

THE MONTE CARLO COMPLEXITY OF FREDHOLM INTEGRAL EQUATIONS

STEFAN HEINRICH AND PETER MATHÉ

ABSTRACT. A complexity study of Monte Carlo methods for Fredholm integral equations is carried out. We analyze the problem of computing a functional $\mu(y)$, where y is the solution of a Fredholm integral equation

$$y(s) = \int_{I^m} k(s, t)y(t) dt + f(s), \quad s \in I^m,$$

on the m -dimensional unit cube I^m , where the kernel k and right-hand side f are given r times differentiable functions. We permit stochastic numerical methods which can make use of function evaluations of k and f only.

All Monte Carlo methods known to the authors for solving the above problem are of the order $n^{-1/2}$, while the optimal deterministic methods yield rate $n^{-r/(2m)}$, thus taking into account the given smoothness of the data. Here, n denotes the (average) number of function evaluations performed. The optimal algorithm we present combines deterministic and stochastic methods in an optimal way. It can be seen that both rates—the standard Monte Carlo rate for general continuous data and the deterministic rate for r -smooth data—multiply. This provides the smallest error that stochastic methods of given computational cost can achieve.

1. INTRODUCTION

The paper is concerned with the efficiency of Monte Carlo methods for the approximate solution of integral equations. We consider Fredholm integral equations of the second kind on the m -dimensional unit cube, i.e.,

$$(1) \quad y(s) = \int_{I^m} k(s, t)y(t) dt + f(s), \quad s \in I^m,$$

where $I^m = [0, 1]^m$. Here, k and f are given continuous functions on $I^m \times I^m$ and I^m , respectively.

Monte Carlo methods are usually applied to find functionals of solutions. To this end, fix a functional $\mu \in C(I^m)'$ which is a Radon measure on I^m . Let us denote the duality between $C(I^m)$ and $C(I^m)'$ by $\langle \cdot, \cdot \rangle$. The problem we study can now be formulated as follows: Given k and f as above, compute $\langle y, \mu \rangle$, where y is the (unique) solution of (1). Of special interest is the case $\mu = \delta_{s_0}$, that is, we seek the value of y at s_0 , $s_0 \in I^m$ fixed. Other cases of interest are μ being the integral, weighted integral, etc.

Received by the editor November 26, 1990.

1991 *Mathematics Subject Classification.* Primary 65C05, 65R20.

We analyze the global behavior of Monte Carlo methods on classes of smooth data, more precisely, on sets of r times continuously differentiable kernels and right-hand sides. We determine the smallest possible error for stochastic methods of given computational cost. To obtain lower bounds, one has to formalize the notion of a Monte Carlo algorithm, which is done in the framework of information-based complexity theory [20]. The upper bounds are established by providing and analyzing a concrete algorithm.

The efficiency of Monte Carlo methods for integration and function approximation was studied in [3, 15, 16, 20]. Optimality and complexity of deterministic methods for Fredholm equations, considered over classes of kernels and right-hand sides, were analyzed in [6, 17, 18, 19, 11, 12]. There is also a vast literature on various stochastic methods for Fredholm integral equations (see [7, 9] and references therein). However, the rigorous mathematical investigation of the efficiency and complexity aspect of Monte Carlo methods is new, in particular the determination of optimal rates.

Our results, combined with those of Emelyanov and Ilin [6], also allow a certain comparison of deterministic and stochastic methods (which can, of course, only be a rough, heuristic one, owing to the different error criteria used). In this light the algorithm providing the best rate has some new features. The usual Monte Carlo methods yield the typical rate of $M^{-1/2}$ for M trials and do not react to any smoothness of the data (see [7]). Deterministic methods, in turn, do so, but yield only a dimension-dependent best rate of $M^{-r/(2m)}$ (M the number of arithmetic operations performed; see [6]). Our algorithm keeps both advantages, that is, respects smoothness and adds the Monte Carlo contribution, in this way giving the optimal rate of $M^{-r/(2m)-1/2}$. This is achieved by suitably combining standard deterministic and stochastic methods. Thus, the algorithm we present might also be of computational interest.

The paper is organized as follows. In §2 we provide the necessary notions and formulate the main results. Proofs of the lower and upper bounds are given in §§3 and 4, respectively. Section 5 contains some discussion of models of computation, the computational cost of the concrete algorithm, and the relation to deterministic methods. Basic references for Monte Carlo methods are [7, 9], and for information-based complexity, [20].

2. THE MAIN RESULT

First we introduce the classes of right-hand sides and kernels we shall consider. Let $C^r(I^m)$ and $C^r(I^m \times I^m)$ be the spaces of r times continuously differentiable functions (that means—to avoid ambiguity—the restrictions of C^r -functions defined on some neighborhood of the respective domains).

