Monte Carlo Complexity of Parametric Integration

STEFAN HEINRICH AND EUGÈNE SINDAMBIWE

Department of Computer Science, University of Kaiserslautern
Postfach 3049
D-67653 Kaiserslautern, Germany

Abstract

The Monte Carlo complexity of computing integrals depending on a parameter is analyzed for smooth integrands. An optimal algorithm is developed on the basis of a multigrid variance reduction technique. The complexity analysis implies that our algorithm attains a higher convergence rate than any deterministic algorithm. Moreover, because of savings due to computation on multiple grids, this rate is also higher than that of previously developed Monte Carlo algorithms for parametric integration.

1 INTRODUCTION

Multivariate integration is a standard field of application for Monte Carlo methods. Usually, a single integral is approximated. In this paper we study the case of parametric integration, that is, the integral depends on a parameter. Since the solution is now a function (of the parameter), this problem carries features of both integration and approximation. In the parametric case a direct pointwise application of Monte Carlo methods may lead to difficulties, e.g. to nonsmoothness of the resulting curves due to fluctuations. Frolov and Chentsov (1962), Sobol (1962, 1973) have developed and analyzed the method of dependent tests to overcome these difficulties. Recent approaches to this problem are contained in Ermakov and Mikhailov (1982), Mikhailov (1991), Prigarin (1995), and Voytishek (1996).

From the point of view of complexity theory this problem has not been studied before. It is the aim of this paper to provide such an analysis. For the class of r-times continuously differentiable functions we determine the order of the minimal error (except for one case in which a logarithmic gap occurs). This gives matching upper and lower complexity bounds. These bounds are of interest also in relation to other complexity results. They represent a kind of "interpolation" between the two boundary cases of parametric integration − pure integration and pure approximation. It is well-known, that in the first case Monte Carlo methods are superior to deterministic ones (for the considered class), while in the the second case they are not. It turns out that parametric integration shows an intermediate

1

behavior: Monte Carlo is still superior, but to a smaller extend, as the dimension of the parameter space increases. A detailed discussion is given in Section 6. But our analysis yields more than the complexity rates. For the proof of the upper bounds we develop a new algorithm, which is directly implementable and is easily extended to other situations than the model class. This algorithm possesses a new feature − the multigrid structure of variance reduction first developed for integral equations in Heinrich (1998a). This allows considerable savings of arithmetic work as compared to the previous Monte Carlo algorithms for parametric integration mentioned above (see Heinrich, 1998b, for a comparative analysis). The paper can be considered as an application and further development of the ideas in Heinrich (1998a) to another (though related) problem. In fact, the parametric integration problem is conceptually simpler than the full solution of integral equation studied in Heinrich (1998a). So it is a side effect of the present paper that it makes the essence of that technically quite involved approach more transparent. On the other hand, only the case of equal dimension parameters $d_1 = d_2$ of approximation (d_1) and integration (d_2) is truly related to integral equations, while it is the various constellations between d_1 and d_2 which provides the full understanding of the complexity of parametric integration.

In Section 2 we present the required notions from information-based complexity theory, recall related previous results and formulate the main result. Section 3 is devoted to the algorithm description. The analysis of this algorithm and the proof of the upper bound can be found in Section 4. The proof of the lower bound is the contents of Section 5, and in the final Section 6 we give some comments and illustrations of the results obtained.

2 PRELIMINARIES

In this section we formulate the problem to be investigated. Then, in order to be precise about the setting, we give some definitions from information-based complexity theory and Monte Carlo methods, and finally we formulate the main result.

2.1 Problem Formulation

Let $d_1, d_2 \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$ be fixed such that $d_1 + d_2 > 0$. By G_i we denote the d_i-dimensional unit cube, i.e. $G_i := [0,1]^{d_i}$ for $i = 1, 2$. We agree upon setting $[0,1]^0 = \{0\}$. Moreover, the length of some multi-index $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_{d_1+d_2}) \in \mathbb{N}_0^{d_1+d_2}$ is $|\alpha| := \alpha_1 + \alpha_2 + \ldots + \alpha_{d_1+d_2}$. For $r \in \mathbb{N}$ fixed, let $X := C^r(G_1 \times G_2)$, i.e. the space X consists of functions f with continuous partial derivatives $D^\alpha f$ on $G_1 \times G_2$, for all α, $|\alpha| \leq r$. Moreover we set $Y := C(G_1)$. Let $\|\cdot\|$ denote the maximum norm both on Y and $C(G_1 \times G_2)$ and let the norm $\|\cdot\|_r$ of the Banach space X be given by

$$\|f\|_r := \max_{|\alpha| \leq r} \|D^\alpha f\|.$$

2

Let X_0 be the unit ball of X. Our aim is to determine the complexity of numerically approximating the solution operator $S : X \to Y$,

$$(Sf)(s) := \int_{G_2} f(s,t)dt, \ s \in G_1. \tag{1}$$

That is, we study parametric integration: Integrate the family of functions $f(s,t)$ parametrized by $s \in G_1$ over $t \in G_2$. The limiting cases where either d_1 or d_2 is equal to zero were formally included because they represent classical problems of numerical mathematics. In fact, $d_1 = 0$ leads to integration on $C^r(G_2)$, whereas $d_2 = 0$ corresponds to function approximation on $C^r(G_1)$ in the norm of $C(G_1)$. Both are well-analyzed for several settings. The aim of this paper is to study the intermediate cases in which $d_1 \neq 0$ and $d_2 \neq 0$. We present a Monte Carlo method approximating S. Then we show it to be order optimal with respect to the Monte Carlo error, among all methods using (randomized) adaptive standard information of varying cardinality.

2.2 Monte Carlo Setting

We use the general terminology of information-based complexity (IBC), which is explained in more details in Traub, Wasilkowski, and Woźniakowski (1988) or Novak (1988). In order to be as selfcontained as possible, we summarize the IBC notions needed in this paper.

First, we have to specify the type of data, or in IBC terms, the information, which is the input for algorithmic approximations of S. In our case, adaptive standard information is used. This is defined as follows:
Let Λ be defined by

$$\Lambda := \left\{ L_{(s,t)}^\alpha \ : \ \alpha \in \mathbb{N}^{d_1+d_2}, \ |\alpha| \leq r, \ (s,t) \in G_1 \times G_2 \right\},$$

where

$$L_{(s,t)}^\alpha(f) := (D^\alpha f)(s,t).$$

Let $\text{ter}_i : \mathbb{R}^i \to \{0,1\}$, $i \in \mathbb{N}$ be some (termination) functions and let a sequence of mappings $(L_i)_{i \in \mathbb{N}}$ be given such that for $i \in \mathbb{N}$,

$$L_i : X \times \mathbb{R}^{i-1} \to \mathbb{R},$$

and

$$L_i(\,\cdot\,; y_1, \dots, y_{i-1}) \in \Lambda \text{ for all } (y_1, \dots, y_{i-1}) \in \mathbb{R}^{i-1}.$$

Let the (standard information) operator $N : X \to \mathbb{R}^\infty := \cup_{i=1}^\infty \mathbb{R}^i$ be defined as

$$N(f) = [L_1(f), L_2(f; y_1), \dots, L_{n_f}(f; y_1, \dots, y_{n_f-1})],$$

where $y_1 = L_1(f)$ and $y_i = \mathrm{L}_i(f; y_1, \dots, y_{i-1})$, $(i > 1)$. Then $N(f)$ is called (adaptive) standard information. Its cardinality is

$$\text{card}(N(f)) := n_f := \min \left\{ n \in \mathbb{N} \ : \ \text{ter}_n(y_1, \dots, y_n) = 1 \right\}.$$

3

Now, an algorithm is simply any mapping $\varphi : \mathbb{R}^\infty \to Y$. Thus, some information $N(f)$ is used as input data to the algorithm φ, so that the composition $\varphi \circ N : X \to Y$ gives some computational approximation to the solution operator S. Let the set of all information operators N be denoted by $\mathcal{N}$ and the set of all algorithms φ by Φ. Now, we can give the definition of an abstract Monte Carlo method as in Novak (1988), Traub, Wasilkowski, and Woźniakowski (1988), and Heinrich (1994, 1998a).

