Wavelet Monte Carlo Methods for the Global Solution of Integral Equations

Stefan Heinrich

Fachbereich Informatik
Universität Kaiserslautern
D-67653 Kaiserslautern, Germany
e-mail: heinrich@informatik.uni-kl.de

Abstract. We study the global solution of Fredholm integral equations of the second kind by the help of Monte Carlo methods. Global solution means that we seek to approximate the full solution function. This is opposed to the usual applications of Monte Carlo, were one only wants to approximate a functional of the solution. In recent years several researchers developed Monte Carlo methods also for the global problem.

In this paper we present a new Monte Carlo algorithm for the global solution of integral equations. We use multiwavelet expansions to approximate the solution. We study the behaviour of variance on increasing levels, and based on this, develop a new variance reduction technique. For classes of smooth kernels and right hand sides we determine the convergence rate of this algorithm and show that it is higher than those of previously developed algorithms for the global problem. Moreover, an information-based complexity analysis shows that our algorithm is optimal among all stochastic algorithms of the same computational cost and that no deterministic algorithm of the same cost can reach its convergence rate.

1 Introduction

We are concerned with the randomized solution of linear integral equations

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

where $G \subseteq \mathbb{R}^d$ is a bounded domain, f and k are given functions on G and G^2, respectively, and u is the unknown function on G (detailed conditions will be given later). Examples of (1) include neutron transport, radiative heat transfer, and light transport. The function f is the source density and $k(s,t)$ describes the transition from state t to state s. Usually, d is high, e. g., $d = 6$ in neutron transport, so Monte Carlo methods are the preferred choice. In their standard form these methods provide statistical estimates of one (or a few) functional(s) of the solution u, such as the value $u(t_0)$ in a given point or a weighted mean

$$\int_G u(t)g(t)\,dt$$

with g a given function. It is well-known and supported both experimentally and theoretically, that for higher d, Monte Carlo is superior to deterministic methods in this task.

But what happens, if we seek to approximate the full solution function u instead of just a functional of it? Monte Carlo approaches to the full solution were developed by Frolov and Chentsov [4], Sobol [13], Mikhailov [10], Voytishek [15,16], Prigarin [11]. There are two basic ways:

1. Fix a grid $\Gamma \subset G$ and obtain Monte Carlo estimates of $u(s)$ for $s \in \Gamma$. Then extend the function from Γ to all of G by an interpolation or approximation procedure.
2. Use a basis sequence, say, an orthonormal system, to approximate

$$u \approx \sum_{i=1}^{n} (u, z_i) z_i$$

and estimate (u, z_i) by Monte Carlo.

Note that in both cases, the overall number of samples is $O(nN)$, where n is the number of functionals to be estimated ($n = |\Gamma|$ in case 1) and N is the number of samples for each functional.

In this paper we are concerned with the second approach (basis sequences). We present a new algorithm which improves upon the above by keeping the same precision but reducing the arithmetic cost considerably. This is achieved by using multiwavelet expansions and variance reduction tuned to their levels. For multilevel versions of the first approach (grid-interpolation) we refer to Heinrich [6,7].

2 Multiwavelets

In this section we briefly present the required facts about multiwavelets, a construction which goes back to Alpert [1]. For first applications to Monte Carlo see Heinrich [5]. We restrict ourselves to $G = [0,1]^d$. Fix an integer $r \geq 0$. For $\ell = 0, 1, \ldots$ (the level index), let

$$\pi_\ell = \{G_{\ell i} : i \in I_\ell = \{0, \ldots, 2^\ell - 1\}^d\}$$

be the partition of G into disjoint cubes of sidelength $2^{-\ell}$, where for $i = (i_1, \ldots, i_d) \in I_\ell$,

$$G_{\ell i} = A_{\ell i_1} \times \cdots \times A_{\ell i_d},$$

with $A_{\ell j} = [2^{-\ell} j, 2^{-\ell}(j+1))$ if $j < 2^\ell - 1$ and $A_{\ell j} = [2^{-\ell} j, 2^{-\ell}(j+1)]$ if $j = 2^\ell - 1$. Let $\mathcal{S}^r(\pi_\ell)$ be the space of splines of maximum degree not exceeding r, with no correlation at the interfaces, that is

$$f \in \mathcal{S}^r(\pi_\ell) \text{ iff } f|_{G_{\ell i}} \in \mathcal{P}^r(G_{\ell i}) \quad (i \in I_\ell),$$

and $\mathcal{P}^r(G_{\ell i})$ is the space of polynomials of maximum-degree $\leq r$ (so, e. g., $\mathcal{P}^1(G_{\ell i})$ is the set of multilinear functions on $G_{\ell i}$). The idea of the construction is the following: Choose any orthonormal basis $\{w_1, \ldots, w_q\}$ of $\mathcal{S}^r(\pi_0)$ (with respect to the

