpoints at $\ell/2^{j-1}$, respectively. Furthermore, (28) implies

$$e(\widehat{\mathcal{S}}_N) \preceq N^{-1/2} \cdot \ln N$$

for $\mathfrak{X} = L_p$ as in Theorem 11, and by (29) we have

$$e(\widehat{\mathcal{S}}_N) \preceq N^{-1/2} \cdot (\ln N)^{3/2}$$

for $\mathfrak{X} = C$.

For full space and fixed subspace sampling we can improve the lower bounds from Theorem 11 in the case $\mathfrak{X} = C$. See Section 9.3 for the corresponding proof.

Theorem 12. *Let $\mathfrak{X} = C$. For full space sampling the minimal errors satisfy*

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

For fixed subspace sampling the minimal errors satisfy

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

Remark 9. Due to Theorems 11 and 12 we have sharp upper and lower bounds on the minimal errors in the different cost models, up to logarithmic terms and up to the fact that, for $\mathfrak{X} = L_p$, some of the lower bounds are established only for an infinite sequence of integers N. Furthermore, the classical Monte Carlo algorithms $\widehat{\mathcal{S}}_N$ presented in Remark 8 are almost optimal in this sense, i.e., there exists a constant $\gamma > 0$ such that

$$e(\widehat{\mathcal{S}}_N) \le \gamma \cdot e^{\mathrm{ran}}_{N,\mathrm{fix}} \cdot \ln N$$

holds for every $N \in \mathbb{N}$ in case of $\mathfrak{X} = C$ and at least for infinitely many integers N in case of $\mathfrak{X} = L_p$. Similarly, the multilevel algorithms $\widehat{S}_N$ presented in Remark 9 satisfy

$$e(\widehat{\mathcal{S}}_N) \le \gamma \cdot e^{\mathrm{ran}}_{N,\mathrm{var}} \cdot \ln N$$

for every $N \in \mathbb{N}$.

9.1. **Preliminaries.** A basic idea in the proofs of Proposition 3 and Theorem 12 is to reduce the case of an m-dimensional diffusion process with properties (i)–(iii) to the particular case of a one-dimensional Brownian motion by means of Lipschitz transformations and stopping.

Let X denote any random element with values in some separable Banach space $\mathfrak{X}$ and consider its distribution μ on this space. We use the notation

$$e^{\mathrm{ran}}_{N,\mathrm{full}}(X, \mathfrak{X}) = e^{\mathrm{ran}}_{N,\mathrm{full}}$$

for the N-th minimal error of randomized algorithms using full space sampling,

$$d^{(p)}_k(X, \mathfrak{X}) = d^{(p)}_k$$

for the k-th average Kolmogorov width of order p, and

$$q^{(r)}_n(X, \mathfrak{X}) = q^{(r)}_n$$

for the n-th quantization number of order r.

Consider a measurable mapping $T : \mathfrak{X} \to \mathfrak{Y}$, where $\mathfrak{Y}$ is a Banach space, too. The following observation is straightforward to verify. We add that an analogous result for Kolmogorov widths is not available.

Lemma 4. *Suppose that T is Lipschitz continuous with a Lipschitz constant $L > 0$. Then*

$$e^{\mathrm{ran}}_{N,\mathrm{full}}(TX, \mathfrak{Y}) \le L \cdot e^{\mathrm{ran}}_{N,\mathrm{full}}(X, \mathfrak{X})$$

and

$$q^{(r)}_n(TX, \mathfrak{Y}) \le L \cdot q^{(r)}_n(X, \mathfrak{X}).$$

We formulate a simplified version of a general relation between quantization numbers and average Kolmogorov widths, which is due to Creutzig (2002, Thm. 4.6.1).

Lemma 5. *For $0 < r < p$*

$$\sup_{n \le 2^\ell} \ln n \cdot q^{(r)}_n(X, \mathfrak{X}) \preceq \sup_{k \le \ell} k \cdot d^{(p)}_k(X, \mathfrak{X}).$$

The following contraction principle holds for best approximation of sums of independent and symmetric random elements.