An abstract Monte Carlo method M is a couple

$$M = \left((\Omega, \Sigma, \mu), (N_\omega, \varphi_\omega)_{\omega \in \Omega} \right),$$

where (Ω, Σ, μ) is a probability space, i.e. Ω is a nonempty set, Σ a σ-algebra of subsets of Ω, and μ a probability measure on Σ. Furthermore $(N_\omega, \varphi_\omega) \in \mathcal{N} \times \Phi$, $(\omega \in \Omega)$ is such that for each $f \in X_0$, $\mathrm{card}(N_\omega(f))$ and $\varphi_\omega(N_\omega(f))$ are Σ-measurabe functions of ω (the latter with respect to the σ-algebra of Borel sets of $C(G_1)$).

Let $\mathcal{M}$ be the class of all Monte Carlo methods. Then the Monte Carlo cardinality of a fixed $M \in \mathcal{M}$ is defined as

$$\mathrm{card}^{mc}(M) := \sup_{f \in X_0} \int_\Omega \mathrm{card}(N_\omega(f)) \, d\mu(\omega).$$

The error of M related to the solution operator S is defined by

$$e^{mc}(S, M) := \sup_{f \in X_0} \int_\Omega \| Sf - \varphi_\omega(N_\omega(f)) \| \, d\mu(\omega). \tag{2}$$

The minimal error among all Monte Carlo methods M with $\mathrm{card}^{mc}(M) \leq n$, or the so-called n-th minimal Monte Carlo error, is

$$e_n^{mc}(S) := \inf_{M \in \mathcal{M}} \left\{ e^{mc}(S, M) : \mathrm{card}^{mc}(M) \leq n \right\}. \tag{3}$$

In our case of standard information, we use the simplest cost model: We assume the functionals L_i, arithmetic operations as well as comparisons to have the same price, and set it to unity. In the following, all of the above operations are referred to as basic operations. Consequently for fixed f and ω, the cost function $\mathrm{cost}(\varphi_\omega, N_\omega(f))$ of some random approximation $\varphi_\omega(N_\omega(f))$ is the sum of $\mathrm{card}(N_\omega(f))$ and the number of basic operations occuring within the computation of $\varphi_\omega(N_\omega(f))$. Therefore, the cost of a Monte Carlo method M is

$$\mathrm{cost}^{mc}(M) := \sup_{f \in X_0} \int_\Omega \mathrm{cost}(\varphi_\omega, N_\omega(f)) d\mu(\omega).$$

Finally, for $\varepsilon \geq 0$ the Monte Carlo complexity $\mathrm{comp}^{mc}(S, \varepsilon)$ is

$$\mathrm{comp}^{mc}(S, \varepsilon) := \inf_{M \in \mathcal{M}} \left\{ \mathrm{cost}^{mc}(M) : e^{mc}(S, M) \leq \varepsilon \right\}. \tag{4}$$

Although we formulate all results in terms of minimal errors, i.e. $e_n^{mc}(S)$, the corresponding complexity statements are easily derived. For the lower bound, this is trivial, since, by definition, we always have $\text{cost}^{mc}(M) \geq \text{card}^{mc}(M)$, while the upper bound follows from the cost analysis of the proposed algorithm given at the end of section 4.

Note, furthermore, that the special case $\Omega = \{\omega_0\}$ corresponds to deterministic methods. Thus, without extra definitions, we consider the deterministic setting as clear. And to avoid any confusion, the deterministic analogues of the above introduced quantities are written with the "det" superscript.

2.3 Main results

First, let us recall the well-known results for the special cases where either $d_1 = 0$ or $d_2 = 0$.

Our formulation uses the asymptotic notation $a_n \preceq b_n$ for sequences of non-negative reals a_n and b_n, which means that there exist some constant $c > 0$ and some $n_0 \in \mathbb{N}$ such that $a_n \leq c\,b_n$ for all $n \geq n_0$. If $a_n \preceq b_n$ and $b_n \preceq a_n$, then we write $a_n \asymp b_n$. We often use the same symbol c for possibly different constants. The following two theorems are folklore.

THEOREM 2.1 *Let $Int : C^r(G_2) \to \mathbb{R}$ denote the integration*

$$Int(f) := \int\limits_{G_2} f(s)ds.$$

Then

$$e_n^{det}(Int) \asymp n^{-r/d_2}, \ \ e_n^{mc}(Int) \asymp n^{-(2r+d_2)/2d_2}. \tag{5}$$

We mention Sard (1949), Bückner (1950) and Nikolskij (1950), who have done the earliest lower bound investigations for quadrature formulae. Bakhvalov (1959, 1961) was the first who pointed out the superiority of Monte Carlo integration over deterministic quadrature formulae. He proved the above result for random information with fixed cardinality. But the same complexity results remain valid for varying cardinality, as Novak (1988) proved. The integration problem was investigated by many other authors for several function spaces and settings. A list of some of them can be found in Traub, Wasilkowski, and Woźniakowski (1988).

On the other hand, Monte Carlo methods do not lead to any improvement for the problem of function approximation.

THEOREM 2.2 *Let Appr denote the function approximation problem, that is the embedding operator from $C^r(G_1)$ to $C(G_1)$. Then*

$$e_n^{det}(Appr) \asymp e_n^{mc}(Appr) \asymp n^{-r/d_1}. \tag{6}$$

The function approximation problem has also been investigated for different function spaces and settings. The reader could consult the above cited sources for more literature.

It is intuitively clear that Monte Carlo methods should lead to some improvement if $d_2 \neq 0$. Moreover, this cannot be as much as for integration as long as $d_1 \neq 0$.

In order to enable comparison, we also give the result in the deterministic setting. We omit the easy proof. The upper bound is obvious from classical function approximation. The lower bound can be obtained by reduction to the case of integration over $G_1 \times G_2$.

THEOREM 2.3 *The n-th minimal deterministic error of the operator S from (1) behaves like*

$$e_n^{det}(S) \asymp n^{-r/(d_1+d_2)}. \tag{7}$$

The following theorem settles the complexity of parametric integration and answers the question of how much better Monte Carlo methods are (as compared to deterministic schemes).

THEOREM 2.4 *For $d_1 \neq 0$ and $d_2 \neq 0$ it holds:*

$$e_n^{mc}(S) \asymp \begin{cases} n^{-(2r+d_2)/2(d_1+d_2)}(\log n)^{1/2}, & \text{if } r > d_1/2 \\ n^{-r/d_1}(\log n)^{r/d_1}, & \text{if } r < d_1/2 \end{cases} \tag{8}$$

and

$$n^{-1/2}(\log n)^{1/2} \preceq e_n^{mc}(S) \preceq n^{-1/2}(\log n)^{3/2} , \text{ if } r = d_1/2 . \tag{9}$$

Note that in asymptotic statements we leave the logarithm unspecified, whereas in cases in which the basis is essential we write e.g. $\log_2 n$ or $\ln n$ to indicate base 2 or the natural logarithm. The proof of the above theorem consists of three parts: First, we give a concrete Monte Carlo approximation A for S. Second, we analyze the behavior of its error $e^{mc}(S, A)$, and gain in this way an upper bound of $e_n^{mc}(S)$. Finally, except for (9), where a gap of $\log n$ will remain, the optimality of the algorithm is established by showing its convergence rate not to be improvable. For this purpose we prove that $e_n^{mc}(S)$ has a lower bound of the same order as the algorithm accuracy.