$L_2(G)$ norm) and extend it by $\{w_{q+1}, \dots, w_{q'}\}$ to an orthonormal basis of $\mathcal{S}^r(\pi_1)$. Now we "repeat" this process on higher levels: Assume that we have already constructed a basis of $\mathcal{S}^r(\pi_\ell)$. Its elements, restricted to $G_{\ell i}$ ($i \in I_\ell$), are just polynomials in $\mathcal{P}^r(G_{\ell i})$. Shrinking $w_{q+1}, \dots, w_{q'}$ to $G_{\ell i}$, we obtain functions which belong to $\mathcal{S}^r(\pi_{\ell+1})$ and which are orthogonal to $\mathcal{S}^r(\pi_\ell)$. Doing this for all i and normalizing, we get the desired orthonormal extension.

More formally, the extensions can be defined as follows: Let $L_\infty(G)$ be the usual space of essentially bounded functions on G with respect to the Lebesgue measure. Let

$$C_{\ell i} : L_\infty(G) \to L_\infty(G)$$

be the contraction operators defined for $f \in L_\infty(G)$ by

$$(C_{\ell i} f)(t) = \begin{cases} f(2^\ell t - i) & \text{for } t \in G_{\ell i} \\ 0 & \text{otherwise.} \end{cases}$$

Put

$$z_{0ij} = w_j \quad (i \in I_0,\, j = 1, \dots, q)$$

and, for $\ell \geq 1$,

$$z_{\ell ij} = 2^{(\ell-1)d/2} C_{\ell-1,i} w_{q+j} \quad (i \in I_{\ell-1},\, j = 1, \dots, q' - q).$$

Clearly, $\{z_{\ell ij} : \ell \leq m,\ i, j \text{ as above}\}$ is an orthonormal basis of $\mathcal{S}^r(\pi_m)$, and $z_{\ell ij}$ vanishes outside of $G_{\ell i}$. Note that for $r = 0$ we get the d-dimensional Haar basis.

We also need related interpolation operators. Let Γ_ℓ be the uniform grid on G of mesh size $r^{-1}2^{-\ell}$ for $r \geq 1$, and the set of centers of the cubes $G_{\ell i}$ ($i \in I_\ell$) for $r = 0$. Let $C(G)$ denote the space of continuous functions on G and let

$$P_\ell : C(G) \to \mathcal{S}^r(\pi_\ell)$$

be the piecewise multivariate Lagrange interpolation on Γ_ℓ of order r (i. e. the tensor product Lagrange interpolation on each $G_{\ell i}$). It is readily checked that

$$\|P_\ell : C(G) \to L_\infty(G)\| \leq c_1$$

and for $f \in C^\nu(G)$,

$$\|f - P_\ell f\|_{L_\infty(G)} \leq c_{2,\nu} 2^{-\nu\ell} \|f\|_{C^\nu(G)}, \tag{2}$$

where $\nu \leq r + 1$ is a positive integer, $C^\nu(G)$ stands for the space of ν times continuously differentiable functions, equipped with the usual norm

$$\|f\|_{C^\nu(G)} = \max_{|\alpha| \leq \nu} \sup_{t \in G} |(D^\alpha f)(t)|,$$

and the constants do not depend on ℓ. Observe that, due to the local structure of the interpolation, $(P_\ell f)(s)$ can be computed in $\mathcal{O}(1)$ operations, the constant not depending on ℓ.