Lemma 6. *Let $X_1, \dots, X_k$ denote a sequence of independent and symmetric random elements with values in $\mathfrak{X}$ and let $p \geq 1$. Then*

$$\mathbb{E}\Big(\operatorname{dist}^p\Big(\sum_{\ell=1}^k \lambda_\ell X_\ell, \mathfrak{X}_0\Big)\Big) \leq \max_{\ell=1,\dots,k} |\lambda_\ell|^p \cdot \mathbb{E}\Big(\operatorname{dist}^p\Big(\sum_{\ell=1}^k X_\ell, \mathfrak{X}_0\Big)\Big)$$

for all $\lambda_1, \dots, \lambda_k \in \mathbb{R}$ and every closed linear subspace $\mathfrak{X}_0 \subset \mathfrak{X}$.

Proof. Take Rademacher variables $\varepsilon_1, \dots, \varepsilon_k$ such that $\varepsilon_1, \dots, \varepsilon_k, X_1, \dots, X_k$ are independent, and consider the quotient mapping $Q : \mathfrak{X} \to \mathfrak{X}/\mathfrak{X}_0$. Since $(X_1, \dots, X_k)$ and $(\varepsilon_1 X_1, \dots, \varepsilon_k X_k)$ coincide in distribution, the same holds true for $(QX_1, \dots, QX_k)$ and $(\varepsilon_1 QX_1, \dots, \varepsilon_k QX_k)$. Hence

$$\mathbb{E}\Big(\operatorname{dist}^p\Big(\sum_{\ell=1}^k \lambda_\ell X_\ell, \mathfrak{X}_0\Big)\Big) = \mathbb{E}\Big\|\sum_{\ell=1}^k \lambda_\ell \cdot QX_\ell\Big\|^p_{\mathfrak{X}/\mathfrak{X}_0} = \mathbb{E}\Big\|\sum_{\ell=1}^k \lambda_\ell \varepsilon_\ell \cdot QX_\ell\Big\|^p_{\mathfrak{X}/\mathfrak{X}_0}.$$

For any choice of elements $y_\ell \in \mathfrak{X}/\mathfrak{X}_0$

$$\mathbb{E}\Big\|\sum_{\ell=1}^k \lambda_\ell \varepsilon_\ell \cdot y_\ell\Big\|^p_{\mathfrak{X}/\mathfrak{X}_0} \leq \max_{\ell=1,\dots,k} |\lambda_\ell|^p \cdot \mathbb{E}\Big\|\sum_{\ell=1}^k \varepsilon_\ell \cdot y_\ell\Big\|^p_{\mathfrak{X}/\mathfrak{X}_0}$$

due to Kahane's contraction principle, see Kahane (1993, p. 21). Thus

$$\mathbb{E}\Big\|\sum_{\ell=1}^k \lambda_\ell \varepsilon_\ell \cdot QX_\ell\Big\|^p_{\mathfrak{X}/\mathfrak{X}_0} \leq \max_{\ell=1,\dots,k} |\lambda_\ell|^p \cdot \mathbb{E}\Big\|\sum_{\ell=1}^k \varepsilon_\ell \cdot QX_\ell\Big\|^p_{\mathfrak{X}/\mathfrak{X}_0},$$

which completes the proof. □

Now we turn to the diffusion process X given by (27).

Lemma 7. *There exists a neighborhood U of u_0 and a function $f \in C^\infty(U)$ such that*

$$(\nabla f)^* bb^* \nabla f = 1.$$

Proof. Choose a radius $r > 0$ such that $\det bb^*(u) \neq 0$ if $|u - u_0| < r$. Furthermore, take $g \in C^\infty(\mathbb{R}^m, \mathbb{R}^{m\times m})$ with symmetric and positive definite values such that

$$g(u) = (bb^*)^{-1}(u)$$

if $|u - u_0| < r/2$ and $g(u)$ is the identity matrix if $|u - u_0| > r$. Then $M = \mathbb{R}^m$ endowed with the metric tensor $\sum_{i,j=1}^m g_{ij}(u) \cdot du^i \otimes du^j$ is a complete C^∞-Riemannian manifold. Here $u^1, \dots, u^m$ are the local coordinates obtained when taking the identity as chart. Moreover, let d_M denote the corresponding Riemannian distance.