3   THE ALGORITHM

3.1   Notation and algorithmic preliminaries

From now on we assume $d_1 \neq 0$ and $d_2 \neq 0$. Let $k \in \mathbb{N}_0$ be fixed. Let $\prod_k^{(j)}$ be the partition of G_j $(j = 1, 2)$ into cubes of sidelength 2^{-k} with disjoint interior. The equidistant mesh of sidelength $r^{-1}2^{-k}$ on G_j, $j = 1, 2$ will be denoted by $\Gamma_k^{(j)}$. This means

$$\Gamma_k^{(j)} := \left\{ r^{-1}2^{-k}(i_1, \ldots, i_{d_j}) \, : \, 0 \leq i_1, \ldots, i_{d_j} \leq r2^k \right\}. \tag{10}$$

Let $P_k^{(j)} : C(G_j) \to C(G_j)$ be the d_j-dimensional composite Lagrange interpolation of degree r on $\Gamma_k^{(j)}$. That is, on each cube $Q \in \prod_k^{(j)}$ the function $P_k^{(j)}f$ is the (d_j-dimensional) Lagrange interpolation over the nodes $Q \cap \Gamma_k^{(j)}$. For a

more detailed definition see Heinrich (1998a). Let $\mathcal{P}^r(\prod_k^{(j)})$ denote the space of all functions $g \in C(G_j)$ such that $g|_Q$ is a polynomial of (maximum) degree less than or equal to r for all $Q \in \prod_k^{(j)}$. Clearly, $P_k^{(j)}$ maps $C(G_j)$ into $\mathcal{P}^r(\prod_k^{(j)})$.

Note that for f fixed $P_k^{(j)}f$ is uniquely defined by $\{f(s) \,:\, s \in \Gamma_k^{(j)}\}$. As a consequence, the operator $P_k^{(j)}$ will also be interpreted as defined on $\ell_\infty(\Gamma_k^{(j)})$.

Finally, we also consider the operator $P_k^{(j)}$ as acting in the space $C(G_1 \times G_2)$, meaning that we interpolate with respect to one variable only, leaving the other one fixed. So $P_k^{(1)}$ is defined by $\left(P_k^{(1)}f\right)(s,t) := \left(P_k^{(1)}f(\,\cdot\,,t)\right)(s)$.

For $x \in \mathbb{R}$ the notation $\lceil x \rceil$ means the smallest integer greater than or equal to, and $[x]$ the greatest one smaller than or equal to x.

The Monte Carlo method given below uses the deterministic meshes $\Gamma_k^{(1)}$ with

$$n_{1,k} := |\Gamma_k^{(1)}| = (r2^k + 1)^{d_1} \tag{11}$$

points for the s-component and $n_{2,k}$ (a number still to be chosen) independent, uniformly distributed samples τ_{jk}, $j = 1, \ldots, n_{2,k}$ for the t component. The resulting algorithm is a multilevel procedure in which the sample number $n_{2,k}$ is a decreasing sequence of k, whereas the meshes $\Gamma_k^{(1)}$ become finer with higher values of k. As a result, the deterministic error gets smaller, the stochastic greater, so that with a careful choice of the parameters both errors are in balance. This leads to the optimal convergence order of the algorithm.

3.2   Algorithm description

1. input:
   Cardinality parameter: n
   Function $f : G_1 \times G_2 \to \mathbb{R}$

2. Level parameters:

   •
   $$m := \left[\frac{1}{d_1 + d_2}(\log_2 n) + 1\right] \tag{12}$$

   • Starting level:
   $$\tilde{m} = \begin{cases} m, & \text{if } r \geq d_1/2 \\ 0 & \text{else} \end{cases} \tag{13}$$

   • Final level:
   $$\ell := \begin{cases} \lceil (1 + d_2/2r)m \rceil, & \text{if } r \geq d_1/2 \\ \lceil (1 + d_2/d_1)m \rceil - p & \text{else, where } p := [(\log_2 m)/d_1] \end{cases} \tag{14}$$

3. Number of samples in level k, $k = \tilde{m}, \ldots, \ell$:
   $$n_{2,k} := \begin{cases} \lceil 2^{d_2 m - (r + d_1/2)(k-m)} \rceil, & \text{if } r \geq d_1/2, \\ \lceil 2^{(d_1+d_2)m - d_1 k - (d_1/2 - r)(\ell - k)} \rceil & \text{else} \end{cases} \tag{15}$$

7

4. `Random variables` $\tau_{kj}(\omega)$:
   Let $\tau_{kj} = \tau_{kj}(\omega)$, $(j = 1, \ldots, n_{2,k}; k = \tilde{m}, \ldots, \ell)$ be independent, uniformly distributed on G_2 random variables on some probability space (Ω, Σ, μ).

5. `Monte Carlo sampling on the starting level` $\tilde{m}$:

   a) If $r \geq d_1/2$, compute for all $s \in \Gamma_{\tilde{m}}^{(1)}$

$$\eta_{\tilde{m}}(s) := \int\limits_{G_2} \left(P_{\tilde{m}}^{(2)} f \right)(s, t) dt + \frac{1}{n_{2,\tilde{m}}} \sum_{j=1}^{n_{2,\tilde{m}}} \left(f(s, \tau_{\tilde{m}j}) - \left(P_{\tilde{m}}^{(2)} f \right)(s, \tau_{\tilde{m}j}) \right)$$

(16)

   (Note that the integral is, in fact, just a deterministic multivariate composite Newton Cotes quadrature. So (16) is Monte Carlo with separation of the main part.)

   b) If $r < d_1/2$, put for $s \in \Gamma_{\tilde{m}}^{(1)}$

$$\eta_{\tilde{m}}(s) := \frac{1}{n_{2,\tilde{m}}} \sum_{j=1}^{n_{2,\tilde{m}}} f(s, \tau_{\tilde{m}j})$$

(17)

   (Note that $\tilde{m} = 0$, so we have here the classical Monte Carlo on the roughest grid.)

6. `Monte Carlo sampling on higher levels`:
   For $\tilde{m} < k \leq \ell$, compute for all $s \in \Gamma_k^{(1)}$,

$$\eta_k(s) := \frac{1}{n_{2,k}} \sum_{j=1}^{n_{2,k}} f(s, \tau_{kj})$$

(18)

7. `Final approximation by levelwise interpolation`:

$$Af := P_{\tilde{m}}^{(1)} \eta_{\tilde{m}} + \sum_{k=\tilde{m}+1}^{\ell} (P_k^{(1)} - P_{k-1}^{(1)}) \eta_k$$

(19)

So the result of the algorithm is an element of $\mathcal{P}^r(\prod_{\ell}^{(1)})$.

4 ALGORITHM ANALYSIS

Clearly, the approximation constructed by the algorithm described in the previous section is a Monte Carlo method in the abstract sense. The aim of this section is to analyze this algorithm and to prove the upper bound in Theorem 2.4. Let $\tau : \Omega \to G_2$ be uniformly distributed on G_2. Then for all $f \in X_0$ and $s \in G_1$

$$\mathbf{E}(f(s, \tau)) = Sf(s)$$

(20)

and

$$\mathbf{E}\left(P_k^{(1)} f(\,\cdot\,, \tau) \right) = P_k^{(1)} Sf.$$

(21)

8

Therefore for all $k > \tilde{m}$

$$\mathbf{E}\left(P_k^{(1)} - P_{k-1}^{(1)}\right)\eta_k = P_k^{(1)}Sf - P_{k-1}^{(1)}Sf.$$

And for $k = \tilde{m}$ we get

$$\mathbf{E}P_{\tilde{m}}^{(1)}\eta_{\tilde{m}} = P_{\tilde{m}}^{(1)}Sf.$$

It follows that

$$\mathbf{E}(Af) = P_\ell^{(1)}Sf. \tag{22}$$

We have

$$\mathbf{E}\|Sf - Af\| \leq \|Sf - P_\ell^{(1)}Sf\| + \mathbf{E}\|Af - P_\ell^{(1)}Sf\|. \tag{23}$$

So the error splits into a deterministic and a stochastic part. Classical polynomial approximation gives for $j = 1, 2$,

$$\|f - P_\ell^{(j)}f\| \preceq 2^{-r\ell} \tag{24}$$

(throughout this chapter the constants c as well as those involved in the $\preceq$ and $\asymp$ notation are assumed to be independent of all occurring variables except for a possible dependence on the problem parameters d_1, d_2, and r). For the deterministic part (24) gives

$$\|Sf - P_\ell^{(1)}Sf\| \preceq 2^{-r\ell}. \tag{25}$$

Next we study the stochastic component. For this the following lemma will be useful:

LEMMA 4.1 *There is a constant $c > 0$ such that if $n_1, n_2 \in \mathbb{N}$ and (ρ_j), $j = 1, \ldots, n_2$ is a sequence of independent $\ell_\infty^{n_1}$–valued random variables with finite second moment, then*

$$\mathbf{Var}\left(\sum_{j=1}^{n_2} \rho_j\right) \leq c \log n_1 \sum_{j=1}^{n_2} \mathbf{Var}(\rho_j), \tag{26}$$

where $\mathbf{Var}(\rho) := \mathbf{E}\|\rho - \mathbf{E}\rho\|_Z^2$ denotes the variance of a random variable ρ with values in a Banach space Z.