3 Multiwavelet Monte Carlo

Here we present the new algorithm. First we specify the conditions on k and f in (1). We consider equation (1) in $L_\infty(G)$, where $G = [0,1]^d$ as above (endowed with the Lebesgue measure). We assume that $f \in C(G)$ and that k is such that $s \to k(s,\cdot)$ defines a continuous function from G to $L_1(G)$. It follows that the integral operator T_k defined by

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

for $g \in L_\infty(G)$, is bounded in $L_\infty(G)$. We assume that $I - T_k$ is invertible in $L_\infty(G)$, where I denotes the identity operator. Hence, (1) is uniquely solvable.

We fix a final level m and approximate u by

$$u \approx u_m = \sum_{\ell=0}^{m} \sum_{i,j} (u, z_{\ell ij}) z_{\ell ij},$$

where $(u, z_{\ell ij})$ denotes the scalar product of $L_2(G)$. Clearly, u_m is the orthogonal projection of u onto $\mathcal{S}^r(\pi_m)$. Our algorithm will approximate u_m by Monte Carlo. For this purpose, suppose that $\eta_{\ell ij}$ is a stochastic estimate of $(u, z_{\ell ij})$, that is, $\eta_{\ell ij}$ is a random variable on some (universal for the whole algorithm) probability space (Ω, Σ, μ) such that

$$\mathbb{E}\eta_{\ell ij} = (u, z_{\ell ij}).$$

The construction of $\eta_{\ell ij}$ is the crucial point and will be described later. We define a vector valued random variable

$$\eta_\ell = \sum_{i,j} \eta_{\ell ij} z_{\ell ij} \ \in \mathcal{S}^r(\pi_\ell) \ominus \mathcal{S}^r(\pi_{\ell-1}) \tag{3}$$

representing the contribution to level ℓ. Assume that the η_ℓ $(\ell = 0, \ldots, m)$ are independent. Fix natural numbers N_ℓ $(\ell = 0, \ldots, m)$ to be chosen later and let $\eta_\ell^{(a)}$ $(a = 1, \ldots, N_\ell)$ be independent copies of η_ℓ (so that $\{\eta_\ell^{(a)} : \ell = 0, \ldots, m,\ a = 1, \ldots, N_\ell\}$ constitutes an independent family of random variables. We approximate u by the $\mathcal{S}^r(\pi_m)$-valued random variable η:

$$u \approx \eta = \sum_{\ell=0}^{m} \frac{1}{N_\ell} \sum_{a=1}^{N_\ell} \eta_\ell^{(a)} = \sum_{\ell=0}^{m} \sum_{i,j} \left(\frac{1}{N_\ell} \sum_{a=1}^{N_\ell} \eta_{\ell ij}^{(a)} \right) z_{\ell ij}, \tag{4}$$

where $\eta_{\ell ij}^{(a)}$ denote the components of $\eta_\ell^{(a)}$. It follows that

$$\mathbb{E}\eta = \sum_{\ell=0}^{m} \sum_{i,j} (u, z_{\ell ij}) z_{\ell ij} = u_m. \tag{5}$$

Hence, our algorithm produces a biased, vector valued estimate. So far this is nothing but a formal splitting over the levels. The key point is the construction of $\eta_{\ell ij}$ tuned to the level ℓ. For this sake we note that since $z_{\ell ij}$ is orthogonal to $\mathcal{S}^r(\pi_{\ell-1})$,

$$(u, z_{\ell ij}) = (u - P_{\ell-1}u, z_{\ell ij}), \tag{6}$$

for $\ell \geq 1$. We use this relation in order to modify the standard von Neumann Ulam collision estimate (see e. g. Spanier and Gelbard [12], Ermakov and Mikhailov [3]) in the following way. First we introduce functions $g_\ell(s)$ and $h_\ell(s,t)$, $\ell = 0, \ldots, m$, by setting $g_0 = f$, $h_0 = k$, and for $\ell \geq 1$,

$$\begin{aligned} g_\ell &= f - P_{\ell-1}f \\ h_\ell(\cdot, t) &= k(\cdot, t) - P_{\ell-1}k(\cdot, t). \end{aligned}$$