Choose $v_0 \in M$ such that $0 < |v_0 - u_0| < r/2$ and $0 < d_M(v_0, u_0) < i_{v_0}(M)$, where $i_{v_0}(M)$ denotes the injectivity radius at v_0, see Sakai (1996, Prop. III.4.13). Define

$$U = \{u \in M : 0 < |v_0 - u| < r/2,\ 0 < d_m(v_0, u) < i_{v_0}(M)\}$$

as well as

$$f(u) = d_M(v_0, u)$$

for $u \in U$. Then $f \in C^\infty(U)$ and $(\nabla f)^* bb^* \nabla f = 1$, see Sakai (1996, Prop.. III.4.8). □

Lemma 8. *Either let $\mathfrak{X} = C$ and $\mathfrak{Y} = C([0,1], \mathbb{R})$ or let $\mathfrak{X} = L_1$ and $\mathfrak{Y} = L_1([0,1], \mathbb{R})$. There exists a Lipschitz continuous mapping $T : \mathfrak{X} \to \mathfrak{Y}$ and a stopping time τ with $\mathbb{P}(\tau > 0) = 1$ such that the stopped process*

$$(TX)^\tau_t = (TX)_{t\wedge\tau}, \qquad t \in [0,1],$$

is a Brownian motion stopped at time τ.

Proof. Due to Lemma 7 there exists a function $h \in C^\infty(\mathbb{R}^m)$ with bounded derivatives that satisfies

$$(\nabla h)^* bb^* \nabla h = 1 \tag{30}$$

on a closed ball with radius $r > 0$ around u_0. Define the stopping time

$$\tau = \inf\{t \in [0,1] : |X_t - u_0| = r\}.$$

Clearly, $\mathbb{P}(\tau > 0) = 1$.

In both cases cases, $\mathfrak{X} = C$ and $\mathfrak{X} = L_1$ we define a Lipschitz continuous mapping $T : \mathfrak{X} \to \mathfrak{Y}$ by

$$(Tx)(t) = h(x(t)) - h(u_0) - \int_0^t \Big((\nabla h)^* a + \tfrac{1}{2} \cdot \sum_{i,j=1}^m (bb^*)_{i,j} \tfrac{\partial^2}{\partial u^i \partial u^j} h \Big)(x(s))\, ds.$$

Itô's formula implies

$$(TX)_t = \int_0^t \big((\nabla h)^* b \big)(X_t)\, dW_t.$$

Observing (30) we conclude that the stopped process $(TX)^\tau$ is a continuous martingale with quadratic variation

$$\langle (TX)^\tau \rangle_t = \int_0^{t \wedge \tau} \big((\nabla h)^* bb^* \nabla h \big)(X_s)\, ds = t \wedge \tau,$$

which completes the proof. □

Remark 10. The assumption that the diffusion coefficient b is of class C^∞ in a neighborhood of the initial value u_0 can be relaxed. For instance, in the one-dimensional case it suffices to assume $b \in C^1([0,1])$ with Lipschitz continuous first derivative. Then

$$f(u) = \int_{u_0}^u |1/b(v)|\, dv$$

is well defined in a neighborhood of u_0, and the statement of Lemma 8 follows with the same proof.

9.2. **Proof of Proposition 3.** We use the contraction principle from Lemma 6 to establish the upper bound for the Kolmogorov widths.

Lemma 9. *For every $p > 0$*

$$d_k^{(p)}(X, C) \preceq k^{-1/2}.$$

Proof. Assume that $p \geq 1$ without loss of generality. Fix $k \in \mathbb{N}$, put $t_\ell = \ell/k$ for $\ell = 0, \dots, k$, and consider the corresponding Euler process $\overline{X}^{(k)}$ defined by $\overline{X}_0^{(k)} = u_0$ and

$$\overline{X}_t^{(k)} = \overline{X}_{t_\ell}^{(k)} + a(\overline{X}_{t_\ell}^{(k)}) \cdot (t - t_\ell) + b(\overline{X}_{t_\ell}^{(k)}) \cdot (W_t - W_{t_\ell})$$

for $t \in [t_\ell, t_{\ell+1}]$. On the space $\mathfrak{X} = C$ we have

$$\mathbb{E}\|X - \overline{X}^{(k)}\|^p \preceq k^{-p/2},$$

see Bouleau, Lépingle (1994, p. 276), and therefore

$$d_k^{(p)}(X, C) \preceq k^{-1/2} + d_k^{(p)}(\overline{X}^{(k)}, C).$$

Let $\widetilde{W}^{(k)}$ denote the piecewise linear interpolation of the Brownian motion W at the breakpoints t_ℓ and define the continuous process $V^{(k)}$ by

