Monte Carlo Complexity of Global Solution of Integral Equations

S. Heinrich
Fachbereich Informatik
Universität Kaiserslautern
Postfach 3049
D-67653 Kaiserslautern
Germany

Abstract

The problem of global solution of Fredholm integral equations is studied. This means that one seeks to approximate the full solution function (as opposed to the local problem, where only the value of the solution in a single point or a functional of the solution is sought). The Monte Carlo complexity is analyzed, i. e. the complexity of stochastic solution of this problem. The framework for this analysis is provided by information-based complexity theory. The investigations complement previous ones on stochastic complexity of local solution and on deterministic complexity of both local and global solution. The results show that even in the global case Monte Carlo algorithms can perform better than deterministic ones, although the difference is not as large as in the local case.

1   Introduction

Monte Carlo methods are a classical tool of solving high dimensional integral equations. Basic applications include neutron transport (Spanier and Gelbard, 1969, Ermakov, 1971) and thermal radiation (Siegel and Howell, 1992). Here we consider this problem from a complexity theoretic point of view. We investigate the minimal possible error among all methods of given cost. Such an analysis helps to understand the potential power of algorithms of a given class (e. g. the class of all deterministic or all randomized algorithms), and hence allows to compare different classes. In particular, it may show for a concrete numerical problem if randomization can bring advantages over deterministic algorithms or not. A framework, notions and methods for such an analysis are provided by information-based complexity theory (see Traub, Wasilkowski, and Woźniakowski, 1988).

In this paper we study the problem of numerical solution of Fredholm integral equations

$$u(s) - \int_G k(s,t)u(t)dt = f(s),$$

with given continuous functions k on G^2 and f on G (the details will be given below). Two subproblems can be distinguished. In the first case we seek to approximate the full solution function u in some way (e. g. by solving on a grid and interpolating or by finite element approximation etc.). We call this the problem of global solution. In the second case we want to approximate the value $u(t_0)$ of the solution in a single point t_0 or the value of a functional of u, e. g. the integral over G. This is called the local solution problem.

While deterministic numerical methods such as Nyström, collocation, FEM usually aim at solving the global problem, the classical Monte Carlo approach is directed to the local solution. Monte Carlo methods are well understood in this situation and are generally acknowledged to bring advantages (at least for high dimensional problems) over the deterministic approaches. The efficiency of randomized methods for the global problem is much less understood. Recent work in this direction is due to Mikhailov (1991a, 1991b, 1995), and Voytishek (1994, 1996), who proposed and analyzed various algorithms.

The present paper is devoted to the complexity theoretic analysis of randomized solution of the global problem. We consider the global problem for the model class of smooth kernels and right hand sides and determine the optimal convergence rates (and thus the complexity). The result complements previous research of Emelyanov and Ilin (1967) on the deterministic complexity of the local and global problem and of Heinrich and Mathé (1993) on the Monte Carlo complexity of the local problem. Our result shows that - as in the local case - randomized methods are superior also for the global problem, but the difference between the optimal stochastic and deterministic convergence rate is smaller than in the local case.

The proof of the lower bound is based on a suitable average case approach, while the upper bound is shown by presenting and analyzing a concrete algorithm. This algorithm is new and different from the classical ones, and also from those proposed in Mikhailov (1991a, 1991b, 1995), Voytishek (1994, 1996) and Heinrich and Mathé (1993), which would not reach the optimal rate. The algorithm starts with the variance reduction technique of Heinrich and Mathé (1993), to provide an approximation on a rough grid. To meet the optimal rate this is, however, not sufficient. Therefore, a new technique is developed - a multigrid Monte Carlo procedure, which provides updates of the solution on successively finer grids. Although the algorithm is tuned to the model class, it is of interest and a topic of future research to extend these ideas to a broader range of problems.

For all material concerning information-based complexity theory, we refer to Traub, Wasilkowski, and Woźniakowski (1988), Novak (1988), Heinrich (1994, 1996), Mathé (1994).

The following section 2 contains basic notions, previous development and the statement of the theorem. The upper bound of the theorem is proved in

section 3, the lower bound in section 4.

2 Notation and formulation of the result

We study the following numerical problem: Approximate the solution u of a Fredholm integral equation of the second kind

$$u(s) - \int_G k(s,t)u(t)dt = f(s). \tag{1}$$

Here $G = [0,1]^d$ denotes the d-dimensional unit cube. We consider this equation in the space $C(G)$ of continuous functions on G, endowed with the supremum norm. This is the standard norm we shall be working with, therefore we denote it simply by $\| \ \|$, while all other norms will be distinguished by subscripts. The given data of the problem, the functions k and f are assumed to belong to sets of functions of a certain smoothness. To introduce them, let $r \in \mathbb{N}$ ($\mathbb{N}$ will always denote the set of positive integers, and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$). Let $C^r(G)$ be the space of r-times continuously differentiable functions endowed with the norm

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

where $\alpha = (\alpha_1, \ldots, \alpha_d)$ is a multi-index and $|\alpha| = \alpha_1 + \ldots + \alpha_d$. Correspondingly we define $C^r(G^2)$ and put $X = C^r(G^2) \oplus C^r(G)$ (the direct sum, endowed with the maximum norm). Now we fix parameters $\kappa_1, \kappa_3 > 0$, $0 < \kappa_2 < 1$ (which will remain fixed throughout the paper), and define

$$\begin{aligned}
\mathcal{K} &= \{k \in C^r(G^2) \quad : \quad \|k\|_r \leq \kappa_1, \ \|k\| \leq \kappa_2\}, \\
\mathcal{F} &= \{f \in C^r(G) \quad : \quad \|f\|_r \leq \kappa_3\}
\end{aligned}$$

and

$$X_0 = \mathcal{K} \times \mathcal{F} \subseteq X.$$

The solution operator $S : X_0 \to C(G)$ (mapping the data onto the exact solution of the problem) is defined as

$$S(k,f) = u = (Id - T_k)^{-1}f.$$

Here Id denotes the identity operator on $C(G)$ and T_k stands for the integral operator acting in $C(G)$ as

$$(T_k g)(s) = \int_G k(s,t)g(t)dt$$

for $g \in C(G)$.

We shall study only this solution operator S, which we call the solution operator of the global problem, meaning that we seek to approximate the full solution u of (1) as a function on G. This should be contrasted with the local solution operator $S_\chi : X_0 \to \mathbb{R}$ studied in Heinrich and Mathé (1993), which is defined as

$$S_\chi(k, f) = \left\langle (Id - T_k)^{-1} f, \chi \right\rangle,$$

where χ is any fixed continuous linear functional on $C(G)$. In this case, for example, only the value of the solution in a fixed point or a certain weighted mean of the solution is sought.