Due to our requirements on f and k made at the beginning of this chapter, g_ℓ and h_ℓ are well-defined. Next we define a Markov chain on G. For this purpose let $p_{\ell ij}$, p_ℓ, and p be arbitrary measurable, non-negative functions on G and G^2, respectively, satisfying

$$\begin{aligned} \int_G p_{\ell ij}(t)\, dt &= 1, \\ \int_G p_\ell(s,t)\, dt &\leq 1, \\ \int_G p(s,t)\, dt &\leq 1, \end{aligned}$$

$$\begin{aligned} p_{\ell ij}(s) &\neq 0 \text{ whenever } z_{\ell ij}(s) \neq 0, \\ p_\ell(s,t) &\neq 0 \text{ whenever } h_\ell(s,t) \neq 0, \text{ and} \\ p(s,t) &\neq 0 \text{ whenever } k(s,t) \neq 0, \end{aligned}$$

for almost all $s \in G$ and $(s,t) \in G^2$, respectively. We also assume that the spectral radius of T_p in $L_\infty(G)$ is less than one.

For each choice of ℓ, i, j these functions define an absorbing Markov chain on G with initial density $p_{\ell ij}$, transition density to the next state p_ℓ, and transition density to all further states p. Almost all trajectories are finite. Let

$$(t_0, t_1, \ldots, t_\gamma)$$

be such a trajectory. That is, t_γ is the last state before absorption, which occurs with probability

$$1 - \int_G p_\ell(s,t)\, dt$$

and

$$1 - \int_G p(s,t)\, dt,$$

respectively. We define the modified collision estimate as follows. Put

$$\eta_{\ell ij} = \frac{z_{\ell ij}(t_0)}{p_{\ell ij}(t_0)} \left(g_\ell(t_0) + \frac{h_\ell(t_0,t_1)}{p_\ell(t_0,t_1)} \sum_{\alpha=1}^{\gamma} f(t_\alpha) \prod_{\beta=2}^{\alpha} \frac{k(t_{\beta-1},t_\beta)}{p(t_{\beta-1},t_\beta)} \right).$$

It is checked in the usual way, that

$$\mathbb{E}\eta_{\ell ij} = (u, z_{\ell ij}),$$

provided the spectral radius of $T_{|k|}$ in $L_\infty(G)$ is less than one, and that $\eta_{\ell ij}$ has finite variance provided $z_{\ell ij}^2/p_{\ell ij} \in L_1(G)$, the kernels h_ℓ^2/p_ℓ and k^2/p define bounded operators in $L_\infty(G)$, and the spectral radius of $T_{k^2/p}$ in $L_\infty(G)$ is less than one.

4 Convergence Analysis

Recall that we approximate the solution u of (1) by the random variable η defined in (4). We shall study the mean square norm error $e(\eta)$ defined as

$$e(\eta) = (\mathbb{E}\|u - \eta\|_{L_\infty(G)}^2)^{1/2}.$$

Denote the norm variance of η_ℓ by

$$v_\ell = \mathbb{E}\|\eta_\ell - \mathbb{E}\eta_\ell\|_{L_\infty(G)}^2,$$

and the metric distance of u to $\mathcal{S}^r(\pi_m)$ in the norm of $L_\infty(G)$ by $\operatorname{dist}(u, \mathcal{S}^r(\pi_m), L_\infty(G))$.

Lemma 1. *There is a constant $c > 0$ depending only on r and d such that*

$$e(\eta) \le c \left(\operatorname{dist}(u, \mathcal{S}^r(\pi_m), L_\infty(G)) + (m \sum_{\ell=0}^{m} v_\ell N_\ell^{-1})^{1/2} \right).$$

Proof. By (5) and the triangle inequality we have

$$e(\eta) \le \|u - u_m\|_{L_\infty(G)} + (\mathbb{E}\|\eta - \mathbb{E}\eta\|_{L_\infty(G)}^2)^{1/2}.$$

To estimate the deterministic component, let Q_m denote the orthogonal projection from L_2 onto $\mathcal{S}^r(\pi_m)$. Using a local representation of Q_m (see [5], section 8), it is readily shown that

$$\sup_m \|Q_m : L_\infty(G) \to L_\infty(G)\| < \infty.$$

It follows that

$$\begin{aligned} &\|u - u_m\|_{L_\infty(G)} = \|(I - Q_m)u\|_{L_\infty(G)} \\ &\le (1 + \|Q_m : L_\infty(G) \to L_\infty(G)\|) \operatorname{dist}(u, \mathcal{S}^r(\pi_m), L_\infty(G)). \end{aligned}$$