$$V_t^{(k)} = b(\overline{X}_{t_\ell}^{(k)}) \cdot (W_t - \widetilde{W}_t^{(k)})$$

for $t \in [t_\ell, t_{\ell+1}]$. Note that $\overline{X}^{(k)} - V^{(k)}$ takes values in the $(k+1)$-dimensional subspace of piecewise linear functions with breakpoints t_ℓ. Hence

$$d_{2k+1}^{(p)}(\overline{X}^{(k)}, C) \le d_k^{(p)}(V^{(k)}, C).$$

Let $\mathfrak{A}$ denote the σ-algebra generated by $W(t_1), \dots, W(t_k)$. The random variables $b(\overline{X}_{t_\ell}^{(k)})$ are measurable with respect to $\mathfrak{A}$, and conditioned on $\mathfrak{A}$ the process $W - \widetilde{W}^{(k)}$ consists of independent Brownian bridges on the subintervals $[t_\ell, t_{\ell+1}]$. We apply Lemma 6 with $X_\ell = 1_{[t_{\ell-1}, t_\ell]} \cdot (W - \widetilde{W}^{(k)})$ to obtain

$$d_{2k}^{(p)}(V^{(k)}, C) \le \left(\mathbb{E}\|b(\overline{X}^{(k)})\|^p\right)^{1/p} \cdot d_{2k}^{(p)}(W - \widetilde{W}^{(k)}, C) \preceq d_k^{(p)}(W, C).$$

From Maiorov (1993) we get $d_k^{(p)}(W, C) \asymp k^{-1/2}$. □

Lemma 10. *For every $r > 0$*

$$q_n^{(r)}(X, L_1) \succeq (\ln n)^{-1/2}.$$

Proof. Observe that, due to Lemma 4 and Lemma 8, it suffices to show that

$$q_n^{(r)}(Y, L_1) \succeq (\ln n)^{-1/2} \tag{31}$$

for every one-dimensional process Y such that

$$Y_{t\wedge\tau} = W_{t\wedge\tau}, \qquad t \in [0,1],$$

with a stopping time τ that satisfies $P(\tau = 0) = 0$.

To this end fix $\varepsilon \in]0,1]$ with $\mathbb{P}(\tau \ge \varepsilon) > 0$ and define a bounded linear operator $T : L_1 \to L_1$ by

$$(Tx)(t) = \varepsilon^{-1/2} \cdot x(\varepsilon \cdot t).$$

Clearly TW is a Brownian motion, too. The quantization problem for Gaussian processes in the space L^1 is analyzed in Dereich, Scheutzow (2005). In particular there exists a constant $\kappa > 0$ such that

$$\lim_{n\to\infty} (\ln n)^{1/2} \cdot q_n^{(r)}(TW, L_1) = \kappa \tag{32}$$

for every $r > 0$, see Dereich, Scheutzow (2005, Thm. 6.1).

For $n \in \mathbb{N}$ let $M_n \subset L_1$ denote any set of cardinality n, fix $\delta \in]0,1[$, and put

$$A_n = \{\mathrm{dist}(TW, M_n) \ge (1-\delta) \cdot q_n^{(r)}(TW, L_1)\}.$$

Due to (32) we can complement the sets M_n to sets $\widetilde{M}_n$ of cardinality $2n$ such that

$$\lim_{n\to\infty} (\ln n)^{1/2} \cdot \left(\mathbb{E}(\mathrm{dist}^{2r}(TW, \widetilde{M}_n))\right)^{1/2r} = \kappa.$$

as well as

$$\lim_{n\to\infty} (\ln n)^{1/2} \cdot \left(\mathbb{E}(\mathrm{dist}^{r}(TW, \widetilde{M}_n))\right)^{1/r} = \kappa.$$