This follows from Proposition 9.11 of Ledoux and Talagrand (1991), see also Heinrich (1998a). Now we prove

LEMMA 4.2

$$\mathbf{E}\|Af - P_\ell^{(1)}Sf\| \preceq (\log n_{1,\ell})^{1/2}\left(\sum_{k=\tilde{m}}^{\ell} (n_{2,k})^{-1}2^{-2rk}\right)^{1/2} \tag{27}$$

Proof. Let us introduce $\mathcal{P}^r(\prod_k^{(1)})$–valued random variables ζ_{kj} for $j = 1, \ldots, n_{2,k}$ and $k = \tilde{m}, \ldots, \ell$. For $k = \tilde{m}$ and $r \geq d_1/2$ we define

$$\zeta_{\tilde{m}j}(s, \omega) := \left(P_{\tilde{m}}^{(1)}f\right)(s, \tau_{\tilde{m}j}(\omega)) - \left(P_{\tilde{m}}^{(1)}P_{\tilde{m}}^{(2)}f\right)(s, \tau_{\tilde{m}j}(\omega)).$$

We note that

$$P_{\tilde{m}}^{(1)}\eta_{\tilde{m}} = P_{\tilde{m}}^{(1)}SP_{\tilde{m}}^{(2)}f + \frac{1}{n_{2,\tilde{m}}}\sum_{j=1}^{n_{2,\tilde{m}}} \zeta_{\tilde{m}j}, \tag{28}$$

9

$$\mathbf{E}\zeta_{\tilde{m}j} = P_{\tilde{m}}^{(1)}Sf - P_{\tilde{m}}^{(1)}SP_{\tilde{m}}^{(2)}f, \tag{29}$$

and from (24),

$$\|\zeta_{\tilde{m}j}\| \preceq 2^{-r\tilde{m}} \tag{30}$$

for all j and ω. For $k = \tilde{m}$ and $r < d_1/2$ we let

$$\zeta_{\tilde{m}j}(s,\omega) := \left(P_{\tilde{m}}^{(1)}f\right)(s, \tau_{\tilde{m}j}(\omega)). \tag{31}$$

Hence

$$P_{\tilde{m}}^{(1)}\eta_{\tilde{m}} = \frac{1}{n_{2,\tilde{m}}} \sum_{j=1}^{n_{2,\tilde{m}}} \zeta_{\tilde{m}j} \tag{32}$$

$$\mathbf{E}\zeta_{\tilde{m}j} = P_{\tilde{m}}^{(1)}Sf \tag{33}$$

and

$$\|\zeta_{\tilde{m}j}\| \preceq 2^{-r\tilde{m}} \tag{34}$$

(the latter means the boundedness of $\zeta_{\tilde{m}j}$, since $\tilde{m} = 0$).
Finally for $k = \tilde{m} + 1, \ldots, \ell$ we define (for all cases)

$$\zeta_{kj}(s,\omega) := \left((P_k^{(1)} - P_{k-1}^{(1)})f\right)(s, \tau_{kj}(\omega)).$$

Obviously

$$(P_k^{(1)} - P_{k-1}^{(1)})\eta_k = \frac{1}{n_{2,k}} \sum_{j=1}^{n_{2,k}} \zeta_{kj}, \tag{35}$$

$$\mathbf{E}\zeta_{kj} = (P_k^{(1)} - P_{k-1}^{(1)})Sf \tag{36}$$

and from (24),

$$\|\zeta_{kj}\| \preceq 2^{-rk}. \tag{37}$$

Now we represent $Af - P_\ell^{(1)}Sf$ by the help of these random variables. Let us first assume $r \geq d_1/2$. Then, using (19), (28) and (35),

$$Af - P_\ell^{(1)}Sf = P_{\tilde{m}}^{(1)}SP_{\tilde{m}}^{(2)}f + \sum_{k=\tilde{m}}^{\ell} \frac{1}{n_{2,k}} \sum_{j=1}^{n_{2,k}} \zeta_{kj}$$

$$- P_{\tilde{m}}^{(1)}Sf - \sum_{k=\tilde{m}+1}^{\ell} (P_k^{(1)} - P_{k-1}^{(1)})Sf,$$

and from (29) and (36)

$$Af - P_\ell^{(1)}Sf = \sum_{k=\tilde{m}}^{\ell} \frac{1}{n_{2,k}} \sum_{j=1}^{n_{2,k}} (\zeta_{kj} - \mathbf{E}\zeta_{kj}). \tag{38}$$

On the basis of (19), (32), (35) and (33), (36) it is checked analogously that (38) also holds in the case $r < d_1/2$. By Hölder's inequality and (38)

$$\left(\mathbf{E}\|Af - P_\ell^{(1)}Sf\|\right)^2 \leq \mathbf{E}\|Af - P_\ell^{(1)}Sf\|^2 = \mathbf{Var}\left(\sum_{k=\tilde{m}}^{\ell} \frac{1}{n_{2,k}} \sum_{j=1}^{n_{2,k}} \zeta_{kj}\right). \tag{39}$$

10

Since $\mathcal{P}^r(\prod_{\tilde{m}}^{(1)}) \subset \ldots \subset \mathcal{P}^r(\prod_{\ell}^{(1)})$, we can consider all ζ_{kj} as $\mathcal{P}^r(\prod_{\ell}^{(1)})$–valued random variables. On the other hand, it is easily checked that there is a linear isomorphism $U : \mathcal{P}^r(\prod_{\ell}^{(1)}) \rightarrow \ell_{\infty}^{n_{1,\ell}}$ with $\|U\| \cdot \|U^{-1}\| \leq c$, where c does not depend on ℓ. Here $\mathcal{P}^r(\prod_{\ell}^{(1)})$ is considered as equipped with the $C(G_1)$ norm. In fact, as such a U we can take, e.g., the operator of restriction to $\Gamma_{\ell}^{(1)}$. Then the inverse U^{-1} is just the interpolation operator $P_{\ell}^{(1)}$. Because of the isomorphism we can apply Lemma 4.1 with $\ell_{\infty}^{n_1}$ replaced by $\mathcal{P}^r(\prod_{\ell}^{(1)})$ and get

$$\mathbf{Var}\Big(\sum_{k=\tilde{m}}^{\ell} \frac{1}{n_{2,k}} \sum_{j=1}^{n_{2,k}} \zeta_{kj} \Big) \leq c \log n_{1,\ell} \sum_{k=\tilde{m}}^{\ell} (n_{2,k})^{-2} \sum_{j=1}^{n_{2,k}} \mathbf{Var}(\zeta_{kj})$$

$$\leq c \log n_{1,\ell} \sum_{k=\tilde{m}}^{\ell} (n_{2,k})^{-1} 2^{-2rk}$$

because of (30), (34) and (37). This proves the lemma. ∎

PROPOSITION 4.1

$$e^{mc}(S, A) \preceq \begin{cases} n^{-(2r+d_2)/2(d_1+d_2)}(\log n)^{1/2}, & \text{if } r > d_1/2 \\ n^{-r/d_1}(\log n)^{r/d_1}, & \text{if } r < d_1/2 \\ n^{-1/2}\log n, & \text{if } r = d_1/2. \end{cases} \qquad (40)$$

Proof. We consider the cases $r \geq d_1/2$ and $r < d_1/2$ separately.