Let $\|\cdot\|_0$ denote the maximum norm on $C(I^m)$ and let $\|\cdot\|_r$ be the norm on $C^r(I^m)$ defined by

$$\|f\|_r = \max_{|a| \leq r} \|D^a f\|_0,$$

where $a = (\alpha_1, \dots, \alpha_m)$ is a multi-index and D^a stands for the respective partial derivative. The corresponding notation applies, of course, to

$$C^r(I^m \times I^m) = C^r(I^{2m}).$$

Given constants $\alpha, \beta > 0$, $0 < \gamma < 1$, and $r \in \mathbf{N}$, we denote

$$\mathcal{H}^r(\alpha, \gamma) = \{k \in C^r(I^m \times I^m), \|k\|_0 \leq \gamma, \text{ and } \|k\|_r \leq \alpha\}$$

and

$$\mathcal{F}^r(\beta) = \{f \in C^r(I^m), \|f\|_r \leq \beta\}.$$

The symbol Id always denotes the identity operating in $C(I^m)$. Let $T_k: C(I^m) \rightarrow C(I^m)$ denote the integral operator defined for $k \in C(I^m \times I^m)$ and $f \in C(I^m)$ by

$$(T_k f)(s) = \int_{I^m} k(s, t) f(t) dt.$$

The requirement $\|k\|_0 \leq \gamma$ implies $\|T_k\|_{C(I^m) \rightarrow C(I^m)} \leq \gamma$. From now on we suppose r, α, β, γ to be fixed. Denote

$$X = X_{\alpha, \beta, \gamma}^r = \mathcal{H}^r(\alpha, \gamma) \times \mathcal{F}^r(\beta).$$

Let finally $\mu \in C(I^m)'$ be fixed. Then the solution operator $S_\mu: X \rightarrow \mathbf{R}$ is the operator which maps the data (k, f) to the solution $\langle y, \mu \rangle$ of our problem, that is,

$$(2) \quad S_\mu(k, f) = \langle (\text{Id} - T_k)^{-1} f, \mu \rangle.$$

Note that the solution operator is nonlinear in k .

As a framework in which to study the numerical approximation of S_μ we use the theory of information-based complexity of Traub, Wasilkowski, and Woźniakowski [20].

Let $\mathcal{N}_{\text{st}}^n(X, \mathbf{R}^n)$ be the class of all standard information operators from X to \mathbf{R}^n , that is, the class of mappings of the form

$$N = (\delta_{(s_1, t_1)}, \delta_{(s_2, t_2)}, \dots, \delta_{(s_l, t_l)}, \delta_{s_{l+1}}, \dots, \delta_{s_n}),$$

i.e.,

$$N(k, f) := (k(s_1, t_1), \dots, k(s_l, t_l), f(s_{l+1}), \dots, f(s_n)),$$

where $0 \leq l \leq n$ and $s_i, t_j \in I^m$ ($i = 1, \dots, n, j = 1, \dots, l$). We adopt the convention that $\mathcal{N}_{\text{st}}^0(X, \mathbf{R}^0) = \{0\}$, 0 the zero mapping. The quantity $N(k, f)$ represents the information about the data (k, f) , which we shall use in the computational process.

Given $N \in \mathcal{N}_{\text{st}}^n(X, \mathbf{R}^n)$, we denote by $\Phi(N, \mathbf{R})$ the class of all mappings $\varphi: N(X) \rightarrow \mathbf{R}$. These mappings are called algorithms using information N . The number $\varphi(N(k, f))$ represents the outcome of the computational process performed at the respective information.

The class $\mathcal{M}^n(X, \mathbf{R})$ is formed by the composition of some information N and algorithm φ :

$$\mathcal{M}^n(X, \mathbf{R}) = \{\varphi \circ N, \varphi \in \Phi(N, \mathbf{R}), N \in \mathcal{N}_{\text{st}}^n(X, \mathbf{R}^n)\}.$$

The elements of $\mathcal{M}^n(X, \mathbf{R})$ are called (deterministic) methods—they symbolize the full computational process. Finally,

$$\mathcal{M}(X, \mathbf{R}) := \bigcup_{n \in \mathbf{N}} \mathcal{M}^n(X, \mathbf{R})$$

constitutes the class of all methods using standard information of some finite cardinality. The elements of $\mathcal{M}(X, \mathbf{R})$ are denoted by u .

Our definition of a random method follows [20, p. 417] (compare also [15]). Define the following σ -algebra:

$$\mathcal{F}(X, \mathbf{R}) := \sigma(\{u \in \mathcal{M}(X, \mathbf{R}), u(k, f) \in A\}, \\ (k, f) \in X \text{ and } A \subset \mathbf{R} \text{ a Borel set}).$$

It follows that for every $(k, f) \in X$ the mapping $u \rightarrow u(k, f)$ is a real random variable on $[\mathcal{M}(X, \mathbf{R}), \mathcal{F}(X, \mathbf{R})]$.

Quite generally, we shall call any probability P on $[\mathcal{M}(X, \mathbf{R}), \mathcal{F}(X, \mathbf{R})]$ a Monte Carlo method. This way we have a large abstract class, containing, in particular, the standard Monte Carlo methods. Furthermore, the cardinality function $\text{card}: \mathcal{M}(X, \mathbf{R}) \rightarrow \mathbf{N}$ is defined by

$$\text{card}(u) = \min\{n \in \mathbf{N}, u \in \mathcal{M}^n(X, \mathbf{R})\}.$$

This function is possibly not measurable. Therefore, we define the MC-cardinality of a Monte Carlo method P via an upper integral

$$\text{MC-card}(P) = \int_{\mathcal{M}(X, \mathbf{R})}^* \text{card}(u) dP(u) \\ := \inf \left\{ \int_{\mathcal{M}(X, \mathbf{R})} z(u) dP(u), z(u) \geq \text{card}(u) \right. \\ \left. \text{for all } u \text{ and } z \text{ is } \mathcal{F}(X, \mathbf{R})\text{-measurable} \right\}.$$

This quantity plays a crucial role, since it is closely tied to the computational cost of the method (see §5). By $\mathcal{P}^n(X, \mathbf{R})$ we denote the class of all Monte Carlo methods with MC-cardinality $\leq n$.