Next we describe the basic setting of information-based complexity theory (Traub, Wasilkowski, and Woźniakowski, 1988) for this problem. We shall fix the following class Λ of information functionals (Λ will be a subset of the dual space X^*):

$$
\begin{aligned}
\Lambda \;\; = \;\; & \left\{ \delta_{(s,t)}^\alpha : \alpha \in \mathbb{N}_0^{2d}, \; |\alpha| \le r, \; (s,t) \in G^2 \right\} \\
& \cup \left\{ \delta_t^\beta : \beta \in \mathbb{N}_0^d, \; |\beta| \le r, \; t \in G \right\},
\end{aligned}
$$

where

$$\delta_{(s,t)}^\alpha(k, f) = D^\alpha k(s, t)$$

and

$$\delta_t^\beta(k, f) = D^\beta f(t)$$

are the corresponding to α and β partial derivatives, taken in the point (s, t) and t, respectively. Hence, we admit values of functions and their derivatives as information. A crucial role in complexity theory is played by the information operator. It is given by two arbitrary sequences of functions $(L_i)_{i \in \mathbb{N}}$ and $(\mathrm{ter}_i)_{i \in \mathbb{N}}$,

$$
\begin{aligned}
L_1 \;\; &: \;\; X \to \mathbb{R} \\
L_i \;\; &: \;\; X \times \mathbb{R}^{i-1} \to \mathbb{R} \quad (i > 1)
\end{aligned}
$$

such that $L_1 \in \Lambda$,

$$L_i(\,\cdot\,, a_1, \ldots, a_{i-1}) \in \Lambda$$

4

for all $(a_1, \ldots, a_{i-1}) \in \mathbb{R}^{i-1}$ and $i > 1$, and

$$\mathrm{ter}_i : \mathbb{R}^i \to \{0, 1\} \quad (i \in \mathbb{N}).$$

Put $\mathbb{R}^\infty = \cup_{i=1}^\infty \mathbb{R}^i$ (the disjoint union). Then define $N : X_0 \to \mathbb{R}^\infty$ as follows: Given $x \in X_0$, let $a_1 = L_1(x)$,

$$a_i = L_i(x, a_1, \ldots, a_{i-1}) \quad (i > 1)$$

and let $n = n(x)$ be the smallest $n \in \mathbb{N}$ with $\mathrm{ter}_n(a_1, \ldots, a_n) = 1$. We shall assume that $n(x) < \infty$ for each $x \in X_0$. Now we define $N : X_0 \to \mathbb{R}^\infty$ by setting

$$N(x) = (a_1, \ldots, a_{n(x)}).$$

The structure of N reflects the process of collecting information (e. g. calling a subroutine) during the computation. As defined above, we consider adaptive information of varying cardinality. For details and background we refer to Traub, Wasilkowski, and Woźniakowski (1988). It is understood that deterministic approximations to S will be sought in the form $\varphi \circ N$, where N is an information operator as above, and φ is an arbitrary mapping $\varphi : \mathbb{R}^\infty \to C(G)$ (representing the computations of the algorithm carried out on the set of obtained information values).

Let $\mathcal{N}$ denote the set of all information operators of the above type and Φ the set of all mappings from $\mathbb{R}^\infty$ to $C(G)$. Given $N \in \mathcal{N}$, $x \in X_0$, we denote

$$\mathrm{card}(N(x)) = n(x),$$

where $n(x)$ is as defined above.

In this paper we are concerned with the randomized complexity, so deterministic approximations will only serve as building blocks of the algorithms. Put in an abstract setting, this looks as follows: An abstract Monte Carlo method

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

consists of a probability space (Ω, Σ, μ), with Ω a nonempty set, Σ a σ-algebra of subsets of Ω, and μ a (σ-additive) probability measure on Σ, and a family $(N_\omega, \varphi_\omega) \in \mathcal{N} \times \Phi$ ($\omega \in \Omega$), such that for each $x \in X_0$, $\varphi_\omega(N_\omega(x))$ and $\mathrm{card}(N_\omega(x))$ are Σ measurable functions of ω (the former one as a function into $C(G)$, endowed with the σ-algebra of Borel sets). Let $\mathcal{M}$ be the class of all such abstract Monte Carlo methods. For background and motivation of this approach see Traub, Wasilkowski, and Woźniakowski (1988), and Heinrich (1994, 1996). Given $M \in \mathcal{M}$, the cardinality of M is defined by

5

$$\operatorname{card}(M) = \sup_{x \in X_0} \int_\Omega \operatorname{card}(N_\omega(x)) \, d\mu(\omega)$$

(we admit $+\infty$ as a possible value of $\operatorname{card}(M)$). The error of M as a randomized approximation to S is given by

$$e(S, M) = \sup_{x \in X_0} \int_\Omega \|S(x) - \varphi_\omega(N_\omega(x)\| \, d\mu(\omega).$$

The minimal error of all Monte Carlo methods of cardinality not exceeding n is defined for $n \in \mathbb{N}$ as

$$e_n^{\mathrm{MC}}(S) = \inf\{e(S, M) : M \in \mathcal{M}, \ \operatorname{card}(M) \leq n\}.$$

This is the crucial quantity of information-based complexity. No randomized method that uses (on the average) at most n information functionals can provide a smaller error than $e_n^{\mathrm{MC}}(S)$. This is the advantage of the generality of the approach: Consider a concrete model of randomized computation over the reals, e. g. the one in Heinrich (1996). Then algorithms on the basis of such a model are easily seen to be a special case of the abstract notion. Hence lower bounds proved for e_n^{MC} lead directly to lower bounds for this model (and certainly also for a variety of other, similar models). But what about upper bounds? Once e_n^{MC} is determined, the definition says that there are abstract methods which reach this error. This may be too little for a concrete model of computation. However, in many situations, including ours here, it is possible to construct special such algorithms, which do not only meet the abstract criteria, but which are fully implementable, with a number of arithmetic operations proportional to the number of information functionals. Hence in these cases the order of the complexity also in the sense of the above model or in the naive, arithmetic sense is completely controlled by the numbers e_n^{MC}.

It is the goal of this paper to determine e_n^{MC} for the problem of full solution of Fredholm integral equations. Before we state the main result, let us recall previous results for the sake of comparison. First we mention the deterministic setting, which was investigated by Emelyanov and Ilin (1967). In the statement, e_n stands for the minimal deterministic error, see Traub, Wasilkowski, and Woźniakowski (1988), Novak (1988). (The deterministic minimal error can be defined as e_n^{MC}, with the difference, that only methods with trivial probability spaces $\Omega = \{\omega_0\}$ are admitted). Furthermore, we use the following notation: $a_n \prec b_n$ means that there are constants $c > 0$ and $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, $a_n \leq cb_n$. We write $a_n \asymp b_n$ if $a_n \prec b_n$ and $b_n \prec a_n$. Concerning the constants appearing throughout this paper let us mention that we often use the same symbol for possibly different constants.

Theorem 1 (Emelyanov and Ilin, 1967)

$$e_n(S) \asymp \sup_{\substack{\chi \in C(G)^* \\ \|\chi\| \leq 1}} e_n(S_\chi) \asymp n^{-r/(2d)}.$$

So in the deterministic setting the global and the local problem are of the same complexity (up to constants). This is no longer the case in the randomized setting, as we shall see below. In this setting the local problem was solved as follows.

Theorem 2 (Heinrich and Mathé, 1993)

$$\sup_{\substack{\chi \in C(G)^* \\ \|\chi\| \leq 1}} e_n^{\mathrm{MC}}(S_\chi) \asymp n^{-r/(2d)-1/2}.$$

Note that both theorems were proved for the case of non-adaptive information, but the proofs can easily be extended to the case of adaptive information. Also, only function values were considered there, but the generalization to values of functions and derivatives is immediate.