Let us now turn to the stochastic part. Here we have to estimate the variance of sums of independent, vector valued random variables. Recall that

$$\eta = \sum_{\ell=0}^{m} \frac{1}{N_\ell} \sum_{a=1}^{N_\ell} \eta_\ell^{(a)},$$

the $\eta_\ell^{(a)}$ being independent random variables with values in $\mathcal{S}^r(\pi_\ell) \subseteq \mathcal{S}^r(\pi_m)$. Since $\mathcal{S}^r(\pi_m)$, considered in the $L_\infty(G)$ norm, is isometric to the ℓ_∞-sum of 2^{dm} copies of $\mathcal{S}^r(\pi_0)$ (just use the fact that there is no correlation between functions on the subcubes of π_m), it follows that the type 2 constant of $\mathcal{S}^r(\pi_m)$ behaves like

$$(\log \dim \mathcal{S}^r(\pi_m))^{1/2} \asymp m^{1/2},$$

see Ledoux and Talagrand [9] for the notion of type 2 constants. (The standard notation $\asymp$ means that both quantities are equal up to a factor which can be bounded form above and below by positive constants not depending on m.) Then Proposition 9.11 of [9] gives

$$\begin{aligned} \mathbb{E}\|\eta - \mathbb{E}\eta\|_{L_\infty(G)}^2 &\le cm \sum_{\ell=0}^{m} \frac{1}{N_\ell^2} \sum_{a=1}^{N_\ell} \mathbb{E}\|\eta_\ell^{(a)} - \mathbb{E}\eta_\ell^{(a)}\|_{L_\infty(G)}^2 \\ &= cm \sum_{\ell=0}^{m} v_\ell N_\ell^{-1}. \end{aligned} \tag{7}$$

This proves the lemma.

The subsequent convergence analysis is carried out for a model class of smooth kernels and right hand sides. It simplifies the analysis, while the essential features of balancing variance over the levels become more transparent. Moreover, this class is well-studied from the point of view of information-based complexity, which allows us to formulate optimality results and comparisons with the deterministic setting. To define the class, we fix a positive integer ν – the degree of smoothness, and real parameters $\kappa_1, \kappa_3 > 0$, $0 < \kappa_2 < 1$, and put

$$\begin{aligned} \mathcal{K} &= \{k \in C^\nu(G^2) : \|k\|_{C^\nu(G^2)} \le \kappa_1, \|k\|_{L_\infty(G^2)} \le \kappa_2\} \\ \mathcal{F} &= \{f \in C^\nu(G) : \|f\|_{C^\nu(G)} \le \kappa_3\}. \end{aligned}$$

We choose the following algorithm parameters:

$$r = \nu - 1, \;\; p_{\ell ij} = |G_{\ell i}|^{-1} \chi_{G_{\ell i}},$$

$p = p_\ell \equiv \theta$, where θ is some constant, $\kappa_2 < \theta < 1$. Hence the initial distribution of the Markov chain is the uniform distribution on $G_{\ell i}$ and all further states are distributed uniformly on G, with $1 - \theta$ being the probability of absorption.

By cost of the algorithm we mean the expected number of arithmetic operations, comparisons, function evaluations (of k and f) and random number generator

calls. The efficiency of algorithms is judged and compared best when the error is expressed as a function of the overall cost. The next result provides this for the multiwavelet algorithm. We shall assume $d > 2\nu$, so the dimension is large compared to the smoothness (we comment on the case $d \leq 2\nu$ at the end).

Theorem 2. *Let $d > 2\nu$. There are constants $c > 0$ and $M_0 \in \mathbb{N}$ such that for each $M \in \mathbb{N}$ with $M > M_0$ there is a choice of the parameters m and $(N_\ell)_{\ell=0}^m$ such that for all $k \in \mathcal{K}$ and $f \in \mathcal{F}$ the multiwavelet Monte Carlo algorithm has cost at most M and error*

$$e(\eta) \leq cM^{-\nu/d}(\log M)^{\nu/d}.$$

Proof. Fix a non-negative integer m, to be chosen later. There is a constant $c > 0$ such that for all $k \in \mathcal{K}$ and $f \in \mathcal{F}$, $u = (I - I_k)^{-1} f$ satisfies

$$\|u\|_{C^\nu(G)} \leq c.$$

(We shall use the same symbol c for possibly different constants, all independent of M and m.) Together with (2) this yields

$$\operatorname{dist}(u, \mathcal{S}^r(\pi_m), L_\infty(G)) \leq c2^{-\nu m}. \tag{8}$$