The error of a Monte Carlo method P for S_μ at $(k, f) \in X$ is defined by

$$e^{mc}(S_\mu, (k, f), P) = \int_{\mathcal{M}(X, \mathbf{R})} |S_\mu(k, f) - u(k, f)| dP(u),$$

and the error over the class X by

$$e^{mc}(S_\mu, X, P) = \sup_{(k, f) \in X} e^{mc}(S_\mu, (k, f), P).$$

The quantity we are interested in is the smallest error a Monte Carlo method of given cardinality can achieve:

$$e_n^{mc}(S_\mu, X) := \inf\{e^{mc}(S_\mu, X, P), P \in \mathcal{P}^n(X, \mathbf{R})\}.$$

The main result of the paper is the following.

Theorem 1. *For all $r \in \mathbf{N}$, $\alpha, \beta > 0$, and $0 < \gamma < 1$ there are constants $0 < c < C < \infty$ such that we have for $X = X_{\alpha, \beta, \gamma}^r$*

$$(3) \quad cn^{-r/(2m)-1/2} \leq \sup_{\|\mu\| \leq 1} e_n^{mc}(S_\mu, X) \leq Cn^{-r/(2m)-1/2}, \quad n \in \mathbf{N}.$$

In particular, we get a general lower bound for all Monte Carlo methods using (on the average) not more than n values of the kernel and right-hand side. Since our estimate is sharp up to a constant factor, it also says that there

is an abstract Monte Carlo method achieving this order of error. One may ask whether there are concrete, practicable algorithms with this property. This is indeed the case. Our proof of the upper bound consists in the description and analysis of such a Monte Carlo method. Theorem 1 also provides the basis of the complexity analysis. Let $\text{MC-comp}(S_\mu, X, \varepsilon)$ denote the complexity of solving the problem S_μ on X with error at most ε (see §5 for a detailed discussion). Then we have

Corollary 1. *For all $r \in \mathbf{N}$, $\alpha, \beta > 0$, and $0 < \gamma < 1$ there are constants $0 < c < C < \infty$ and $\varepsilon_0 > 0$ such that we have for $X = X_{\alpha, \beta, \gamma}^r$*

$$(4) \quad c \left(\frac{1}{\varepsilon}\right)^{2m/(r+m)} \leq \sup_{\|\mu\| \leq 1} \text{MC-comp}(S_\mu, X, \varepsilon) \leq C \left(\frac{1}{\varepsilon}\right)^{2m/(r+m)}, \quad 0 < \varepsilon < \varepsilon_0.$$

For a comparison, the corresponding deterministic counterpart of Theorem 1 is of interest, which follows from the results of Emelyanov and Ilin [6] (see §5 for more details). The deterministic rate is $n^{-r/(2m)}$ (and the corresponding complexity is $(\frac{1}{\varepsilon})^{2m/r}$). So the optimal Monte Carlo rate is indeed the standard Monte Carlo rate $n^{-1/2}$, improved by the rate of the optimal deterministic approximation, showing how to combine randomization and approximation in the most efficient way.

3. THE LOWER BOUND

The proof of the lower bound in (3) of Theorem 1 is based on a technique developed by Bakhvalov [3].

Let ν be any discrete probability measure on X . Denote by

$$e_n(S_\mu, X, \nu) = \inf \left\{ \int_X |S_\mu(k, f) - u(k, f)| d\nu(k, f), u \in \mathcal{M}^n(X, \mathbf{R}) \right\}$$

the error usually called the average-case error with respect to the measure ν . Since ν is discrete, no measurability problems occur.

Lemma 1. *For each $n \in \mathbf{N}$ we have*

$$e_n^{mc}(S_\mu, X) \geq \frac{1}{2} e_{2n}(S_\mu, X, \nu).$$

Proof (compare [15, Lemma 2(ii)]). Let P be any Monte Carlo method with MC-cardinality $\text{MC-card}(P) \leq n$. Given $0 < \varepsilon < 1$, let $z: \mathcal{M}(X, \mathbf{R}) \rightarrow \mathbf{R}$ be a measurable function such that $z(u) \geq \text{card}(u)$, $u \in \mathcal{M}(X, \mathbf{R})$, and

$$(5) \quad \int_{\mathcal{M}(X, \mathbf{R})} z(u) dP(u) \leq (1 + \varepsilon)n.$$

Denote $Z^{2n} = \{u \in \mathcal{M}(X, \mathbf{R}), z(u) \leq 2n\}$. Then $Z^{2n} \subset \mathcal{M}^{2n}(X, \mathbf{R})$ and Z^{2n} is measurable. Using Chebyshev's inequality, we can conclude from (5)

that $P(Z^{2n}) \geq (1 - \varepsilon)/2$. Moreover, using Fubini's theorem, we obtain

$$\begin{aligned} & \sup_{(k, f) \in X} \int_{\mathcal{M}(X, \mathbf{R})} |S_\mu(k, f) - u(k, f)| dP(u) \\ & \geq \sup_{(k, f) \in X} \int_{Z^{2n}} |S_\mu(k, f) - u(k, f)| dP(u) \\ & \geq \int_X \int_{Z^{2n}} |S_\mu(k, f) - u(k, f)| dP(u) d\nu(k, f) \\ & = \int_{Z^{2n}} \int_X |S_\mu(k, f) - u(k, f)| d\nu(k, f) dP(u) \\ & \geq P(Z^{2n}) \cdot \inf \left\{ \int_X |S_\mu(k, f) - u(k, f)| d\nu(k, f), u \in Z^{2n} \right\} \\ & \geq ((1 - \varepsilon)/2) \inf \left\{ \int_X |S_\mu(k, f) - u(k, f)| d\nu(k, f), u \in \mathcal{M}^{2n}(X, \mathbf{R}) \right\}. \end{aligned}$$