In this paper we solve the problem of Monte Carlo complexity of global solution and prove the following

Theorem 3

(i) If $r > d/2$, then $e_n^{\mathrm{MC}}(S) \asymp n^{-r/(2d)-1/4}(\log n)^{1/2}$.

(ii) If $r < d/2$, then $e_n^{\mathrm{MC}}(S) \asymp n^{-r/d}(\log n)^{r/d}$.

(iii) If $r = d/2$, then $n^{-1/2}(\log n)^{1/2} \prec e_n^{\mathrm{MC}}(S) \prec n^{-1/2}(\log n)^{3/2}$.

3 An optimal algorithm and the upper bound

Here we shall develop a Monte Carlo algorithm and show that it reaches the rate required by the theorem, thus proving the upper bound. We need some further notation. We set for $\ell \in \mathbb{N}_0$

$$I_\ell = \left\{ (i_1, \ldots, i_d) : i_1, \ldots, i_d \in \left\{ 0, \ldots, 2^\ell - 1 \right\} \right\}.$$

For each $i \in I_\ell$ we define mappings in $\mathbb{R}^d$

$$\begin{aligned} \sigma_{\ell i}(s) &= 2^{-\ell}(i + s) \\ \tau_{\ell i}(s) &= 2^\ell s - i \end{aligned} \qquad (s \in \mathbb{R}^d).$$

So $\sigma_{\ell i}$ shrinks G to the subcube

$$G_{\ell i} = \sigma_{\ell i}(G)$$

and $\tau_{\ell i} : G_{\ell i} \rightarrow G$ extends $G_{\ell i}$ to G. Let π_ℓ be the partition of G into these subcubes (of sidelength $2^{-\ell}$), i. e.

$$\pi_\ell = \{G_{\ell i} : i \in I_\ell\}.$$

For a function f on $\mathbb{R}^d$ define the contraction and the extension operator by

$$\begin{aligned}(C_{\ell i}f)(s) &= f(\tau_{\ell i}(s)) \\ (E_{\ell i}f)(s) &= f(\sigma_{\ell i}(s)).\end{aligned}$$

Let $\mathcal{P}^r(\pi_\ell)$ be the space of continuous piecewise polynomial functions of degree $\leq r$ on the partition π_ℓ, i. e. $f \in \mathcal{P}^r(\pi_\ell)$ iff $f \in C(G)$ and $f|_{G_{\ell i}}$ is a polynomial of (maximum) degree $\leq r$ for all $i \in I_\ell$. We shall use the following standard interpolation operators: Let

$$\Gamma_\ell = \left\{r^{-1}2^{-\ell}(i_1, \ldots, i_d) :\ 0 \leq i_1, \ldots, i_d \leq r2^\ell\right\}$$

be the uniform grid of sidelength $r^{-1}2^{-\ell}$ on G. Let P_0 be the d-dimensional Lagrange interpolation on Γ_0 (that is, the tensor product of one-dimensional Lagrange interpolation operators of order r). Let P_ℓ be the composition of applying P_0 to each subcube of the partition π_ℓ, i. e.

$$P_\ell = \sum_{i \in I_\ell} C_{\ell i} P_0 E_{\ell i}.$$

We understand P_ℓ as an operator from $\ell_\infty(\Gamma_\ell)$ to $\mathcal{P}^r(\pi_\ell) \subset C(G)$. When we write $P_\ell g$ for $g \in C(G)$, we mean $P_\ell\left((g(t))_{t \in \Gamma_\ell}\right)$. It is well-known that

$$\|P_\ell : \ell_\infty(\Gamma_\ell) \rightarrow C(G)\| \leq c \tag{2}$$

(this notation means the operator norm), and for $f \in C^r(G)$

$$\|f - P_\ell f\| \leq c2^{-r\ell}\|f\|_r. \tag{3}$$

Let ρ be a Z-valued random variable, where Z is a Banach space. We define the second moment as $\mathbb{E}\|\rho\|_Z^2$ and, in case that this is finite,

$$\mathrm{Var}(\rho) = \mathbb{E}\|\rho - \mathbb{E}\rho\|_Z^2.$$

8

To state the next result let us first recall that the type 2 constant $T_2(Z)$ of a Banach space Z is the smallest c with $0 < c \leq +\infty$, such that for all n and all sequences $(z_i)_{i=1}^n \subset Z$,

$$\mathbb{E}\|\sum_{i=1}^n \epsilon_i z_i\|^2 \leq c^2 \sum_{i=1}^n \|z_i\|^2,$$

where (ϵ_i) denotes a sequence of independent Bernoulli random variables on some probability space (Ω, Σ, μ), i. e. $\mu\{\epsilon_i = 1\} = \mu\{\epsilon_i = -1\} = \frac{1}{2}$ (see, e. g. Ledoux and Talagrand, 1991, ch. 9.2, for this definition).

Lemma 1 *a) For each sequence $(\rho_i)_{i=1}^m$ of independent Z-valued random variables of finite second moment*

$$\mathrm{Var}\left(\sum_{i=1}^m \rho_i\right) \leq (2T_2(Z))^2 \sum_{i=1}^m \mathrm{Var}(\rho_i).$$

b) There is a constant $c > 0$ such that for each n,

$$T_2(\ell_\infty^n) \leq c(\log n)^{1/2}.$$

Both results are well-known in Banach space theory. Part a) of Lemma 1 is Proposition 9.11 of Ledoux and Talagrand (1991), part b) is stated in Tomczak-Jaegermann (1988), p. 16. A proof can be found in Linde and Pietsch (1974), Lemma 6, or can be obtained directly by combining relations (3.13), (4.8) and Theorem 4.7 from Ledoux and Talagrand, 1991.

Now we can describe the algorithm. It consists of a deterministic and a stochastic part. Fix $n \in \mathbb{N}$ and let $m \in \mathbb{N}$ be such that

$$2^{2d(m-1)} \leq n < 2^{2dm}. \tag{4}$$

Deterministic part

First we use an algorithm of approximate deterministic solution of equation (1) with the following property. For each $(k, f) \in X_0$ the algorithm provides an approximation $v_0 \in \ell_\infty(\Gamma_m)$ to the true solution $(u(s))_{s \in \Gamma_m}$ on the grid Γ_m. This is a rough approximation, which will be used to precondition the finer Monte Carlo approximation.

We define

$$\begin{array}{rcl} v & = & P_m v_0 \quad\quad \in \quad C(G) \\ h & = & (P_m \otimes P_m)k \quad \in \quad C(G^2) \end{array}$$

where the tensor product has the canonical meaning of applying P_m with respect to both s and t to $k(s,t)$. Finally, we set

$$g = v - T_h v.$$

Since v is piecewise polynomial on π_m and h is such on π_m^2, the computation of g can be accomplished explicitly. Observe that we have achieved that v is an exact solution of the integral equation with kernel h and right hand side g. This is the key ingredient of the separation of main part in the Monte Carlo scheme below.