Again from (2) and the assumptions on k and f we derive

$$\|g_\ell\|_{L_\infty(G)} \leq c2^{-\nu\ell}$$

and

$$\|h_\ell\|_{L_\infty(G^2)} \leq c2^{-\nu\ell}.$$

Since, by our choice,

$$\left|\frac{k(s,t)}{p(s,t)}\right| \leq \frac{\kappa_2}{\theta} < 1$$

and

$$\|z_{\ell ij}\|_{L_\infty(G)} \leq c2^{d\ell/2},$$

we get

$$|\eta_{\ell ij}| \leq c2^{-d\ell/2-\nu\ell}$$

and hence, by the disjointness (up to sets of measure zero) of the supports of $z_{\ell ij}$,

$$\|\eta_\ell\|_{L_\infty(G)} \leq c2^{-\nu\ell}.$$

This implies $v_\ell \leq c2^{-2\nu\ell}$, and the lemma above gives

$$e(\eta) \leq c\left(2^{-\nu m} + (m\sum_{\ell=0}^{m} 2^{-2\nu\ell} N_\ell^{-1})^{1/2}\right). \tag{9}$$

Since the expected length of the Markov chain depends only on θ, a realization of the random variable $\eta_{\ell ij}$ can be computed at expected cost $\mathcal{O}(1)$. The variable η_ℓ has $\mathcal{O}(2^{d\ell})$ components, so the cost of computing η_ℓ is $\mathcal{O}(2^{d\ell})$. Fixing an upper bound M of the overall cost, we minimize the right hand side of (9). In a first step we leave m fixed and minimize the second summand,

$$m \sum_{\ell=0}^{m} 2^{-2\nu\ell} N_\ell^{-1}$$

subject to the condition

$$\sum_{\ell=0}^{m} 2^{d\ell} N_\ell \leq M. \tag{10}$$

Note that since we are only interested in the order of these quantities, we can neglect constant factors. With this in mind, we can write the solution of the minimization as

$$N_\ell \asymp 2^{-(d/2-\nu)(m-\ell)-d\ell} M.$$

This choice gives

$$m \sum_{\ell=0}^{m} 2^{-2\nu\ell} N_\ell^{-1} \leq cmM^{-1} 2^{(d-2\nu)m}.$$

Next, m has to be chosen in such a way that deterministic and stochastic error (i. e. both summands on the right hand side of (9)) are in balance:

$$mM^{-1} 2^{(d-2\nu)m} \asymp 2^{-2\nu m},$$

thus

$$m 2^{dm} \asymp M,$$

which is satisfied iff

$$2^{dm} \asymp M(\log M)^{-1},$$

and as the final error bound we obtain

$$e(\eta) \leq cM^{-\nu/d} (\log M)^{\nu/d},$$

which proves the theorem.

This result should be compared with the lower bound on the same class of ν-smooth kernels and right hand sides obtained by the author in [6]. For the precise framework we refer to that paper and to Traub, Wasilkowski and Woźniakowski [14].

Theorem 3 ([6]). *Let $d > 2\nu$. There is a constant $c > 0$ such that for all $M \in \mathbb{N}$ the following holds: No stochastic algorithm of cost at most M can have a smaller error than*

$$cM^{-\nu/d}(\log M)^{\nu/d}.$$

It follows that our algorithm is of optimal order on the class $(\mathcal{K}, \mathcal{F})$. A different optimal multilevel algorithm using interpolation instead of wavelet decompositions was given by the author in [6], [7].

We are also able to compare the behaviour of stochastic and deterministic algorithms. The following lower bound (which is, in fact, the optimal order) was obtained by Emelyanov and Ilin [2]. For the framework we refer to [2], [6] or [14].

Theorem 4 ([2]). *There is a constant $c > 0$ such that for all $M \in \mathbb{N}$ the following holds: No deterministic algorithm of cost at most M can have a smaller error than*

$$cM^{-\nu/(2d)}.$$

Hence the rate of convergence of the multiwavelet Monte Carlo algorithm on the class of smooth inputs is roughly the square of that of the best deterministic algorithm (this assumes, of course, that we accept the comparison between deterministic and stochastic error criteria).