With $\varepsilon \rightarrow 0$ we arrive at

$$e_n^{mc}(S_\mu, X) \geq \frac{1}{2} e_{2n}(S_\mu, X, \nu). \quad \square$$

Thus, we have reduced the problem of Monte Carlo methods to an average-case problem with deterministic methods of fixed cardinality. Moreover, to prove lower bounds, we can specify the functional μ to be $\sigma = \delta_{s_0}$ for some fixed $s_0 \in I^m$.

In fact we shall prove the lower bound for all such δ_{s_0} . We can assume without loss of generality that $1/2 \leq s_{01} \leq 1$, where $s_0 = (s_{01}, \dots, s_{0m})$ (otherwise we modify the subsequent constructions in an obvious way). The following lemma makes it easier to handle the error. Let λ denote the integration functional on $C(I^m \times I^m)$, that is,

$$\langle k, \lambda \rangle = \int_{I^m} \int_{I^m} k(s, t) ds dt, \quad k \in C(I^m \times I^m).$$

Let $\theta = \frac{1}{2} \min(\alpha, \gamma)$. Then θ , as a constant function, belongs to $\mathcal{H}^r(\alpha, \gamma)$, and $T_\theta f = \theta \int_{I^m} f(t) dt$ is a constant function for every function f . Put

$$I_1 = [0, \frac{1}{2}] \times I^{m-1}, \quad J_1 = [\frac{1}{2}, 1] \times I^{m-1}$$

and define

$$\mathcal{B} = \{k \in C^r(I^m \times I^m), \|k\|_r \leq \theta, \text{supp}(k) \subset I_1 \times J_1\}.$$

Observe that for $k \in \mathcal{B}$ we have $\theta + k \in \mathcal{H}^r(\alpha, \gamma)$, $T_k^2 = 0$, and $k(s_0, t) = 0$ for all $t \in I^m$.

Lemma 2. *Let β, γ , and $\sigma = \delta_{s_0}$ be fixed as above. For all $k_1, k_2 \in \mathcal{B}$ we have*

$$|S_\sigma(\theta + k_1, \beta) - S_\sigma(\theta + k_2, \beta)| \geq \beta \theta |\langle k_1 - k_2, \lambda \rangle|.$$

Proof. Let $k \in \mathcal{B}$. Since $T_k^2 = 0$ and $T_\theta T_k \beta = \beta \theta \langle k, \lambda \rangle$, we easily verify the equation

$$\begin{aligned} (\text{Id} - T_{\theta+k})(\text{Id} + T_k)\beta &= (\text{Id} - T_\theta(\text{Id} + T_k))\beta \\ &= \beta(1 - \theta - \theta \langle k, \lambda \rangle). \end{aligned}$$

Since $1 - \theta - \theta\langle k, \lambda \rangle$ cannot vanish, we see that

$$(\text{Id} - T_{\theta+k})^{-1} \beta = \frac{1}{1 - \theta - \theta\langle k, \lambda \rangle} (\text{Id} + T_k) \beta.$$

Exploiting, moreover, $k(s_0, t) = 0, t \in I^m$, we see that

$$S_\sigma(\theta + k, \beta) = \frac{\beta}{1 - \theta - \theta\langle k, \lambda \rangle}.$$

Given kernels $k_1, k_2 \in \mathcal{B}$, we conclude

$$\begin{aligned} & |S_\sigma(\theta + k_1, \beta) - S_\sigma(\theta + k_2, \beta)| \\ &= \frac{\beta \theta |\langle k_1 - k_2, \lambda \rangle|}{(1 - \theta - \theta\langle k_1, \lambda \rangle)(1 - \theta - \theta\langle k_2, \lambda \rangle)} \geq \beta \theta |\langle k_1 - k_2, \lambda \rangle|. \quad \square \end{aligned}$$

Lemma 3. *There is a constant $C_1 > 0$ such that for all $n, q \in \mathbf{N}$ with $q \geq 2n$ and all choices of kernels $k_1, \dots, k_q \in \mathcal{B}$ with mutually disjoint supports there is a discrete measure ν on X such that*

$$e_{2n}(S_\sigma, X, \nu) \geq C_1 (q - 2n)^{1/2} \min_{1 \leq i \leq q} |\langle k_i, \lambda \rangle|.$$

Proof. Let $(\varepsilon_i)_{i=1}^q$ be a sequence of independent $\{+1, -1\}$ -valued random variables on some probability space $[\Omega, \rho]$ satisfying

$$\rho\{\varepsilon_i = -1\} = \rho\{\varepsilon_i = +1\} = \frac{1}{2} \quad \text{for all } i$$

(i.e., they are Bernoulli). Define a mapping $\psi: \Omega \rightarrow X$ by

$$\psi(\omega) = \left(\theta + \sum_{i=1}^q \varepsilon_i(\omega) k_i, \beta \right).$$