1. Case $r \geq d_1/2$:

   By (25), (14) and (12) it holds:

   $$\|Sf - P_{\ell}^{(1)}Sf\| \preceq 2^{-r(\log_2 n)(1+d_2/2r)/(d_1+d_2)} \asymp n^{-(2r+d_2)/2(d_1+d_2)}. \qquad (41)$$

   On the other side (11), (14) and (12) imply

   $$\log(n_{1,\ell}) \asymp \ell \asymp \log n,$$

   and (27) together with (12)–(15) yields finally

   $$(\mathbf{E}\|Af - P_{\ell}^{(1)}Sf\|)^2 \preceq \log n \sum_{k=m}^{\ell} 2^{-d_2 m + (r+d_1/2)(k-m) - 2rk}.$$

   We rewrite the exponent on the right hand side as

   $$-d_2 m + (r+d_1/2)(k-m) - 2rk = -d_2 m - 2rm + (r+d_1/2)(k-m) - 2r(k-m)$$

   so that we get

   $$(\mathbf{E}\|Af - P_{\ell}^{(1)}Sf\|)^2 \preceq (\log n)\, 2^{-(2r+d_2)m} \sum_{k=m}^{\ell} 2^{-(r-d_1/2)(k-m)}. \qquad (42)$$

   For $r > d_1/2$ the sum on the right hand side of (42) is a geometric series, bounded from above by some ℓ–independent (therefore also independent from m and n) constant. So, combining (12), (23), (41) and (42) completes the proof for $r > d_1/2$.

For $r = d_1/2$ the sum on the right-hand side of (42) is bounded by ℓ, which leads to

$$(\mathbf{E}\|Af - P_\ell^{(1)}Sf\|)^2 \preceq (\log n)\,\ell\,2^{-(2r+d_2)m}.$$

As final result we get

$$\mathbf{E}\|Af - P_\ell^{(1)}Sf\| \preceq n^{-1/2}\log n,$$

completing the proof of the case $r = d_1/2$.

2. Case $r < d_1/2$:

We proceed analogously. First, (25), (14) and (12) lead to

$$\|Sf - P_\ell^{(1)}Sf\| \preceq 2^{-r\ell} \asymp 2^{-rm(1+d_2/d_1)+r(\log_2 m)/d_1} \asymp n^{-r/d_1}(\log n)^{r/d_1}.$$

Furthermore, by the same steps as above we get

$$(\mathbf{E}\|Af - P_\ell^{(1)}Sf\|)^2 \preceq \log(n_{1,\ell})2^{-(d_1+d_2)m}\sum_{k=0}^{\ell}2^{d_1k+(d_1/2-r)(\ell-k)-2rk}.$$

For the exponent we have $d_1k + (d_1/2 - r)(\ell - k) - 2rk = (d_1/2 - r)(k + \ell)$. Since $r < d_1/2$,

$$(\mathbf{E}\|Af - P_\ell^{(1)}Sf\|)^2 \preceq \log(n_{1,\ell})2^{-(d_1+d_2)m+(d_1-2r)\ell}.$$

We use that $2^{d_1\ell - d_2 m} \asymp 2^{d_1 m - \log_2 m}$ and finally get

$$(\mathbf{E}\|Af - P_\ell^{(1)}Sf\|)^2 \preceq \log(n_{1,\ell})\,2^{-\log_2 m - 2r\ell} \asymp 2^{-2r\ell}$$

as $2^{\log_2 m} \asymp \log(n_{1,\ell})$ by (14). Thus, up to a constant factor, the deterministic and the stochastic errors are bounded from above by $2^{-r\ell} \asymp n^{-r/d_1}(\log n)^{r/d_1}$. This completes the proof of the proposition. ∎

Now we estimate the cost of the algorithm. Let us fix $s \in \Gamma_k^{(1)}$. The computation of $\eta_k(s)$ requires $\mathcal{O}(n_{2,k})$ function values and operations. Indeed, if $k = \tilde{m}$ and $r \geq d_1/2$, then the deterministic part of (16) requires $\mathcal{O}(2^{d_2 m}) = \mathcal{O}(n_{2,\tilde{m}})$ operations and function values. The stochastic parts of (16), (17) and (18) are done using $\mathcal{O}(n_{2,k})$ function values and operations. This has to be multiplied by the cardinality of $\Gamma_k^{(1)}$, which is $\mathcal{O}(n_{1,k})$. The final interpolation procedure (19) can obviously be accomplished in $\mathcal{O}(n_{1,\ell})$ operations. Hence the overall number of function values and arithmetic operations is of the order $\sum_{k=\tilde{m}}^{\ell} n_{1,k}n_{2,k}$. For $r > d_1/2$ we get

$$\sum_{k=\tilde{m}}^{\ell} n_{1,k}n_{2,k} \asymp \sum_{k=\tilde{m}}^{\ell} 2^{d_1 k + d_2 m - (r+d_1/2)(k-m)} \asymp n. \qquad (43)$$

12

The latter step follows from

$$d_1 k + d_2 m - (r + d_1/2)(k - m) = (d_1 + d_2)m - (r - d_1/2)(k - m).$$

For $r = d_1/2$ the same relation (43) gives an order of $n \log n$. For the remaining case $r < d_1/2$ an analogous argument gives

$$\sum_{k=\tilde{m}}^{\ell} n_{1,k} n_{2,k} \asymp \sum_{k=0}^{\ell} 2^{d_1 k} 2^{(d_1+d_2)m - d_1 k - (d_1/2 - r)(\ell - k)} \asymp 2^{(d_1+d_2)m} \asymp n.$$

This easily yields the upper bound of Theorem 2.4. ∎

5   LOWER BOUND

In this part, we present our approach to the lower bound. First, a relationship between the Monte Carlo and the average case error is established.

Let ν be any probability measure with finite support in the unit ball X_0 of X. Recall the following notions of the average case setting. For $N \in \mathcal{N}$ and $\varphi \in \Phi$, let

$$\text{card}^{avg}(N, \nu) := \int_{X_0} \text{card}(N(f)) d\nu(f).$$

For $n \in \mathbb{N}$ we set

$$e^{avg}(S, N, \varphi, \nu) := \int_{X_0} \|Sf - \varphi(N(f))\| d\nu(f)$$

and

$$e_n^{avg}(S, \nu) := \inf \left\{ e^{avg}(S, N, \varphi, \nu) \ : \ \text{card}^{avg}(N, \nu) \le n, \ \varphi \in \Phi \right\}.$$

The first step of our approach consists in reducing the Monte Carlo error to the average error. This idea is due to Bakhvalov (1959). A proof of the following lemma can be found, for example, in Heinrich (1998a).