Employing Lemma A.1 in Dereich, Scheutzow (2005) we conclude that

$$\lim_{n\to\infty} \mathbb{P}(A_n) = 1.$$

Consequently

$$\begin{aligned}
\mathbb{E}(\mathrm{dist}^r(TY, M_n)) &\ge \mathbb{E}(1_{\{\tau\ge\varepsilon\}} \cdot \mathrm{dist}^r(TW, M_n)) \\
&\ge (1-\delta)^r \cdot \mathbb{P}(\{\tau \ge \varepsilon\} \cap A_n) \cdot \left(q_n^{(r)}(TW, L_1)\right)^r \\
&\succeq (\ln n)^{-r/2},
\end{aligned}$$

which yields

$$q_n^{(r)}(TY, L_1) \succeq (\ln n)^{-1/2}.$$

The latter bound implies (31) by Lemma 4. □

Proof of Proposition 3. In view of Lemma 9 and Lemma 10 it suffices to show that

$$q_n^{(r)}(X, C) \preceq (\ln n)^{-1/2} \tag{33}$$

and

$$d_k^{(p)}(X, L_1) \succeq k^{-1/2}. \tag{34}$$

By Lemma 5 and Lemma 9 we have

$$\ln n \cdot q_n^{(r)}(X, C) \preceq \sup_{k \le 2 \ln n} k \cdot d_k^{(2r)}(X, C) \preceq (\ln n)^{1/2},$$

which yields (33). From Lemma 9 we also get

$$d_k^{(p)}(X, L_1) \le c \cdot k^{-1/2} \tag{35}$$

with some constant $c > 0$. Moreover, by Lemma 10,

$$\sup_{n \le 2^\ell} \ln n \cdot q_n^{(p/2)}(X, L_1) \succeq \sup_{n \le 2^\ell} (\ln n)^{1/2} \succeq \ell^{1/2}.$$

Consequently, by Lemma 5

$$\sup_{k \le \ell} k \cdot d_k^{(p)}(X, L_1) \ge \widetilde{c} \cdot \ell^{1/2}$$

with some constant $\widetilde{c} \in]0, c[$. Put $c = (\widetilde{c}/c)^2$. Since

$$\sup_{k < c \cdot \ell} k \cdot d_k^{(p)}(X, L_1) < \widetilde{c} \cdot \ell^{1/2}$$

by (35), we conclude that

$$\ell \cdot d_{\lfloor c \cdot \ell \rfloor}^{(p)}(X, L_1) \ge \sup_{c \cdot \ell \le k \le \ell} k \cdot d_k^{(p)}(X, L_1) \ge \widetilde{c} \cdot \ell^{1/2},$$

which yields (34). □

9.3. **Proof of Theorem 12.** Consider a one-dimensional Brownian motion W. Given $\ell \in \mathbb{N}$ and $\varepsilon \in]0, 1]$ let $s_i = i/\ell \cdot \varepsilon$ and put

$$B_{i,0}^{\ell,\varepsilon} = \{x \in C([0,1]) : x(s_i) - x(s_{i-1}) \ge \varepsilon^{1/2}/\ell^{3/2}\}$$

as well as

$$B_{i,1}^{\ell,\varepsilon} = \{x \in C([0,1]) : x(s_i) - x(s_{i-1}) < -\varepsilon^{1/2}/\ell^{3/2}\}$$

for $i = 1, \dots, \ell$. Moreover, define

$$A_\alpha^{\ell,\varepsilon} = \bigcap_{i=1}^{\ell} \{W \in B_{i,\alpha_i}^{\ell,\varepsilon}\}$$

for any multi-index $\alpha \in \{0,1\}^\ell$.

Lemma 11. *There exists a constant $c_0 \in]0, 1[$ such that*

$$c_0 \cdot 2^{-\ell} \le \mathbb{P}(A_\alpha^{\ell,\varepsilon}) \le 2^{-\ell}$$

for all $\ell \in \mathbb{N}$, $\varepsilon \in]0, 1]$, and $\alpha \in \{0,1\}^\ell$.

Proof. Obviously, the probability $\mathbb{P}(A_\alpha^{\ell,\varepsilon})$ does not depend on α. Hence

$$\begin{aligned}\mathbb{P}(W \in B_{i,0}^{\ell,\varepsilon}) &= \tfrac{1}{2} - \mathbb{P}(0 \le W_{s_i} - W_{s_{i-1}} \le \varepsilon^{1/2}/\ell^{3/2}) \\ &= \tfrac{1}{2} - \int_0^{1/\ell} (2\pi)^{-1/2} \exp(-x^2/2)\, dx \\ &\ge \tfrac{1}{2} \cdot \bigl(1 - \sqrt{2/\pi} \cdot \ell^{-1}\bigr)\end{aligned}$$

implies

$$2^\ell \cdot \mathbb{P}(A_\alpha^{\ell,\varepsilon}) \ge \bigl(1 - \sqrt{2/\pi} \cdot \ell^{-1}\bigr)^\ell.$$