That this is indeed a mapping into X follows from the definition of θ and the assumptions on the k_i . Let ν be the image measure $\rho \cdot \psi^{-1}$ induced on X by ψ . Now we take any $u \in \mathcal{M}^{2n}(X, \mathbf{R})$ and represent it as $u = \varphi \cdot N$, $N \in \mathcal{N}_{st}^{2n}(X, \mathbf{R}^{2n})$, $\varphi \in \Phi(N, \mathbf{R})$,

$$N = (\delta_{(s_1, t_1)}, \dots, \delta_{(s_l, t_l)}, \delta_{s_{l+1}}, \dots, \delta_{s_{2n}}).$$

Let $A \subset \{1, \dots, q\}$ be the set of those i for which the support of k_i contains at least one of the pairs $(s_1, t_1), \dots, (s_l, t_l)$. By assumption, the supports of the k_i are disjoint, which implies $\text{card}(A) \leq 2n$. Let $B = \{1, \dots, q\} \setminus A$. Then $\text{card}(B) \geq q - 2n$ and

$$k_i(s_j, t_j) = 0 \quad (i \in B, j = 1, \dots, l).$$

Let us denote

$$\begin{aligned} g_1(\omega) &= \theta + \sum_{i \in A} \varepsilon_i(\omega) k_i, & (\omega \in \Omega), \\ g_2(\omega) &= \sum_{i \in B} \varepsilon_i(\omega) k_i & (\omega \in \Omega). \end{aligned}$$

Thus, we have $\psi(\omega) = (g_1(\omega) + g_2(\omega), \beta)$ for all $\omega \in \Omega$. Observe that $g_1 + g_2$ and $g_1 - g_2$ have the same distribution and that, by construction,

$$N(g_1(\omega) + g_2(\omega), \beta) = N(g_1(\omega) - g_2(\omega), \beta), \quad \omega \in \Omega.$$

Now we get

$$\begin{aligned}
 & \int_X |S_\sigma(k, f) - u(k, f)| d\nu(k, f) \\
 &= \int_\Omega |S_\sigma(g_1(\omega) + g_2(\omega), \beta) - \varphi(N(g_1(\omega) + g_2(\omega), \beta))| d\rho(\omega) \\
 &= \frac{1}{2} \sum_{\tau=\pm 1} \int_\Omega |S_\sigma(g_1(\omega) + \tau g_2(\omega), \beta) - \varphi(N(g_1(\omega) + \tau g_2(\omega), \beta))| d\rho(\omega) \\
 &\geq \frac{1}{2} \int_\Omega |S_\sigma(g_1(\omega) + g_2(\omega), \beta) - S_\sigma(g_1(\omega) - g_2(\omega), \beta)| d\rho(\omega) \\
 (6) \quad &\geq \beta\theta \int_\Omega |\langle g_2(\omega), \lambda \rangle| d\rho(\omega) = \beta\theta \int_\Omega \left| \sum_{i \in B} \varepsilon_i(\omega) \langle k_i, \lambda \rangle \right| d\rho(\omega) \\
 (7) \quad &\geq C_1 \left(\sum_{i \in B} \langle k_i, \lambda \rangle^2 \right)^{1/2} \geq C_1 (q - 2n)^{1/2} \min_{1 \leq i \leq q} |\langle k_i, \lambda \rangle|,
 \end{aligned}$$

where we used Lemma 2 to derive (6) and Khintchine's inequality with constant K_2 (see [13, Chapter IV, §5]) to derive (7) with $C_1 = \beta\theta K_2$. \square

It remains to produce suitable sets of k_i 's. Let $\varphi \in C^\infty(\mathbf{R})$, $\text{supp}(\varphi) \subset (0, 1)$, and $\int_0^1 \varphi \neq 0$. Let $l \in \mathbf{N}$ and define

$$\mathcal{I}_l = \{(i_1, \dots, i_m), 0 \leq i_1 \leq l-1, 0 \leq i_p \leq 2l-1, p = 2, \dots, m\},$$

and

$$\mathcal{J}_l = \{(j_1, \dots, j_m), l \leq j_1 \leq 2l-1, 0 \leq j_p \leq 2l-1, p = 2, \dots, m\}.$$

For $(\mathbf{i}, \mathbf{j}) \in \mathcal{I}_l \times \mathcal{J}_l$, $\mathbf{i} = (i_1, \dots, i_m)$, $\mathbf{j} = (j_1, \dots, j_m)$, and $(s, t) \in I^m \times I^m$, $s = (s_1, \dots, s_m)$, $t = (t_1, \dots, t_m)$, set

$$\tilde{k}_{(\mathbf{i}, \mathbf{j})}^l(s, t) := \prod_{p=1}^m \varphi(2ls_p - i_p) \prod_{p=1}^m \varphi(2lt_p - j_p).$$

It is easily seen that there are constants $C_2, C_3 > 0$ (independent of l) such that for $(\mathbf{i}, \mathbf{j}) \in \mathcal{I}_l \times \mathcal{J}_l$ we have

$$(8) \quad |\langle \tilde{k}_{(\mathbf{i}, \mathbf{j})}^l, \lambda \rangle| \geq C_2 l^{-2m}$$

and

$$(9) \quad \|\tilde{k}_{(\mathbf{i}, \mathbf{j})}^l\|_r \leq C_3 l^r.$$

Finally, let

$$(10) \quad k_{(\mathbf{i}, \mathbf{j})}^l = \theta \frac{\tilde{k}_{(\mathbf{i}, \mathbf{j})}^l}{\|\tilde{k}_{(\mathbf{i}, \mathbf{j})}^l\|_r}.$$