Let us now compare our multiwavelet algorithm with the previously developed one-level procedures mentioned in section 1 under points 1 and 2. If the ingredients are optimally chosen, the deterministic part of the error has the same estimate $n^{-\nu/d}$ for ν-smooth functions. The variance of the estimators for $u(s)$ or $(u, z_i)z_i$ is usually $\Theta(1)$ (even if one uses a wavelet expansion and applies the standard estimators without the modifications described in section 3). Hence the total error is of the order

$$n^{-\nu/d} + N^{-1/2}(\log n)^{1/2}.$$

Minimizing with respect to the cost constraint $nN \leq M$ gives

$$M^{-\frac{\nu}{d+2\nu}}(\log M)^{\frac{\nu}{d+2\nu}},$$

a rate worse than that of our new algorithm, but because of the condition $d > 2\nu$ still better than that of the best deterministic algorithm.

Let us finally mention that the convergence analysis in the case $d \leq 2\nu$ can be carried out similarly. However, for $d < 2\nu$ the algorithm is no longer optimal. An optimal algorithm for this case is given in [6]. To produce a corresponding multiwavelet algorithm, we need to combine the approach of the present paper with the separation of main part technique in [6], first developed in [8].

References

1. B. K. Alpert. A class of bases in L^2 for the sparse representation of integral operators. *SIAM J. Math. Anal.*, **24**: 246 – 262, 1993.

2. K. V. Emelyanov and A. M. Ilin. On the number of arithmetic operations, necessary for the approximate solution of Fredholm integral equations of the second kind. *Zh. Vychisl. Mat. i Mat, Fiz.*, **7**: 905 – 910, 1967 (in Russian).
3. S. M. Ermakov and G. A. Mikhailov. *Statistical Modelling*. Nauka, Moscow, 1982 (in Russian).
4. A. S. Frolov and N. N. Chentsov. On the calculation of certain integrals dependent on a parameter by the Monte Carlo method. *Zh. Vychisl. Mat. Mat. Fiz.*, **2**(4): 714 – 717, 1962 (in Russian).
5. S. Heinrich. Random approximation in numerical analysis. In: K. D. Bierstedt, A. Pietsch, W. M. Ruess, and D. Vogt, editors, *Functional Analysis*, pages 123 – 171. Marcel Dekker, 1994.
6. S. Heinrich. Monte Carlo complexity of global solution of integral equations. *J. Complexity*, **14**: 151 – 175, 1998.
7. S. Heinrich. A multilevel version of the method of dependent tests. In: S. M. Ermakov, Y. N. Kashtanov, and V. B. Melas, editors, *Proceedings of the 3rd St. Petersburg Workshop on Simulation*, pages 31 – 35. St. Petersburg University Press, 1998.
8. S. Heinrich and P. Mathé. The Monte Carlo complexity of Fredholm integral equations. *Math. of Computation*, **60**: 257 – 278, 1993.
9. M. Ledoux and M. Talagrand. *Probability in Banach Spaces*. Springer, Berlin–Heidelberg–New York, 1991.
10. G. A. Mikhailov. *Minimization of Computational Costs of Non-Analogue Monte Carlo Methods*. World Scientific, Singapore, 1991.
11. S. M. Prigarin. Convergence and optimization of functional estimates in statistical modelling in Sobolev's Hilbert spaces. *Russian J. Numer. Anal. Math. Modelling*, **10**(4): 325 – 346, 1995.
12. J. Spanier and E. M. Gelbard, *Monte Carlo Principles and Neutron Transport Problems*. Addison-Wesley, Reading, Massachusetts, 1969.
13. I. M. Sobol. *Computational Monte Carlo Methods*. Nauka, Moscow, 1973 (in Russian).
14. J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski. *Information-Based Complexity*. Academic Press, New York, 1988.
15. A. V. Voytishek. 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, 1994. (in Russian).
16. A. V. Voytishek. On the errors of discetely stochastic procedures in estimating globally the solution of an integral equation of the second kind. *Russian J. Numer. Anal. Math. Modelling*, **11**: 71 – 92, 1996.