Stochastic part

Fix θ with $\kappa_2 < \theta < 1$ and consider a stationary absorbing Markov chain on G with density of initial distribution $p_0(s) \equiv 1$ and density of transition probability $p(s,t) \equiv \theta$. (In other words, the initial and consecutive states are distributed uniformly on G and absorption occurs with probability $1 - \theta$.) We assume that the Markov chain together with a countable number of independent copies is defined on some basic probability space (Ω, Σ, μ). Almost all realizations of the Markov chain are of finite length. Let

$$\xi = (t_0, \ldots, t_q) \tag{5}$$

be such a realization. First we define a random variable $\eta_m(s, \xi)$ for $s \in \Gamma_m$ by setting

$$\eta_m(s, \xi) = (1 - \theta)^{-1} \theta^{-q} [k(s, t_q) k(t_q, t_{q-1}) \ldots k(t_1, t_0) f(t_0) \\ - h(s, t_q) h(t_q, t_{q-1}) \ldots h(t_1, t_0) g(t_0)]. \tag{6}$$

This is the absorption estimate in the von Neumann - Ulam scheme (see, e. g., Ermakov, 1971), with the variance reduction by separation of the main part as suggested in Heinrich and Mathé (1993). We define a vector-valued random variable by setting

$$\eta_m(\xi) = (\eta_m(s, \xi))_{s \in \Gamma_m} \in \ell_\infty(\Gamma_m).$$

This exhausts the approach of Heinrich and Mathé (1993) - but is not yet sufficient for our purposes. We need approximations of the solution on grids finer than Γ_m. This will be accomplished by a multilevel updating procedure. For this purpose we choose the final level $m^* > m$.

Fix ℓ with $m < \ell \le m^*$, and set

$$h_\ell(\,\cdot\,, t) = k(\,\cdot\,, t) - P_{\ell-1} k(\,\cdot\,, t), \tag{7}$$

10

and for $s \in \Gamma_\ell$,

$$\eta_\ell(s, \xi) = (1 - \theta)^{-1} \theta^{-q} h_\ell(s, t_q) k(t_q, t_{q-1}) \ldots k(t_1, t_0) f(t_0). \tag{8}$$

Now put

$$\eta_\ell(\xi) = (\eta_\ell(s, \xi))_{s \in \Gamma_\ell} \in \ell_\infty(\Gamma_\ell).$$

Next choose a natural number q_ℓ for each ℓ with $m \leq \ell \leq m^*$. Let $\xi_{i\ell}$ ($i = 1, \cdots, q_\ell$, $\ell = m, \ldots, m^*$) be independent realizations of the Markov chain. For each ℓ, we set

$$\zeta_\ell = \frac{1}{q_\ell} \sum_{i=1}^{q_\ell} \eta_\ell(\xi_{i\ell}). \tag{9}$$

Observe that ζ_ℓ is an $\ell_\infty(\Gamma_\ell)$ valued random variable. The final approximation to u is computed from the variables above by interpolation:

$$\zeta = v - g + P_{m^*} f + \sum_{\ell=m}^{m^*} P_\ell \zeta_\ell. \tag{10}$$

Hence ζ is a piecewise polynomial function from $\mathcal{P}^r(\pi_{m^*})$, the coefficients of the polynomial pieces being random. This accomplishes the description of the algorithm.

In the sequel we analyze the algorithm and specify the so far undefined parameters so as to suit the smoothness class under consideration. In the deterministic part we choose the approximation in such a way that

$$\|u|_{\Gamma_m} - v_0\|_{\ell_\infty(\Gamma_m)} \leq c_2 2^{-rm}, \tag{11}$$

where $c_1, c_2 > 0$ do not depend on k, f and m.

In a way analogous to (3) we have

$$\|k - h\| \leq c 2^{-rm}. \tag{12}$$

Relations (2), (3), (11), and (12) imply

$$\|f - g\| \leq c 2^{-rm}. \tag{13}$$

Fix a κ_2' with $\kappa_2 < \kappa_2' < \theta$. By (12), there is an m_0 such that for $m \geq m_0$ and for all $k \in \mathcal{K}$ the resulting h satisfies $\|h\| \leq \kappa_2'$. Then it can be checked as in Heinrich and Mathé (1993), that

$$\mathbb{E}\eta_m(s, \xi) = (T_k u)(s) - (T_h v)(s) = (T_k u)(s) - v(s) + g(s) \tag{14}$$

and, using (12) and (13),

$$|\eta_m(s, \xi)| \leq c2^{-rm} \tag{15}$$

(the constant being independent of m, k, f, ξ, and s).

In a similar way one verifies that for all $\ell > m$,

$$\mathbb{E}\eta_\ell(s, \xi) = \int_G h_\ell(s, t) u(t) \, dt = (T_k u)(s) - (P_{\ell-1} T_k u)(s), \tag{16}$$

and, in view of (7) and (3)

$$|\eta_\ell(s, \xi)| \leq c2^{-r\ell}. \tag{17}$$

We define

$$m^* = \lceil m(1 + d/(2r) \rceil. \tag{18}$$

if $r \geq d/2$ and

$$m^* = 2m - p, \tag{19}$$

$$p = [(\log_2 m)/d] \tag{20}$$

if $r < d/2$. (Here $\lceil a \rceil$ and $[a]$ have the usual meaning of the smallest integer $\geq a$ and the largest integer $\leq a$, respectively.) Note that $m < m^* \leq 2m$.

For $\ell = m, m + 1, \ldots, m^*$ we set

$$q_\ell = \lceil 2^{dm - (r + d/2)(\ell - m)} \rceil \tag{21}$$

if $r \geq d/2$, and

$$q_\ell = \lceil 2^{d(2m - \ell) - (d/2 - r)(2m - \ell - p)} \rceil \tag{22}$$

12

if $r < d/2$.

In view of (14) and (16) we have

$$
\begin{aligned}
\mathbb{E}\zeta &= v - g + P_{m^*}f + P_m T_k u - v + g + \sum_{\ell=m+1}^{m^*} (P_\ell T_k u - P_{\ell-1} T_k u) \\
&= P_{m^*}f + P_{m^*}T_k u = P_{m^*}u.
\end{aligned}
\tag{23}
$$

Hence we have a biased random approximation. In the sequel we estimate the precision of the approximation

$$
\mathbb{E}\|\zeta - u\| \le \mathbb{E}\|\zeta - P_{m^*}u\| + \|P_{m^*}u - u\|.
\tag{24}
$$

By (3) (and the standard conclusion about the smoothness of u for $(k, f) \in X_0$) we have

$$
\|u - P_{m^*}u\| \le c2^{-rm^*}.
\tag{25}
$$

Moreover, by Hölder's inequality and (23)

$$
\begin{aligned}
(\mathbb{E}\|\zeta - P_{m^*}u\|)^2 &\le \mathbb{E}\|\zeta - P_{m^*}u\|^2 = \mathbb{E}\|\zeta - \mathbb{E}\zeta\|^2 \\
&= \mathbb{E}\left\|\sum_{\ell=m}^{m^*}(P_\ell \zeta_\ell - \mathbb{E}P_\ell \zeta_\ell)\right\|^2.
\end{aligned}
\tag{26}
$$