Since

$$\text{supp}(k_{(\mathbf{i}, \mathbf{j})}^l) \subset \prod_{p=1}^m \left(\frac{i_p}{2l}, \frac{i_p+1}{2l} \right) \times \prod_{p=1}^m \left(\frac{j_p}{2l}, \frac{j_p+1}{2l} \right),$$

it follows that $k_{(i,j)}^l \in \mathcal{B}$ and that they are of mutually disjoint support. Moreover,

$$|\langle k_{(i,j)}^l, \lambda \rangle| \geq \frac{C_2}{C_3} l^{-r-2m}, \quad (i, j) \in \mathcal{I}_l \times \mathcal{I}_l.$$

Note also that

$$\text{card}(\mathcal{I}_l \times \mathcal{I}_l) = 4^{m-1} l^{2m}.$$

We can accomplish the proof of the lower bound as follows. For each $n \in \mathbb{N}$ let l_n be the unique natural number satisfying

$$4^{m-1} l_n^{2m} \geq 4n > 4^{m-1} (l_n - 1)^{2m}.$$

Then $l_n < (2^{(2-m)/m} + 1)n^{1/(2m)}$. Moreover, setting $q_n = 4^{m-1} l_n^{2m}$, we get $q_n - 2n \geq q_n/2$. Now we apply Lemma 3 to $q = q_n$ and to the set $\{k_{(i,j)}^{l_n}, (i, j) \in \mathcal{I}_{l_n} \times \mathcal{I}_{l_n}\}$. Combined with Lemma 1 this yields

$$\begin{aligned} e_n^{mc}(S_\sigma, X) &\geq \frac{C_1}{2} \left(\frac{q_n}{2}\right)^{1/2} \frac{C_2}{C_3} l_n^{-r-2m} \\ &= 2^{m-5/2} C_1 C_2 C_3^{-1} l_n^{-r-m} \geq C_4 n^{-r/(2m)-1/2}, \end{aligned}$$

which finishes the proof of the lower bound.

4. THE UPPER BOUND—VARIANCE REDUCTION FOR LIPSCHITZ ESTIMATORS

Stochastic methods to solve Fredholm equations of the second kind are well known and used frequently (see [7, 9, 8]). All known methods yield the typical Monte Carlo rate $M^{-1/2}$ (for M trials), but do not respect the smoothness of the kernel and the right-hand side, respectively.

For the integration problem, among various techniques for variance reduction, the method of “separation of the main part” can be used to include smoothness properties of the integrand (see [7, 8]). This section provides an application of these ideas to Fredholm equations of the second kind, more generally, to numerical problems which allow estimators satisfying some Lipschitz condition.

Precisely, let us introduce the following notion.

Definition 1. Let (Y, d) be a metric space and $\xi = (\xi_x)_{x \in Y}$ be a family of (real-valued) square integrable random variables on a probability space $[\Omega, \mathcal{F}, P]$.

The family ξ is said to be Lipschitz (in mean) if there is a constant $C \geq 0$ such that

$$(\mathbf{E}|\xi_x - \xi_y|^2)^{1/2} \leq Cd(x, y).$$

(The smallest constant C satisfying the above estimate is called the Lipschitz constant.)

Conditions of this type are well known in the theory of stochastic processes and are related to the path continuity (see [14, §35.3]). The following lemma is an immediate consequence of the definition and contains the main idea of the variance-reduction technique.

Lemma 4. Let (Y, d) be a metric space and $\xi = (\xi_x)_{x \in Y}$ be Lipschitz with constant C . Then we have

1. $T(x) := \mathbf{E}\xi_x, x \in Y$, is Lipschitz with constant C on Y .

2. For $M \in \mathbf{N}$ and independent copies ξ^j of ξ , $1 \leq j \leq M$, define

$$\zeta_M(x) := \frac{1}{M} \sum_{j=1}^M \xi_x^j, \quad x \in Y.$$

Then we have

$$(11) \quad (\mathbf{E}|T(x) - [T(y) + \zeta_M(x) - \zeta_M(y)]|^2)^{1/2} \leq \frac{C}{\sqrt{M}} d(x, y), \quad x, y \in Y.$$

Proof. Since ξ_x is square integrable, the expectation $\mathbf{E}\xi_x$ exists. Moreover, we have

$$|T(x) - T(y)| = |\mathbf{E}\xi_x - \mathbf{E}\xi_y| \leq \mathbf{E}|\xi_x - \xi_y| \leq (\mathbf{E}|\xi_x - \xi_y|^2)^{1/2} \leq Cd(x, y).$$

To prove (11), we use standard properties of the variance of the sum of independent mean zero random variables $T(x) - \xi_x^j$:

$$\begin{aligned} & \mathbf{E}|T(x) - [T(y) + \zeta_M(x) - \zeta_M(y)]|^2 \\ &= \frac{1}{M^2} \mathbf{E} \left| \sum_{j=1}^M |(T(x) - \xi_x^j) - (T(y) - \xi_y^j)| \right|^2 \\ &= \frac{1}{M^2} \sum_{j=1}^M \mathbf{E}|(\xi_y^j - \xi_x^j) - (T(y) - T(x))|^2 \\ &\leq \frac{1}{M} \mathbf{E}|\xi_x - \xi_y|^2 \leq \frac{C^2}{M} d^2(x, y), \quad x, y \in Y. \quad \square \end{aligned}$$