LEMMA 5.1 *Let ν be any probability measure on X_0 with finite support. Then for any $n \in \mathbb{N}$*

$$e_n^{mc}(S) \ge \frac{1}{2} e_{2n}^{avg}(S, \nu).$$

For the construction of ν we introduce some additional notation. For $m_1, m_2 \in \mathbb{N}$ let

$$I := \{0, \ldots, m_1 - 1\}^{d_1} , \ J := \{0, \ldots, m_2 - 1\}^{d_2}.$$

For $i = (i_1, \ldots, i_{d_1}) \in I$ and $j = (j_1, \ldots, j_{d_2}) \in J$, let the subcubes $G_{1,i}$ of G_1 and $G_{2,j}$ of G_2 be defined by

$$G_{1,i} := \left[\frac{i_1}{m_1}, \frac{i_1 + 1}{m_1} \right] \times \ldots \times \left[\frac{i_{d_1}}{m_1}, \frac{i_{d_1} + 1}{m_1} \right],$$

$$G_{2,j} := \left[\frac{j_1}{m_2}, \frac{j_1 + 1}{m_2} \right] \times \ldots \times \left[\frac{j_{d_2}}{m_2}, \frac{j_{d_2} + 1}{m_2} \right].$$

Moreover let $s_i \in G_{1,i}$ and $t_j \in G_{2,j}$ be given by

$$s_i = (i_1/m_1, \ldots, i_{d_1}/m_1), \; t_j = (j_1/m_1, \ldots, j_{d_2}/m_2).$$

Let ψ_1 and ψ_2 be infinitely differentiable functions with support in G_1 and G_2, respectively, such that

$$\|\psi_1\|_r = \|\psi_2\|_r = 1$$

and

$$\gamma := \int_{G_2} \psi_2(t)dt \neq 0.$$

Let $\gamma_1 := \|\psi_1\|$. Cleary $\gamma_1 \neq 0$. For $i \in I$ and $j \in J$ we set

$$\psi_{1,i}(s) := \psi_1(m_1(s - s_i)), \; \psi_{2,j}(t) := \psi_2(m_2(t - t_j))$$

and define ψ_{ij} by

$$\psi_{ij}(s,t) := \min\{m_1^{-r}, m_2^{-r}\} \, \psi_{1,i}(s) \, \psi_{2,j}(t).$$

Then the following lemma can be shown easily.

LEMMA 5.2 *Let $\lambda_{ij}, i \in I, j \in J$ be real numbers. Then the function f,*

$$f := \sum_{i \in I} \sum_{j \in J} \lambda_{ij} \psi_{ij}$$

fulfills:

$$\|Sf\| = |\gamma|\gamma_1 \min\{m_1^{-r}, m_2^{-r}\} \, m_2^{-d_2} \max_{i \in I} \left| \sum_{j \in J} \lambda_{ij} \right|.$$

LEMMA 5.3 *Let $n_1, n_2 \in \mathbb{N}$ and let $\varepsilon_{ij}, \; i = 1, \ldots, n_1, j = 1, \ldots, n_2$ be independent symmetric $\{-1, 1\}$–valued Bernoulli random variables, i. e. $\mu\{\varepsilon_{ij} = 1\} = \mu\{\varepsilon_{ij} = -1\} = 1/2$. Then*

$$\mathbf{E}\Big(\max_{1 \le i \le n_1} \Big| \sum_{j=1}^{n_2} \varepsilon_{ij} \Big| \Big) \asymp \Big(n_2 \min(n_2, \log(n_1 + 1)) \Big)^{1/2}$$

Proof. The upper bound n_2 is obvious, while $(n_2 \log(n_1 + 1))^{1/2}$ follows from Lemma 4.1.

For the lower bound, we first show

$$\mathbf{E}\Big(\max_{1 \le i \le n_1} \Big| \sum_{j=1}^{n_2} \varepsilon_{ij} \Big| \Big) \ge c(n_2 \log(n_1 + 1))^{1/2}, \;\; \text{for} \;\; 1 \le n_1 \le 2^{n_2}, \qquad (44)$$

with some constant $c > 0$. A proof of (44) for the special case $n_1 = n_2$ can be found on page 120 of Ledoux and Talagrand (1991). The argument immediately carries over to our case. For the sake of completeness we sketch the proof. Relation (4.2) of Ledoux and Talagrand (1991) states that there is a constant $c \ge 1$ such that for all $n \in \mathbb{N}$ and t satisfying

$$c \, n^{1/2} \le t \le c^{-1} n \qquad (45)$$

14

$$\mu\Big\{\sum_{j=1}^{n}\varepsilon_j > t\Big\} \geq \exp(-c\,t^2/n), \tag{46}$$

where ε_j, $j = 1,\ldots,n$ are independent symmetric Bernoulli random variables as above (an elementary proof of this relation is given right there: Ledoux and Talagrand, 1991, p. 90). We choose $n = n_2$ and $t = (c^{-2}(\ln n_1)n_2)^{1/2}$. Then $\exp(c^4) \leq n_1 \leq 2^{n_2}$ implies (45). Consequently, (46) gives for each i

$$\mu\Big\{\sum_{j=1}^{n_2}\varepsilon_{ij} > \big(c^{-2}(\ln n_1)n_2\big)^{1/2}\Big\} \geq \Big(\frac{1}{n_1}\Big)^{1/c} \geq \frac{1}{n_1}.$$

By independence

$$\mu\Big\{\max_{1\leq i\leq n_1}\sum_{j=1}^{n_2}\varepsilon_{ij} > (c^{-2}(\ln n_1)n_2)^{1/2}\Big\} \geq 1 - (1 - \frac{1}{n_1})^{n_1} \geq 1 - e^{-1}.$$

This proves (44) and hence the lemma for $1 \leq n_1 \leq 2^{n_2}$. If $n_1 > 2^{n_2}$, we apply statement (44) with $\tilde{n}_1 = 2^{n_2}$ and get

$$\mathbf{E}\Big(\max_{1\leq i\leq n_1}\Big|\sum_{j=1}^{n_2}\varepsilon_{ij}\Big|\Big) \geq \mathbf{E}\Big(\max_{1\leq i\leq 2^{n_2}}\Big|\sum_{j=1}^{n_2}\varepsilon_{ij}\Big|\Big) \asymp n_2. \qquad \blacksquare$$

PROPOSITION 5.1

$$e_n^{mc}(S) \succeq \begin{cases} n^{-(2r+d_2)/2(d_1+d_2)}(\log n)^{1/2}, & \text{if } r \geq d_1/2 \\ n^{-r/d_1}(\log n)^{r/d_1} & \text{else.} \end{cases} \tag{47}$$

Proof. Let $n \in \mathbb{N}$ be arbitrary. Fix $m_1, m_2 \in \mathbb{N}$ (to be chosen later) in such a way that

$$m_1^{d_1} m_2^{d_2} \geq 4n. \tag{48}$$

Let the index sets I, J and the the functions ψ_{ij}, $i \in I$, $j \in J$ be as introduced above. Let ε_{ij}, $i \in I$, $j \in J$ be independent, symmetric Bernoulli random variables with values in $\{-1, 1\}$, defined on some probability space (Ω, Σ, μ). Let

$$X_1 := \Big\{ f = \sum_{i\in I}\sum_{j\in J}\alpha_{ij}\psi_{ij} \; : \; \alpha_{ij} = \pm 1 \Big\}. \tag{49}$$

By the construction of ψ_{ij}, X_1 is a subset of the unit ball X_0. Let now ν denote the uniform distribution on X_1. Clearly, ν is the distribution of

$$\sum_{i\in I}\sum_{j\in J}\varepsilon_{ij}(\omega)\,\psi_{ij}.$$

We start by applying Lemma 5.1 and estimate the quantity $e_n^{avg}(S, \nu)$ from below. Let N be such that $\text{card}^{avg}(N, \nu) \leq n$. Then for any algorithm $\varphi \in \Phi$ we have

$$e^{avg}(S, N, \varphi, \nu) := \int_{X_1} \|Sf - \varphi(N(f))\| d\nu(f).$$

15

Denote by $\nu(\cdot\,|a)$ the conditional measure defined by the condition $N(f) = a$, with support in $N^{-1}(a) \cap X_1$. Then

$$e^{avg}(S, N, \varphi, \nu) = \int\limits_{N(X_1)} \int\limits_{N^{-1}(a)} \|Sf - \varphi(N(f))\| d\nu(f|a) d\tilde{\nu}(a), \qquad (50)$$

where $\tilde{\nu}$ denotes the measure, which is induced by ν on $N(X_1)$, i.e. $\tilde{\nu} = \nu \circ N^{-1}$. Let $a \in N(X_1)$ be some fixed standard information. Then there exists some set

$$Z_a := \{(s_1, t_1), \ldots, (s_{\hat{n}}, t_{\hat{n}})\}$$

with $(s_i, t_i) \in G_1 \times G_2$ for all $i = 1, \ldots, \hat{n}$, such for all $f \in X_0$ with $N(f) = a$

$$a = (f(s_1, t_1), \ldots, f(s_{\hat{n}}, t_{\hat{n}})).$$

For fixed a, let the set K_a be defined by

$$K_a := \{(i, j) \in I \times J \,:\, Z_a \cap (G_{1,i} \times G_{2,j})^0 = \emptyset\}. \qquad (51)$$