Define the restriction operator

$$
R_{m^*} : C(G) \to \ell_\infty(\Gamma_{m^*})
$$

by

$$
R_{m^*}f = f|_{\Gamma_{m^*}}
$$

for $f \in C(G)$. Then clearly

$$
P_{m^*}R_{m^*}P_\ell = P_\ell
$$

for $\ell \le m^*$. Using this, we continue (26) as

$$
= \mathbb{E}\left\|\sum_{\ell=m}^{m^*}(P_{m^*}R_{m^*}P_\ell \zeta_\ell - \mathbb{E}P_{m^*}R_{m^*}P_\ell \zeta_\ell)\right\|^2
$$

13

$$
\begin{aligned}
&= \quad \mathbb{E} \left\| P_{m^*} \sum_{\ell=m}^{m^*} (R_{m^*} P_\ell \zeta_\ell - \mathbb{E} R_{m^*} P_\ell \zeta_\ell) \right\|^2 \\
&\leq \quad c\, \mathbb{E} \left\| \sum_{\ell=m}^{m^*} (R_{m^*} P_\ell \zeta_\ell - \mathbb{E} R_{m^*} P_\ell \zeta_\ell) \right\|_{\ell_\infty(\Gamma_{m^*})}^2 \\
&= \quad c\, \mathrm{Var} \left(\sum_{\ell=m}^{m^*} R_{m^*} P_\ell \zeta_\ell \right) \\
&= \quad c\, \mathrm{Var} \left(\sum_{\ell=m}^{m^*} q_\ell^{-1} \sum_{i=1}^{q_\ell} R_{m^*} P_\ell \eta_\ell(\xi_{i\ell}) \right) \\
&\leq \quad cm^* \sum_{\ell=m}^{m^*} q_\ell^{-2} \sum_{i=1}^{q_\ell} \mathrm{Var}(R_{m^*} P_\ell \eta_\ell(\xi_{i\ell})) \\
&\leq \quad cm \sum_{\ell=m}^{m^*} q_\ell^{-2} \sum_{i=1}^{q_\ell} \mathrm{Var}(\eta_\ell(\xi_{i\ell})) \\
&= \quad cm \sum_{\ell=m}^{m^*} q_\ell^{-1} \mathrm{Var}(\eta_\ell), \quad\quad\quad\quad\quad\quad\quad (27)
\end{aligned}
$$

where we used Lemma 1. In the sequel we distinguish between three cases $r > d/2$, $r = d/2$ and $r < d/2$. In the first case we use (15), (17), and (21) to continue (27) as follows:

$$
\begin{aligned}
&\leq \quad cm \sum_{\ell=m}^{m^*} 2^{-dm+(r+d/2)(\ell-m)-2r\ell} \\
&\leq \quad cm \sum_{\ell=m}^{m^*} 2^{-(2r+d)m-(r-d/2)(\ell-m)} \\
&\leq \quad cm 2^{-(2r+d)m}. \quad\quad\quad\quad\quad\quad\quad\quad (28)
\end{aligned}
$$

Combining (18), (24) - (28), and (4), we obtain the desired estimate

$$
\begin{aligned}
\mathbb{E}\|\zeta - u\| \quad &\leq \quad cm^{1/2} 2^{-rm-dm/2} \\
&\leq \quad c(\log n)^{1/2} n^{-r/(2d)-1/4}.
\end{aligned}
$$

If $r = d/2$, we argue in the same way, with the exception that, due to the summation of inequality (28), another factor m appears, which results in

$$
\mathbb{E}\|\zeta - u\| \leq cn^{-1/2} \log n. \quad\quad\quad\quad\quad\quad (29)
$$

Finally, if $r < d/2$, we use (15), (17), and (22) and continue (27) as

$$
\begin{aligned}
&\leq & cm \sum_{\ell=m}^{m^*} 2^{-d(2m-\ell)+(d/2-r)(2m-\ell-p)-2r\ell} \\
&= & cm \sum_{\ell=m}^{m^*} 2^{-d(2m-\ell-p)-dp+(d/2-r)(2m-\ell-p)+2r(2m-\ell-p)-2r(2m-p)} \\
&= & cm \sum_{\ell=m}^{m^*} 2^{-dp-2r(2m-p)-(d/2-r)(2m-\ell-p)} \\
&\leq & cm2^{-dp-2r(2m-p)} \\
&\leq & c2^{-2r(2m-p)}.
\end{aligned}
\tag{30}
$$

The last two inequalities were consequences of (19) and (20). We combine (19), (20), (24 – 27), (30), and (4) and get

$$
\begin{aligned}
\mathbb{E}\|\zeta - u\| &\leq & c2^{-r(2m-p)} \\
&\leq & c(\log n)^{r/d} n^{-r/d}.
\end{aligned}
$$

To complete the proof of the upper bound of the theorem, we have to estimate the cardinality of the method, i. e. the expected number of function values. We shall, in addition, estimate the number of arithmetic operations, thus giving a complete analysis of the complexity also in the sense of Heinrich (1996).

It is known from Emelyanov and Ilin (1967) that the algorithm of deterministic computation of v_0 can be chosen in such a way that it requires $O(2^{2md})$ function values and arithmetic operations.

The same is true for the computation of v, h and g. The expected length of the random walk (5) is easily seen to be finite. Hence the expected number of function values and operations to compute (6) for one random walk ξ and for all $s \in \Gamma_m$ is $O(2^{dm})$. To compute (8) for one walk ξ and for all $s \in \Gamma_\ell$ we first compute the piecewise polynomial function

$$
P_{\ell-1}k(\,\cdot\,,t_q) \in \mathcal{P}^r(\pi_{\ell-1})
$$

from the kernel values $k(s,t_q)$ $(s \in \Gamma_{\ell-1})$ in $O(2^{d\ell})$ operations and then

$$
h(s,t_q) = k(s,t_q) - (P_{\ell-1}k(\,\cdot\,,t_q))(s)
$$

for all $s \in \Gamma_\ell$, again in $O(2^{d\ell})$ operations. Multiplying this with the number of samples q_ℓ in level ℓ, we obtain the cost of computing the vector $\zeta_\ell \in \ell_\infty(\Gamma_\ell)$ for $\ell = m, \ldots, m^*$. The final approximation (10) involves a summation of piecewise polynomial functions on $\pi_m, \pi_{m+1}, \ldots, \pi_{m^*}$. It is easily seen that this can be accomplished in

$$O(2^{dm} + \ldots + 2^{dm^*}) = O(2^{dm^*})$$

operations. Note also that we need $O(\sum_{\ell=m}^{m^*} q_\ell)$, that is, not more than $O(2^{dm})$ calls of a standard random number generator providing independent uniformly distributed on $[0, 1]$ samples. So the overall expected number $\bar{n}$ of function values and arithmetic operations for the stochastic part satisfies

$$\bar{n} \le c \sum_{\ell=m}^{m^*} q_\ell 2^{d\ell}.$$

In the case $r > d/2$, we obtain from (21)

$$\begin{aligned} \bar{n} &\le& c \sum_{\ell=m}^{m^*} 2^{dm-(r+d/2)(\ell-m)+d\ell} \\ &=& c \sum_{\ell=m}^{m^*} 2^{2dm+d(\ell-m)-(r+d/2)(\ell-m)} \\ &=& c \sum_{\ell=m}^{m^*} 2^{2dm-(r-d/2)(\ell-m)} \\ &\le& c2^{2dm} \le cn. \end{aligned}$$