The Monte Carlo method suggested by the above lemma can be described as follows. Suppose we are given a Monte Carlo method P on $\mathcal{M}(Y, \mathbf{R})$ for a solution operator S . This method gives rise to a family ξ of random variables on $[\mathcal{M}(Y, \mathbf{R}), \mathcal{F}(Y, \mathbf{R}), P]$ defined by

$$(12) \quad \xi_x(u) := u(x), \quad x \in Y, \quad u \in \mathcal{M}(Y, \mathbf{R}).$$

Further, if P is unbiased for S , that means,

$$S(x) = T(x) := \mathbf{E}\xi_x = \int_{\mathcal{M}(Y, \mathbf{R})} u(x) dP(u), \quad x \in Y,$$

then we can apply the above lemma. Given $x \in Y$, we choose a good approximation y to x for which $T(y)$ can be computed exactly. Knowing $T(y)$, we can improve the approximation by averaging independent trials with Monte Carlo method P , each realization of which is applied both to x and y . This procedure is working whenever the family $(\xi_x)_{x \in Y}$ of random variables defined by (12) is Lipschitz in mean on an appropriate space (Y, d) .

The Monte Carlo methods, applied to solve Fredholm equations of the second kind, are normally based on the use of absorbing Markov chains (see [7, 8, 9]). For our purposes it can be seen that the Markov chain we shall employ has a very simple description. Let $\Delta \in \mathbf{R}^m \setminus I^m$ be any element, the so-called absorbing state. Let $D := I^m \cup \{\Delta\}$ be the state space, equipped with the Borel σ -algebra \mathcal{D} . Given $0 < p < 1$ and the (initial) distribution μ on I^m , we introduce the distribution of a random sequence $(X_i)_{i=1}^\infty$ with values in $[D, \mathcal{D}]$

as probability P on $[\Omega, \mathcal{F}]$ with $\Omega = D^{\mathbb{N}}$ and $\mathcal{F} = \mathcal{D}^{\mathbb{N}}$. This is done by giving the transition probabilities. For $n \in \mathbb{N}$, $x \in I^m$, and λ the Lebesgue measure on I^m , let

$$P(X_1 \in A) = \mu(A),$$

$$P(X_{n+1} \in A | X_n = x) = p \cdot \lambda(A), \quad A \subset I^m, A \in \mathcal{D}, n \geq 1,$$

and

$$P(X_{n+1} = \Delta | X_n = x) = 1 - p.$$

Further,

$$P(X_{n+1} = \Delta | X_n = \Delta) = 1, \quad n \geq 1,$$

meaning that Δ is absorbing. In the notation of stochastic processes, $(X_i)_{i=1}^{\infty}$ is a stationary Markov chain with absorbing state Δ and transition densities $p(x, y) = p$. The Markov chain on $[D, \mathcal{D}]$ thus introduced can be completely described by p and the (initial) distribution μ . The lifetime τ of such a Markov chain $(X_i)_{i=1}^{\infty}$ is defined to be $\tau((X_i)_{i=1}^{\infty}) := \min\{k, X_k = \Delta\} - 1$ if there is a k with $X_k = \Delta$, and $\tau((X_i)_{i=1}^{\infty}) = \infty$ otherwise. It will be seen from Proposition 1 below that τ is finite with probability 1. The Monte Carlo method will be defined using this Markov chain.

Given $r \in \mathbb{N}$, α, β, γ , and $X = X'_{\alpha, \beta, \gamma}$ as above, we introduce for technical reasons another space X_0 . To this end, fix $\tilde{\gamma}$ with $\gamma < \tilde{\gamma} < 1$ and $\tilde{\beta}$ with $\beta < \tilde{\beta} < \infty$. Put

$$X_0 := X_0(\tilde{\beta}, \tilde{\gamma}) = \{(k, f), k \in C(I^m \times I^m) \text{ with } \|k\|_0 \leq \tilde{\gamma},$$

$$f \in C(I^m) \text{ with } \|f\|_0 \leq \tilde{\beta}\},$$

equipped with the metric $d((k, f), (h, g)) := \|k - h\|_0 + \|f - g\|_0$. The space X_0 contains X as a subset, and S_μ can be extended to X_0 in a natural way, since the defining equation (2) for S_μ also applies to $(k, f) \in X_0$. We shall use the same notation S_μ for the extended operator.

First we introduce and analyze the standard Monte Carlo method, based on this Markov chain (compare [7, 9]). This will be done for elements of X_0 . For this purpose we will specify p and the initial distribution μ . Suppose we are going to approximate S_μ for some $\mu \in C'(I^m)$. Since S_μ is linear in μ , and since each μ can be decomposed into its positive and its negative part, it is easy to see that it is enough to find Monte Carlo methods based on probabilities μ on I^m . So, given a probability μ_0 on I^m , we want to approximate S_{μ_0} . We choose the Markov chain $(X_i)_{i=1}^{\infty}$ with initial distribution μ_0 and transition density p with $\tilde{\gamma} < p < 1$, fixed from now on. The distribution of this chain is denoted by P_0 . For each $(k, f) \in X_0$ we define a random variable ξ on the probability space $[\Omega, \mathcal{F}, P_0]$ in the following way:

$$(13) \quad \xi((X_i)_{i=1}^{\infty}, k, f) := \begin{cases} \frac{1}{(1-p)} f(X_1) & \text{if } \tau((X_i)_{i=1}^{\infty}) = 1, \\ \frac{1}{(1-p)p^{\tau-1}} \left(\prod_{i=1}^{\tau-1} k(X_i, X_{i+1}) \right) f(X_\tau) & \text{if } 1 < \tau((X_i)_{i=1}^{\infty}) < \infty, \\ 0 & \text{otherwise.} \end{cases}$$

The following proposition contains all technical details needed below.