The latter bound tends to $\exp(-\sqrt{2/\pi})$ as ℓ tends to infinity, which completes the proof. □

Let A be any event with $\mathbb{P}(A) \ge 1 - c_0/2$ and put

$$N(\varepsilon, \ell) = \#\{\alpha \in \{0,1\}^\ell : \mathbb{P}(A_\alpha^{\ell,\varepsilon} \cap A) > c_0 \cdot 2^{-\ell-2}\}.$$

Lemma 12. *For all $\varepsilon > 0$ and $\ell \in \mathbb{N}$*

$$N(\varepsilon, \ell) \ge c_1 \cdot 2^\ell,$$

where $c_1 = c_0/(4 - c_0)$.

Proof. Due to Lemma 11

$$\begin{aligned}\mathbb{P}\Bigl(\bigcup_{\alpha\in\{0,1\}^\ell} A_\alpha^{\ell,\varepsilon} \cap A\Bigr) &\ge \mathbb{P}(A) + \mathbb{P}\Bigl(\bigcup_{\alpha\in\{0,1\}^\ell} A_\alpha^{\ell,\varepsilon}\Bigr) - 1 \\ &\ge \mathbb{P}(A) + c_0 - 1 \ge c_0/2.\end{aligned}$$

On the other hand, by the definition of $N(\varepsilon, \ell)$ and Lemma 11

$$\mathbb{P}\Bigl(\bigcup_{\alpha\in\{0,1\}^\ell} A_\alpha^{\ell,\varepsilon} \cap A\Bigr) \le (2^\ell - N(\varepsilon,\ell)) \cdot c_0 \cdot 2^{-\ell-2} + N(\varepsilon,\ell) \cdot 2^{-\ell}.$$

It remains to combine both estimates. □

Proof of Theorem 12. First we establish the lower bound for $e_{N,\mathrm{full}}^{\mathrm{ran}}$. Because of Lemma 4 and Lemma 8 it suffices to prove that

$$e_{N,\mathrm{full}}^{\mathrm{ran}}(Y, C) \succeq N^{-1/2} \cdot (\ln N)^{-3/2} \tag{36}$$

for every one-dimensional process Y such that

$$Y_{t\wedge\tau} = W_{t\wedge\tau}, \qquad t \in [0,1],$$

with a stopping time τ that satisfies $P(\tau = 0) = 0$. To this end we use Proposition 1.

Put

$$B_\alpha^\ell = \bigcap_{i=1}^{\ell} B_{i,\alpha_i}^{\ell,0}$$

and define $f_\alpha^\ell \in F$ by

$$f_\alpha^\ell(x) = \operatorname{dist}\Bigl(x, \bigl(B_\alpha^\ell\bigr)^c\Bigr)$$

for $\alpha \in \{0,1\}^\ell$. Note that

$$f_\alpha^\ell(x) \ge \tfrac{1}{2} \cdot \min_{i=1,\dots,\ell} |x(s_i) - x(s_{i-1})|$$

for $x \in B_\alpha^\ell$. Choose $\varepsilon \in]0,1]$ with $\mathbb{P}(\tau \geq \varepsilon) \geq 1 - c_0/2$, and let $A = \{\tau \geq \varepsilon\}$. Then

$$\begin{aligned} S(f_\alpha^\ell) &\geq \tfrac{1}{2} \cdot \mathbb{E}\Big(1_A \cdot 1_{B_\alpha^\ell}(W) \cdot \min_{i=1,\dots,\ell} |W_{s_i} - W_{s_{i-1}}|\Big) \\ &\geq \tfrac{1}{2} \cdot \mathbb{E}\Big(1_{A_\alpha^{\ell,\varepsilon} \cap A} \cdot \min_{i=1,\dots,\ell} |W_{s_i} - W_{s_{i-1}}|\Big) \\ &\geq \tfrac{1}{2} \cdot \varepsilon^{1/2} / \ell^{3/2} \cdot \mathbb{P}(A_\alpha^{\ell,\varepsilon} \cap A). \end{aligned}$$

Take $N = \lfloor c_1 \cdot 2^{\ell-1} \rfloor$ and use Lemma 12 to conclude that

$$S(f_\alpha^\ell) \succeq N^{-1} \cdot (\ln N)^{-3/2}$$

holds uniformly for at least $2N$ multi-indices $\alpha \in \{0,1\}^\ell$. Finally, apply Proposition 1 to complete the proof of (36).