This means that no element of the set Z_a lies in the interior of $G_{1,i} \times G_{2,j}$ for all $(i, j) \in K_a$. For a fixed $f \in X_1$ as defined in (49) let

$$\tilde{f}_a := \sum_{(i,j) \in K_a} \alpha_{ij} \psi_{ij}, \quad \bar{f}_a := f - \tilde{f}_a.$$

Then

$$f = \tilde{f}_a + \bar{f}_a, \quad \text{and} \quad N(-\tilde{f}_a + \bar{f}_a) = a.$$

$\nu(\cdot\,|a)$ is invariant with respect to the mapping $\tilde{f}_a + \bar{f}_a \to -\tilde{f}_a + \bar{f}_a$. Therefore

$$\int\limits_{N^{-1}(a)} \|Sf - \varphi(N(f))\| d\nu(f|a) = \frac{1}{2} \int\limits_{N^{-1}(a)} \sum_{\beta = \pm 1} \|S(\beta \tilde{f}_a + \bar{f}_a) - \varphi(N(\bar{f}_a))\| d\nu(f|a)$$

$$\geq \frac{1}{2} \int\limits_{N^{-1}(a)} \|S(\tilde{f}_a + \bar{f}_a) - S(-\tilde{f}_a + \bar{f}_a)\| d\nu(f|a)$$

$$= \int\limits_{N^{-1}(a)} \|S(\tilde{f}_a)\| d\nu(f|a).$$

Hence

$$e^{avg}(S, N, \varphi, \nu) \geq \int\limits_{N(X_1)} \int\limits_{N^{-1}(a)} \|S(\tilde{f}_a)\| d\nu(f|a) d\tilde{\nu}(a).$$

Let now $\widehat{X_1}$ be defined by

$$\widehat{X_1} := \{f \in X_1 \,:\, \mathrm{card}(N(f)) \leq 2n\}.$$

As we deal only with those N with $\mathrm{card}^{avg}(N, \nu) \leq n$,

$$n \geq \int\limits_{X_1 \setminus \widehat{X_1}} \mathrm{card}(N(f)) \, d\nu(f) \geq 2\nu(X_1 \setminus \widehat{X_1})n.$$

Therefore $\nu(\widehat{X}_1) \geq 1/2$. From this we deduce

$$e^{avg}(S, N, \varphi, \nu) \geq \frac{1}{2} \inf_{a \in N(\widehat{X}_1)} \int_{N^{-1}(a)} \|S(\tilde{f}_a)\| d\nu(f|a). \tag{52}$$

By the construction of $\tilde{f}_a$, we have

$$\int_{N^{-1}(a)} \|S(\tilde{f}_a)\| d\nu(f|a) = \mathbf{E} \Big\| S\Big(\sum_{(i,j) \in K_a} \varepsilon_{ij} \psi_{ij} \Big) \Big\|.$$

Combining this with (52) and Lemma 5.2, we get

$$e^{avg}(S, N, \varphi, \nu) \geq \frac{1}{2} |\gamma| \gamma_1 \min (m_1^{-r}, m_2^{-r}) \, m_2^{-d_2} \min_{a \in N(\widehat{X}_1)} \mathbf{E} \Big(\max_{i \in I} \big| \sum_{j \in K_{a,i}} \varepsilon_{ij} \big| \Big), \tag{53}$$

with

$$K_{a,i} := \{j \in J \,:\, (i,j) \in K_a\}.$$

Observe that for $a \in N(\widehat{X}_1)$, $\mathrm{card}(a) \leq 2n$, and hence, by the construction of K_a in (51),

$$|K_a| \geq m_1^{d_1} m_2^{d_2} - 2n. \tag{54}$$

Let

$$I_a := \{i \in I \,:\, |K_{a,i}| \geq m_2^{d_2}/4\}. \tag{55}$$

Then

$$|I_a| \geq m_1^{d_1}/4. \tag{56}$$

In fact, assuming the opposite, i.e. $|I_a| < m_1^{d_1}/4$, we have

$$|K_a| = \sum_{i \in I} |K_{a,i}| = \sum_{i \in I_a} |K_{a,i}| + \sum_{i \notin I_a} |K_{a,i}| < \frac{1}{2} m_1^{d_1} m_2^{d_2}.$$

But this is a contradiction to (54) and (48), which proves (56). Now we set

$$n_1 = \big\lceil m_1^{d_1}/4 \big\rceil, \; n_2 = \big\lceil m_2^{d_2}/4 \big\rceil.$$

An elementary direct argument or the use of the contraction principle (see Theorem 4.4 of Ledoux and Talagrand, 1991) yields

$$\mathbf{E} \max_{i \in I} \big| \sum_{j \in K_{a,i}} \varepsilon_{ij} \big| \geq \mathbf{E} \max_{1 \leq \tilde{i} \leq n_1} \big| \sum_{\tilde{j}=1}^{n_2} \varepsilon_{\tilde{i}\tilde{j}} \big|, \tag{57}$$

where $\varepsilon_{\tilde{i}\tilde{j}}$ $(\tilde{i} = 1, \ldots, n_1, \tilde{j} = 1, \ldots, n_2)$ are again independent symmetric Bernoulli random variables. Since N was arbitrary, (53), (57) and Lemma 5.3 imply

$$e_n^{avg}(S, \nu) \succeq \min (m_1^{-r}, m_2^{-r}) m_2^{-d_2} \big(n_2 \min(n_2, \log(n_1 + 1)) \big)^{1/2}$$

$$\asymp \min (m_1^{-r}, m_2^{-r}) m_2^{-d_2/2} \min \big(m_2^{d_2}, \log(m_1 + 1) \big)^{1/2}. \tag{58}$$

We treat the cases $r \geq d_1/2$ and $r < d_1/2$ separately.

1. Case $r \geq d_1/2$:

   We choose $m_1 = m_2 = 2\lceil n^{1/(d_1+d_2)} \rceil$. Then (48) is satisfied. Moreover

   $$\min\left(m_2^{d_2}, \log(m_1+1)\right) \asymp \log n,$$

   which gives together with (58)

   $$e_n^{avg}(S, \nu) \succeq n^{-r/(d_1+d_2)-d_2/(2(d_1+d_2))}\left(\log n\right)^{1/2},$$

   completing the proof in this case.

2. Case $r < d_1/2$:

   Here we set $m_1 = 2\lceil n^{1/d_1}\left(\log n\right)^{-1/d_1}\rceil$ and $m_2 = 2\lceil\left(\log n\right)^{1/d_2}\rceil$. Again (48) is satisfied. Furthermore,

   $$\min\left(m_1^{-r}, m_2^{-r}\right) \asymp n^{-r/d_1}\left(\log n\right)^{r/d_1}$$

   and

   $$\min\left(m_2^{d_2}, \log(m_1+1)\right) \asymp \log n.$$

   Combining this with (58), we obtain

   $$e_n^{avg}(S, \nu) \succeq n^{-r/d_1}\left(\log n\right)^{r/d_1},$$

   ending the proof of Proposition 5.1 and of Theorem 2.4. ∎

6 COMMENTS

In the sequel we want to illustrate the Monte Carlo rates obtained, their relation to the parameter constellation and to the deterministic case. Let us first mention the following view on parametric integration: Let $d := d_1 + d_2$. Then we want to integrate functions on the d-dimensional unit cube partially, that means, we integrate over d_2 dimensions and leave d_1 dimensions "untouched". Let us now try to analyze the dependence on d_1.