In the case $r = d/2$, the argument is the same, but the summation gives

$$\bar{n} \le cm2^{2dm} \le cn \log n. \tag{31}$$

Finally, for $r < d/2$, we get from (19) and (22)

$$\begin{aligned} \bar{n} &\le& \sum_{\ell=m}^{m^*} 2^{d(2m-\ell)-(d/2-r)(2m-\ell-p)+d\ell} \\ &=& c \sum_{\ell=m}^{m^*} 2^{2dm-(d/2-r)(2m-\ell-p)} \\ &\le& c2^{2dm} \le cn. \end{aligned}$$

This obviously proves the upper bound of the theorem in the cases $r \ne d/2$. For $r = d/2$, we put $\tilde{n} = n\lceil \log n \rceil$ and obtain from (31) and (29) that our method needs an expected number of $O(\tilde{n})$ operations and function values and has expected error

$$\le cn^{-1/2} \log n \le c\tilde{n}^{-1/2} (\log \tilde{n})^{3/2},$$

yielding the upper bound also for $r = d/2$.

16

4 The lower bound

In this chapter we prove the lower estimate of the theorem. The general approach to such estimates consists in the reduction to the average case, a procedure first applied by Bakhvalov (1959). We shall consider only probability measures of finite (discrete) support, hence no measurability questions arise. So let ν be such a measure on X_0, let $N \in \mathcal{N}, \varphi \in \Phi$. Put

$$
\begin{aligned}
\mathrm{card}^{\mathrm{avg}}(N, \nu) &= \int_{X_0} \mathrm{card}(N(x)) \, d\nu(x), \\
e^{\mathrm{avg}}(S, N, \varphi, \nu) &= \int_{X_0} \|S(x) - \varphi(N(x))\| \, d\nu(x), \\
e_n^{\mathrm{avg}}(S, \nu) &= \inf\{e^{\mathrm{avg}}(S, N, \varphi, \nu) : \mathrm{card}^{\mathrm{avg}}(N) \leq n, \ \varphi \in \Phi\}.
\end{aligned}
$$

Lemma 2 *For each probability measure ν on X_0 of finite support and each $n \in \mathbb{N}$,*

$$
e_n^{\mathrm{MC}}(S) \geq \frac{1}{2} e_{2n}^{\mathrm{avg}}(S, \nu).
$$

Proof: Let $M = ((\Omega, \Sigma, \mu), (N_\omega, \varphi_\omega)_{\omega \in \Omega})$ be a Monte Carlo method with $\mathrm{card}(M) \leq n$. Hence

$$
\begin{aligned}
n &\geq \sup_{x \in X_0} \int_\Omega \mathrm{card}(N_\omega(x)) \, d\mu(\omega) \\
&\geq \int_{X_0} \int_\Omega \mathrm{card}(N_\omega(x)) \, d\mu(\omega) \, d\nu(x) \\
&= \int_\Omega \int_{X_0} \mathrm{card}(N_\omega(x)) \, d\nu(x) \, d\mu(\omega) \\
&= \int_\Omega \mathrm{card}^{\mathrm{avg}}(N_\omega, \nu) \, d\mu(\omega).
\end{aligned}
$$

We set

$$
\Omega_0 = \{\omega \in \Omega : \mathrm{card}^{\mathrm{avg}}(N_\omega, \nu) \leq 2n\}.
$$

The inequalities above imply $\mu(\Omega_0) \geq 1/2$. Now we have

$$
\begin{aligned}
e(S, M) &= \sup_{x \in X_0} \int_\Omega \|S(x) - \varphi_\omega(N_\omega(x))\| \, d\mu(\omega) \\
&\geq \int_{X_0} \int_\Omega \|S(x) - \varphi_\omega(N_\omega(x))\| \, d\mu(\omega) \, d\nu(x)
\end{aligned}
$$

$$
\begin{aligned}
&= \int_\Omega \int_{X_0} \| S(x) - \varphi_\omega(N_\omega(x)) \| \; d\nu(x) \; d\mu(\omega) \\
&= \int_\Omega e^{\mathrm{avg}}(S, N_\omega, \varphi_\omega, \nu) \; d\mu(\omega) \\
&\geq \mu(\Omega_0) \inf_{\omega \in \Omega_0} e^{\mathrm{avg}}(S, N_\omega, \varphi_\omega, \nu) \\
&\geq \frac{1}{2} e^{\mathrm{avg}}_{2n}(S, \nu).
\end{aligned}
$$

This proves the lemma.

Next let $n \in \mathbb{N}$, and let ε_{ij} $(i, j = 1, \dots, n)$ be independent Bernoulli random variables. We shall use the following result.

Lemma 3 *For $n \in \mathbb{N}$,*

$$
\mathbb{E} \max_{1 \leq i \leq n} \left| \sum_{j=1}^n \varepsilon_{ij} \right| \asymp (n \log n)^{1/2}
$$

and for $m, n \in \mathbb{N}$ with $2^{m-1} \leq n$,

$$
\mathbb{E} \max_{1 \leq i \leq n} \left| \sum_{j=1}^m \varepsilon_{ij} \right| \asymp m.
$$

Proof: The upper bounds follow from Lemma 1 above. The lower bound in the first relation is proved in Ledoux and Talagrand (1991), p. 120. The lower bound in the second relation follows from

$$
\begin{aligned}
\mu \left\{ \max_{1 \leq i \leq n} \left| \sum_{j=1}^m \varepsilon_{ij} \right| < m \right\} &\leq (1 - 2^{-(m-1)})^n \\
&\leq (1 - 2^{-(m-1)})^{2^{m-1}} < e^{-1}.
\end{aligned}
$$

This proves Lemma 3.

We shall construct a measure ν on X_0 and estimate $e^{\mathrm{avg}}_n(S, \nu)$. For this purpose, fix $n \in \mathbb{N}$ and choose $m \in \mathbb{N}$ such that

$$
2^{2d(m-1)-2} < 4n \leq 2^{2dm-2}. \tag{32}
$$

Put $p = m$ if $r \geq d/2$ and $p = [(\log_2 m)/d] + 1$ if $r < d/2$. Note that $1 \leq p \leq m$. Let

$$\begin{aligned}
I'_{2m-p} &= \{(i_1, \ldots, i_d) \in I_{2m-p} : \ 0 \le i_1 < 2^{2m-p-1}\} \\
I''_p &= \{(i_1, \ldots, i_d) \in I_p : \ 2^{p-1} \le i_1 < 2^p\}, \\
D &= I'_{2m-p} \times I''_p.
\end{aligned}$$

Then

$$|D| = 2^{2dm-2}. \tag{33}$$

Let ψ_0 be a C^∞ function on $\mathbb{R}$ with $\operatorname{supp}(\psi_0) \subseteq (0, 1)$ and

$$\int_{\mathbb{R}} \psi_0(s)ds = 1. \tag{34}$$

Put

$$\psi(s_1, \ldots, s_d) = \prod_{\ell=1}^{d} \psi_0(s_\ell).$$

Define

$$\begin{aligned}
\psi_{2m-p,i}(s) &= \psi(\tau_{2m-p,i}(s)) & (i \in I'_{2m-p}) \\
\psi_{p,j}(t) &= \psi(\tau_{p,j}(t)) & (j \in I''_p) \\
\widetilde{k}_{ij}(s,t) &= \psi_{2m-p,i}(s)\psi_{p,j}(t)
\end{aligned}$$

and

$$k_{ij}(s,t) = \min(\kappa_1, \kappa_2)\|\widetilde{k}_{ij}\|_r^{-1}\widetilde{k}_{ij}(s,t).$$