We combine the lower bound for $e_{N,\mathrm{full}}^{\mathrm{ran}}$ together with Theorem 5 and Proposition 3 to obtain the lower bound for $e_{N,\mathrm{fix}}^{\mathrm{ran}}$. □

Acknowledgments

We are grateful to Karsten Große-Brauckmann for valuable discussions and remarks. Moreover, we thank two anonymous referees for their comments.

References

Ayache, A., Taqqu, M. S. (2003), Rate optimality of wavelet series approximations of fractional Brownian motion, J. Fourier Anal. Appl. **9**, 451–471.

Bakhvalov, N. S. (1959), On approximate computation of integrals (in Russian), Vestnik MGV, Ser. Math. Mech. Astron. Phys. Chem. **4**, 3–18.

Bakhvalov, N. S. (1971), On the optimality of linear methods for operator approximation in convex classes of functions, USSR Comput. Math. Math. Phys. **11**, 244–249.

Bally, V., Talay, D. (1996), The law of the Euler scheme for stochastic differential equations. I: Convergence rate of the distribution function, Probab. Theory Relat. Fields **104**, 43–60.

Belinsky, E., Linde, W. (2002), Small ball probabilities of fractional Brownian sheets via fractional integration operators, J. Theoret. Probab. **15**, 589–612.

Bouleau, N., Lépingle, D. (1994), Numerical Methods for Stochastic Processes, Wiley, New York.

Chen, G., Fang, G., (2004), Probabilistic and average widths of multivariate Sobolev spaces with mixed derivative equipped with the Gaussian measure. J. Complexity **20**, 858–875.

Csáki, E. (1984), 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.

Creutzig, J. (2002), Approximation of Gaussian random vectors in Banach spaces, Ph.D. Dissertation, Universität Jena.

Dereich, S. (2003), High resolution coding of stochastic processes and small ball probabilities. Ph.D. Dissertation, TU Berlin.

Dereich, S. (2007a), The coding complexity of diffusion processes under supremum norm distortion, to appear in Stochastic Process. Appl., doi:10.1016/j.spa.2007.07.003

Dereich, S. (2007b), The coding complexity of diffusion processes under $L^p[0,1]$-norm distortion, to appear in Stochastic Process. Appl., doi:10.1016/j.spa.2007.07.002

Dereich, S., Fehringer, F., Matoussi, A., and Scheutzow, M. (2003), On the link between small ball probabilities and the quantization problem, J. Theoret. Probab. **16**, 249–265.

Dereich, S., Scheutzow, M. (2005), High-resolution quantization and entropy coding for fractional Brownian motion, Electron. J. Probab. **11**, 700–722.

Dzhaparidze, K., van Zanten, H. (2005), Optimality of an explicit series expansion of the fractional Brownian sheet, Stat. Prob. Letters **71**, 295–301.

Fill, J. A., Torcaso, F. (2004), Asymptotic analysis via Mellin transforms for small deviations in L^2-norm of integrated Brownian sheets, Probab. Theory Relat. Fields **130**, 259–288.

Giles, M. B. (2006), Multi-level Monte Carlo path simulation, Report NA-06/03, Oxford Univ. Computing Lab., to appear in Oper. Res.

Graf, S., Luschgy, H. (2000), Foundations of Quantization for Probability Distributions, Lect. Notes in Math. **1730**, Springer-Verlag, Berlin.

Gray, R. M., Neuhoff, D. L., Shields, P. C. (1975), A generalization of Ornstein's $\overline{d}$ distance with applications to information theory, Ann. Appl. Prob. **3**, 315–328.

Heinrich, S. (1998), Monte Carlo complexity of global solution of integral equations, J. Complexity **14**, 151–175.

Heinrich, S. (2001), 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.

Heinrich, S., Sindambiwe, E. (1999), Monte Carlo complexity of parametric integration, J. Complexity **15**, 317–341.

Kahane, J.-P. (1993), Some Random Series of Functions, Cambridge Univ. Press, Cambridge.