Clearly the deterministic exponent of decay, r/d, is independent of d_1. The Monte Carlo exponent for $d_1 = 0$ is $r/d + 1/2$, the well-known rate for integration, while for $d_1 = d$ it is r/d, the rate of pure approximation. Now let us consider the Monte Carlo exponent $\alpha(d_1)$ of the intermediate situation (we neglect logarithms). Let us introduce

$$\alpha_1(d_1) = \frac{r}{d} + \frac{d - d_1}{2d} \tag{59}$$

and

$$\alpha_2(d_1) = \frac{r}{d_1}. \tag{60}$$

18

For all d_1 the following relation between (59) and (60) is readily checked

$$\alpha_1(d_1) = \alpha_2(d_1) + \frac{(d - d_1)(d_1 - 2r)}{2dd_1}. \tag{61}$$

By Theorem 2.4 we have $\alpha(d_1) = \alpha_1(d_1)$ for $d_1 \leq 2r$ and $\alpha(d_1) = \alpha_2(d_1)$ for $d_1 > 2r$. Taking into account (61), this means that, in fact, $\alpha(d_1) = \min(\alpha_1(d_1), \alpha_2(d_1))$. The behavior is illustrated in Figure 1.

Figure 1: Exponent of convergence as a function of d_1

| .... | deterministic rate | − − | $\alpha_1(d_1)$ |
| ——— | Monte Carlo rate | −·— | $\alpha_2(d_1)$ |
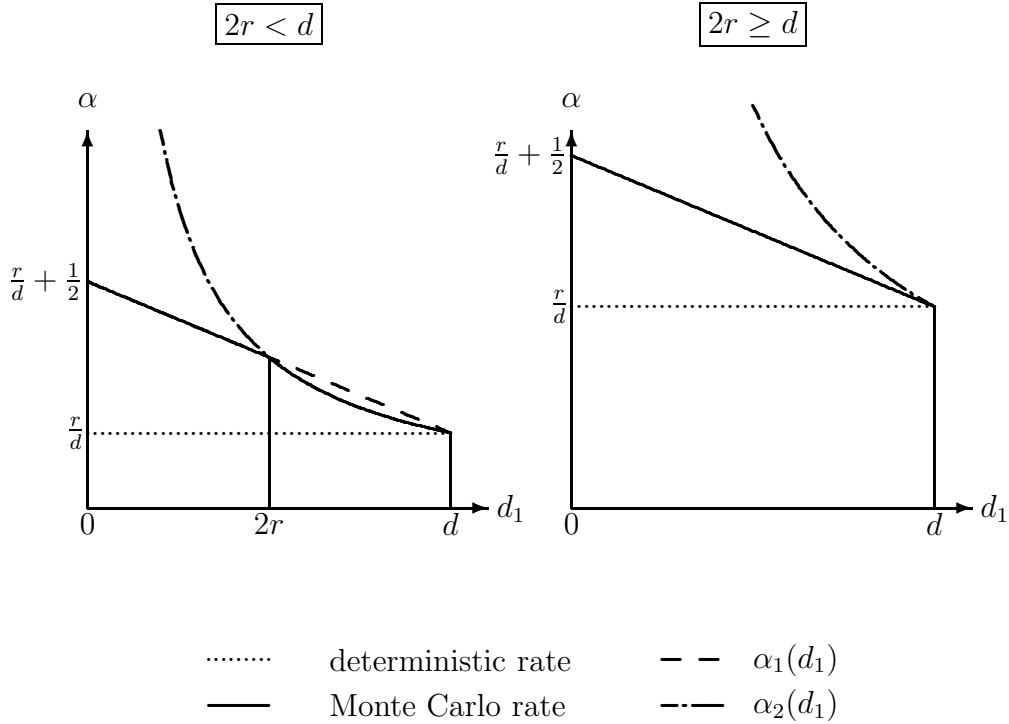
Note that the rate r/d_1 (for $d_1 > 2r$), in other words, the error n^{-r/d_1}, is the well-known one of approximation of r-smooth functions in d_1 dimensions using n points. However, our task is much more complicated – before we approximate the solution function using n node points we first have to determine the values in these points, which amounts to the approximate computation of n integrals over the d_2-dimensional cube. So the result is far from being trivial. It shows that (in the Monte Carlo setting) integration over the remaining variables can be done at the same cost as just d_1-dimensional approximation (up to a loss of a logarithmic factor).

References

[1] BAKHVALOV, N. S. (1959), On approximate calculation of integrals, *Vestnik Moskow. Gos. Univ., Ser. Mat. Mekh. Astronom, Fiz. Khim* **4**, 3–18 (in Russian).

[2] BAKHVALOV, N. S. (1961), An estimate of the main remainder in quadrature formulas, *Zh. Vychisl. Mat. Mat. Fiz.* **1**(1961), 64–77 (in Russian) [*English transl.: U.S.S.R. Comput. Math. and Math. Phys.* **1** (1961), 68– 82].

[3] BÜCKNER, H. (1950), Bemerkungen zur numerischen Quadratur, *Math. Nachr.* **3**, 142–151.

[4] ERMAKOV, S. M. and MIKHAILOV, G. A. (1982), "Statistical Modelling", Nauka, Moskow, 1982 (in Russian).

[5] FROLOV, A. S. and CHENTSOV, N. N. (1962), On the calculation of definite integrals dependent on a parameter by the Monte Carlo method, *Zh. Vychisl. Mat. Fiz.*, **2**(4), 714–717 (in Russian) [*English transl.: U.S.S.R. Comput. Math. and Math. Phys.* **2**(2), 714–717].

[6] HEINRICH, S. (1994), Random approximation in numerical analysis, *Functional Analysis (K.D. Bierstedt, A. Pietsch, W. M. Ruess, and D. Vogt, eds.)*, Marcel Dekker, pp. 123 – 171.

[7] HEINRICH, S. (1998a), Monte Carlo Complexity of Global Solution of Integral Equations, *J. Complexity* **14**, 151 – 175.

[8] HEINRICH, S. (1998b), A multilevel version of the method of dependent tests, *Proceedings of the 3rd St. Petersburg Workshop on Simulation (S. M. Ermakov, Y. N. Kashtanov, and V. B. Melas, editors)*, St. Petersburg University Press, pp. 31 – 35.

[9] LEDOUX, M. and TALAGRAND, M. (1991), "Probability in Banach Spaces", Springer, Berlin–Heidelberg–New York.

[10] MIKHAILOV, G. A., "Minimization of Computational Costs of Non–Analogue Monte Carlo Methods". World Scientific, Singapore, 1991.

[11] NOVAK, E. (1988), "Deterministic and Stochastic Error Bounds in Numerical Analysis", Lecture Notes in Mathematics 1349, Springer, Berlin, Heidelberg, New York, London, Paris and Tokio.

[12] NIKOLSKIJ, S. M. (1950), On the problem of approximation estimate by quadrature formulas, *Usp. Mat. Nauk.* **5**, 165–177 (in Russian).

[13] PRIGARIN, S. M. (1995), Convergence and optimization of functional estimates in statistical modelling in Sobolev's Hilbert spaces. *Russian J. Numer. Math. Modelling*, **10**(4), 325–346.

[14] SARD, A. (1949), Best approximate integration formulas; best approximation formulas, *American J. Math.* **71**, 80–91.

[15] SOBOL, I. M. (1962), The use of ω^2–distribution for error estimation in the calculation of integrals by the Monte Carlo method, *Zh. Vychisl. Mat. Fiz.*, **2**(4), 717–723 (in Russian) [*English transl.: U.S.S.R. Comput. Math. and Math. Phys..* **2**(2), 717–723].

[16] SOBOL, I. M. (1973), "Computational Monte Carlo Methods", Nauka, Moscow, 1973 (in Russian).

[17] TRAUB, J. F., WASILKOWSKI, G. W. and WOŹNIAKOWSKI, H. (1988), "Information–based Complexity", Academic Press, New York.

[18] VOYTISHEK, A. V. (1996), Discrete–stochastic procedures for the global estimation of an integral which depends on a parameter, *Comp. Maths Math. Physics* **36** (8), 997–1009.