It is easily checked that $k_{ij} \in \mathcal{K}$,

$$\operatorname{supp}(k_{ij}) \subseteq G^0_{2m-p,i} \times G^0_{p,j}, \tag{35}$$

where A^0 denotes the interior of a set A, and

$$\|\widetilde{k}_{ij}\|_r \le c2^{r(2m-p)} \tag{36}$$

for $(i,j) \in I'_{2m-p} \times I''_p = D$. Now we are ready to define the measure ν. Let ε_{ij} $((i,j) \in D)$ be independent Bernoulli variables on some probability space

19

(Ω, Σ, μ). Let f_0 be the function on G with $f_0 \equiv \kappa_3$. Hence $f_0 \in \mathcal{F}$. Define a $\mathcal{K}$-valued discrete random variable h on (Ω, Σ, μ) by setting

$$h(\omega) = \sum_{(i,j) \in D} \varepsilon_{ij}(\omega) k_{ij}$$

and an X_0-valued variable by

$$z(\omega) = (h(\omega), f_0).$$

We set $\nu = \mu \circ z^{-1}$, so ν is a probability measure on X_0 with finite support X_1, where $X_1 = \mathcal{K}_1 \times \{f_0\}$,

$$\mathcal{K}_1 = \left\{ \sum_{(i,j) \in D} \beta_{ij} k_{ij} : \beta_{ij} = \pm 1 \right\}.$$

Observe that for all $k \in \mathcal{K}_1$,

$$\operatorname{supp}(k) \subseteq G' \times G'', \tag{37}$$

with

$$G' \quad = \quad \{(s_1, \ldots, s_d) \in G : 0 \le s_1 \le 1/2\}$$

and

$$G'' \quad = \quad \{(s_1, \ldots, s_d) \in G : 1/2 \le s_1 \le 1\}.$$

It follows from (37) that $T_k^2 = 0$ and hence

$$S(k, f) = (Id - T_k)^{-1} f = (Id + T_k) f = f + T_k f. \tag{38}$$

We shall use this relation later on. Now we estimate $e_n^{\mathrm{avg}}(S, \nu)$ form below. For this sake, we let $N \in \mathcal{N}$ with

$$\operatorname{card}^{\mathrm{avg}}(N, \nu) \le n, \tag{39}$$

and $\varphi \in \Phi$. We shall bound $e^{\mathrm{avg}}(S, N, \varphi, \nu)$ from below. Define

$$X_2 = \{x \in X_1 : \operatorname{card}(N(x)) \le 2n\}.$$

From (39) we infer

$$n \geq \int_{X_1 \backslash X_2} \mathrm{card}(N(x)) d\nu(x) \geq 2n\nu(X_1 \backslash X_2),$$

and consequently,

$$\nu(X_2) \geq 1/2. \tag{40}$$

Put $N(X_1) = A_1$, $N(X_2) = A_2$. Then we get

$$\mu\{N(z) \in A_2\} \geq 1/2. \tag{41}$$

We have

$$\int_{X_0} \|S(x) - \varphi(N(x))\| d\nu(x) = \mathbb{E}\|S(z) - \varphi(N(z))\|$$

$$= \sum_{a \in A_1} \mathbb{E}(\|S(z) - \varphi(N(z))\| \mid N(z) = a)\mu\{N(z) = a\}$$

$$\geq \sum_{a \in A_2} \mathbb{E}(\|S(z) - \varphi(N(z))\| \mid N(z) = a)\mu\{N(z) = a\}, \tag{42}$$

where the conditional expectation is just the expectation of $\|S(z) - \varphi(N(z))\|$ with respect to the conditional measure $(\mu|\{N(z) = a\})$. Next we fix $a \in A_2$, $a = (a_1, \ldots, a_{n(a)})$. By the definition of A_2 and X_2, $n(a) \leq 2n$. Let

$$\begin{aligned} t_1 & \\ t_2 &= t_2(a_1) \\ \ldots & \\ t_{n(a)} &= t_{n(a)}(a_1, \ldots, a_{n(a)-1}) \end{aligned}$$

be the support points (in $G^2 \cup G$) of information N produced for those $x \in X$ with $N(x) = a$ (recall the we admit only information consisting of function and derivative values). Define

$$D_a = \{(i,j) \in D : k_{ij}(t_q) = 0 \text{ for all } q \in \{1, \ldots, n(a)\} \text{ with } t_q \in G^2\}, \tag{43}$$

$$D'_a = D \backslash D_a.$$

It follows from the above, from (32) and (33) that

$$|D_a| \geq |D| - 2n \geq |D|/2 = 2^{2dm-3}. \tag{44}$$

Put

$$\begin{aligned} h_a(\omega) &= \sum_{(i,j) \in D_a} \varepsilon_{ij}(\omega) k_{ij}, \\ h'_a(\omega) &= \sum_{(i,j) \in D'_a} \varepsilon_{ij}(\omega) k_{ij}. \end{aligned}$$

Observe that for all $\omega \in \Omega$ with $N(z(\omega)) = a$,

$$\begin{aligned} N(h_a(\omega) + h'_a(\omega), f_0) &= N(-h_a(\omega) + h'_a(\omega), f_0) \\ &= N(h'_a(\omega), f_0), \end{aligned}$$

because of (43), and that $-h_a(\omega) + h'_a(\omega)$ has the same conditional distribution with respect to $\{N(z) = a\}$ as $h_a(\omega) + h'_a(\omega)$. Hence

$$\begin{aligned} &\mathbb{E}(\|S(z) - \varphi(N(z))\| \mid N(z) = a) \\ =\ & \mathbb{E}(\|S(h_a + h'_a, f_0) - \varphi(N(h_a + h'_a, f_0))\| \mid N(h, f_0) = a) \\ =\ & \mathbb{E}(\|S(h_a + h'_a, f_0) - \varphi(N(h'_a, f_0))\| \mid N(h'_a, f_0) = a) \\ =\ & \mathbb{E}(\|S(-h_a + h'_a, f_0) - \varphi(N(h'_a, f_0))\| \mid N(h'_a, f_0) = a) \\ \geq\ & \frac{1}{2}\mathbb{E}(\|S(h_a + h'_a, f_0) - S(-h_a + h'_a, f_0)\| \mid N(h'_a, f_0) = a) \end{aligned}$$

and in view of (38)

$$\begin{aligned} =\ & \frac{1}{2}\mathbb{E}(\|T_{h_a + h'_a} f_0 - T_{(-h_a + h'_a)} f_0\| \mid N(h'_a, f_0) = a) \\ =\ & \mathbb{E}(\|T_{h_a} f_0\| \mid N(h'_a, f_0) = a) \\ =\ & \mathbb{E}\|T_{h_a} f_0\|, \tag{45} \end{aligned}$$

because of the independence of h_a and h'_a.