Kantorovich, L. V., Rubinstein, G. S. (1958), On a space of completely additive functions (in Russian), Vestnik Leningrad Univ. 13, no. 7, Ser. Mat. Astron. Phys. **2**, 52–59.

Kebaier, A. (2005), Statistical Romberg extrapolation: a new variance reduction method and applications to option pricing, Ann. Appl. Prob. **15**, 2681–2705.

Kühn, T., Linde, W. (2002), Optimal series representation of fractional Brownian sheets, Bernoulli **8**, 669–696.

Kuelbs, J., Li, W. V., Shao Q. M. (1995), Small ball estimates for fractional Brownian motion under Hölder norm and Chung's functional LIL, J. Theoret. Probab. **8**, 361–386.

Li, W.V., Shao, Q.-M. (1999), Small ball estimates for Gaussian processes under the Sobolev norm, J. Theoret. Probab. **12**, 699–720.

Li, W. V., Shao, Q.-M. (2001), 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.

Luschgy, H., Pagès, G. (2004), Sharp asymptotics of the functional quantization problem for Gaussian processes, Ann. Appl. Prob. **32**, 1574–1599.

Luschgy, H., Pagès, G. (2006), Functional quantization of a class of Brownian diffusion: a constructive approach, Stochastic Processes Appl. **116**, 310–336.

Mathé, P. (1990), s–numbers in information–based complexity, J. Complexity **6**, 41–66.

Maiorov, V. (1992), Widths of spaces endowed with a Gaussian measure, Russian Acad. Sci. Dokl. Math. **45**, 305–309.

Maiorov, V. (1993), Average n-widths of the Wiener space in the L_∞-norm, J. Complexity **9**, 222–230.

Nemirovsky, A. S., Yudin, D. B. (1983), Problem Complexity and Method Efficiency in Optimization, Wiley, New York.

Novak, E. (1988), Deterministic and Stochastic Error Bounds in Numerical Analysis, Lect. Notes in Math. **1349**, Springer-Verlag, Berlin.

Pagès, G., Printems, J. (2005), Functional quantization for numerics with an application to option pricing, Monte Carlo Meth. Appl. **11**, 407–446.

Rachev, S. T. (1991), Probability Metrics and the Stability of Stochastic Models, Wiley, Chichester.

Ritter, K. (2000), Average-Case Analysis of Numerical Problems, Lect. Notes in Math. **1733**, Springer-Verlag, Berlin.

Sakai, T. (1996), Riemannian Geometry, Transl. Math. Monogr. **149**, AMS, Rhode Island.

Smolyak, S. A. (1965), On optimal restoration of functions and functionals of them (in Russian), Candidate Dissertation, Moscow State University.

Talay, D., Tubaro, L. (1990), Expansion of the global error for numerical schemes solving stochastic differential equations, Stochastic Anal. Appl. **8**, 483–509.

Traub, J. F., Wasilkowski, G. W., Woźniakowski, H. (1988), Information-Based Complexity, Academic Press, New York.

Wasilkowski, G. W. (1989), Randomization for continuous problems, J. Complexity **5**, 195–218.

Wasilkowski, G. W., Woźniakowski, H. (1996), On tractability of path integration, J. Math. Phys. **37**, 2071–2088.

Wasilkowski, G. W., Woźniakowski, H. (2001), Complexity of weighted approximation over $\mathbb{R}^d$, J. Complexity **17**, 722–740.

Jakob Creutzig, Fachbereich Mathematik, Technische Universität Darmstadt, Schlossgartenstr. 7, 64289 Darmstadt, Germany
E-mail address: creutzig@mathematik.tu-darmstadt.de

Steffen Dereich, Institut für Mathematik, MA 7-5, Fakultät II, Technische Universität Berlin, Strasse des 17. Juni 136, 10623 Berlin, Germany
E-mail address: dereich@math.tu-berlin.de

Thomas Müller-Gronbach, Institut für Mathematische Stochastik, Fakultät für Mathematik, Universität Magdeburg, Postfach 4120, 39016 Magdeburg, Germany
E-mail address: gronbach@mail.math.uni-magdeburg.de

Klaus Ritter, Fachbereich Mathematik, Technische Universität Darmstadt, Schlossgartenstr. 7, 64289 Darmstadt, Germany
E-mail address: ritter@mathematik.tu-darmstadt.de