In view of (34), (35) and (36) we can continue relation (45) above as follows:

$$\begin{aligned} =\ & \kappa_3 \mathbb{E} \max_{s \in G} \left| \sum_{(i,j) \in D_a} \varepsilon_{ij} \int_G k_{ij}(s,t)\, dt \right| \\ =\ & \kappa_3 \min(\kappa_1, \kappa_2) \|\widetilde{k}_{ij}\|_r^{-1} \mathbb{E} \max_{s \in G} \left| \sum_{(i,j) \in D_a} \varepsilon_{ij} \psi_{2m-p,i}(s) \int_G \psi_{p,j}(t)\, dt \right| \\ \geq\ & c 2^{-r(2m-p)-dp} \mathbb{E} \max_{i \in I'_{2m-p}} \left| \sum_{j:(i,j) \in D_a} \varepsilon_{ij} \right|. \tag{46} \end{aligned}$$

Denote $q_1 = 2^{d(2m-p)-3}$ and $q_2 = 2^{dp-3}$. Without loss of generality we can assume n to be so large that q_1 and q_2 are integers. Furthermore, we set

$$D_{a,i} = \{j : (i,j) \in D_a\}.$$

Then

$$|\{i : |D_{a,i}| \geq q_2\}| \geq q_1. \tag{47}$$

To show this, we assume the contrary. Hence

$$
\begin{aligned}
|D_a| &= \sum_{i \in I'_{2m-p}} |D_{a,i}| \\
&= \sum_{i:|D_{a,i}| \geq q_2} |D_{a,i}| + \sum_{i:|D_{a,i}| < q_2} |D_{a,i}| \\
&< q_1 2^{dp-1} + 2^{d(2m-p)-1} q_2 = 2^{2dm-3},
\end{aligned}
$$

contradicting (44), which proves (47).

Observe that the independence of the Bernoulli variables ε_{ij} implies

$$\mathbb{E} \max_i \left| \sum_{j:(i,j) \in D_a} \varepsilon_{ij} \right| \geq \mathbb{E} \max_i \left| \sum_{j:(i,j) \in B} \varepsilon_{ij} \right|$$

for any subset $B \subset D_a$. This together with (47) implies

$$\mathbb{E} \max_i \left| \sum_{j:(i,j) \in D_a} \varepsilon_{ij} \right| \geq \mathbb{E} \max_{1 \leq i \leq q_1} \left| \sum_{j=1}^{q_2} \varepsilon_{ij} \right|, \tag{48}$$

where ε_{ij}, $i = 1, \ldots, q_1$, $j = 1, \ldots, q_2$ are new independent Bernoulli variables. If $r \geq d/2$ and hence $p = m$, $q_1 = q_2 = 2^{dm-3}$, Lemma 3 gives

$$\mathbb{E} \max_{1 \leq i \leq q_1} \left| \sum_{j=1}^{q_2} \varepsilon_{ij} \right| \geq cm^{1/2} 2^{dm/2}. \tag{49}$$

In this case we can continue (46) by the help of (48) and (49) as follows:

$$\geq cm^{1/2} 2^{-rm-dm/2} \geq c(\log n)^{1/2} n^{-r/(2d)-1/4}. \tag{50}$$

23

Relations (45), (46), (50) together with (41) and (42) prove the lower bound in the case $r \geq d/2$. Now we assume $r < d/2$. Then $q_1 = 2^{d(2m-p)-3}$ and $q_2 = 2^{dp-3}$, with $p = [(\log_2 m)/d] + 1$. Put $q_2' = \min(q_2, d(2m-p) - 2)$. Then $2^{q_2'-1} \leq q_1$, and Lemma 3 gives

$$\mathbb{E} \max_{1 \leq i \leq q_1} \left| \sum_{j=1}^{q_2} \varepsilon_{ij} \right| \geq \mathbb{E} \max_{1 \leq i \leq q_1} \left| \sum_{j=1}^{q_2'} \varepsilon_{ij} \right| \geq cq_2'. \tag{51}$$

Using (48) and (51), we continue (46) in the following way

$$\begin{aligned}
&\geq & cq_2' 2^{-r(2m-p)-dp} \\
&\geq & cm 2^{-2rm+(r/d)\log_2 m - \log_2 m} \\
&\geq & cm^{r/d} 2^{-2rm} \geq c(\log n)^{r/d} n^{-r/d}.
\end{aligned} \tag{52}$$

Now relations (41), (42), (45), (46), and (52) imply the lower bound. This completes the proof of the theorem.

References

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

EMELYANOV, K. V. and ILIN, A. M. (1967), On the number of arithmetic operations, necessary for the approximate solution of Fredholm integral equations of the second kind (in Russian), *Zh. Vychisl. Mat. i Mat, Fiz.* **7**, 905 – 910.

ERMAKOV, S. M. (1971), "The Monte Carlo Method and Related Problems (in Russian)", Nauka, Moscow, 1971.

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.

HEINRICH, S. (1996), Complexity of Monte Carlo Algorithms, in: The Mathematics of Numerical Analysis, *Lectures in Applied Mathematics*, **32** (Park City, Utah), AMS-SIAM Summer Seminar, pp. 1621 – 1633.

HEINRICH, S. and MATHÉ, P. (1993), The Monte Carlo complexity of Fredholm integral equations, *Math. of Computation* **60**, 257 – 278.

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

LINDE, W. and PIETSCH, A. (1974), Mappings of Gaussian measures of cylindrical sets in Banach spaces, Theory Prob. and its Appl. **19**, 445 – 460.

MATHÉ, P. (1994), "Approximation Theory of Stochastic Numerical Methods", Habilitation Thesis, Berin.

MIKHAILOV, G. A. (1991a), "Minimization of Computational Costs of Non-Analogue Monte Carlo Methods", World Scientific, Singapore.

MIKHAILOV, G. A. (1991b), "Optimization of Weighted Monte Carlo Methods", Springer, Berlin-Heidelberg-New York.

MIKHAILOV, G. A. (1995), "New Monte Carlo Methods with Estimating Derivatives", VSP, Utrecht.

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.

SPANIER, J. and HOWELL, J. R. (1969), "Monte Carlo Principles and Neutron Transport Problems", Addison-Wesley, Reading, Massachusetts.

SIEGEL, R. and HOWELL, J. R. (1992), "Thermal Radiation Heat Transfer", 3rd ed., Hemisphere Publishing Corporation, New York.

TOMCZAK-JAEGERMANN, N. (1988), "Banach-Mazur Distances and Finite-Dimensional Operator Ideals, Longman, Essex, 1988.

TRAUB, J. F., WASILKOWSKI, G. W. and WOŹNIAKOWSKI, H. (1988), "Information-Based Complexity", Academic Press, New York.

VOYTISHEK, A. V. (1994), Asymptotics of convergence of discretely stochastic numerical methods of the global estimate of the solution to an integral equation of the second kind, *Sibirsk. Mat. Zh.* **35**, 728 – 736, (in R ussian).

VOYTISHEK, A. V. (1996), On the errors of discretely stochastic procedures in estimating globally the solution of an integral equation of the second kind, *Russian J. Numer. Anal. Math. Modelling* **11**, 71 – 92.