Proposition 1. *There holds*

1. $\mathbf{E}\tau((X_i)_{i=1}^\infty) = \frac{1}{1-p}$;
2. $\mathbf{E}\xi((X_i)_{i=1}^\infty, k, f) = S_{\mu_0}(k, f)$, $(k, f) \in X_0$;
3. *The family $(\xi_{(k,f)})_{(k,f) \in X_0}$ of random variables on $[\Omega, \mathcal{F}, P_0]$, defined by*

$$\xi_{(k,f)}((X_i)_{i=1}^\infty) := \xi((X_i)_{i=1}^\infty, k, f), \quad (k, f) \in X_0,$$

is Lipschitz with constant

$$C_5 \leq \sqrt{\frac{p}{(1-p)(p-\tilde{\gamma}^2)}} \max \left\{ 1, \frac{\sqrt{2}\tilde{\beta}}{p-\tilde{\gamma}^2} \right\}.$$

Proof. The calculus needed to prove the assertions is standard and explained in [9, Lemma 1.1.1]. Let us recall the following relation, exploited several times below. For convenience, let us introduce the following function:

$$\chi_{\{n\}}(\tau) = \begin{cases} 1, & \tau((X_i)_{i=1}^\infty) = n, \\ 0 & \text{elsewhere.} \end{cases}$$

If $F: [D^{\mathbf{N}}, \mathcal{D}^{\mathbf{N}}] \rightarrow \mathbf{R}$ is a measurable function and $F_n: (I^m)^n \rightarrow \mathbf{R}$ defined by $F_n(x_1, \dots, x_n) = F(x_1, \dots, x_n, \Delta, \Delta, \dots)$ is bounded, then we have

$$(14) \quad \begin{aligned} & \mathbf{E}(F((X_i)_{i=1}^\infty) \cdot \chi_{\{n\}}(\tau)) \\ &= \underbrace{\int \cdots \int}_{n\text{-times}} F_n(x_1, \dots, x_n) p^{n-1} (1-p) dx_2 \cdots dx_n \mu_0(dx_1). \end{aligned}$$

Using (14), we obtain

$$\mathbf{E}\tau((X_i)_{i=1}^\infty) = \sum_{n=1}^\infty n \mathbf{E}\chi_{\{n\}}(\tau) = \sum_{n=1}^\infty n p^{n-1} (1-p) = \frac{1}{1-p},$$

which proves 1.

Further,

$$\begin{aligned} \mathbf{E}\xi((X_i)_{i=1}^\infty, k, f) &= \sum_{n=1}^\infty \mathbf{E}(\xi((X_i)_{i=1}^\infty, k, f) \chi_{\{n\}}(\tau)) \\ &= \int f(x_1) \mu_0(dx_1) \\ &\quad + \sum_{n=2}^\infty \underbrace{\int \cdots \int}_{n\text{-times}} \left(\prod_{i=1}^{n-1} k(x_i, x_{i+1}) \right) f(x_n) dx_n \cdots dx_2 \mu_0(dx_1) \\ &= \sum_{n=1}^\infty \langle T_k^{n-1} f, \mu_0 \rangle \\ &= \langle (\text{Id} - T_k)^{-1} f, \mu_0 \rangle = S_{\mu_0}(k, f), \quad (k, f) \in X_0, \end{aligned}$$

proving 2.

It remains to prove 3. First observe that $\xi((X_i)_{i=1}^\infty, k, f)$ is linear in f , so that we can conclude

$$\begin{aligned} &\xi((X_i)_{i=1}^\infty, k, f) - \xi((X_i)_{i=1}^\infty, h, g) \\ &= \xi((X_i)_{i=1}^\infty, h, f - g) + \xi((X_i)_{i=1}^\infty, k, f) - \xi((X_i)_{i=1}^\infty, h, f), \end{aligned}$$

and it is sufficient to show that $\xi((X_i)_{i=1}^\infty, k, f)$ is Lipschitz with constant C_5 in the second and first arguments, separately. An application of (14) shows

$$\begin{aligned} \mathbf{E}|\xi((X_i)_{i=1}^\infty, h, f - g)|^2 &= \frac{1}{1 - p} \sum_{n=1}^\infty \langle T_{h^2/p}^{n-1}(f - g)^2, \mu_0 \rangle \\ &= \frac{1}{1 - p} \langle (\text{Id} - T_{h^2/p})^{-1}(f - g)^2, \mu_0 \rangle \\ &\leq \frac{1}{1 - p} \|(\text{Id} - T_{h^2/p})^{-1}\|_{C(I^m) \rightarrow C(I^m)} \|f - g\|_0^2. \end{aligned}$$

Since $\|h^2/p\|_0 \leq \tilde{\gamma}^2/p$, we obtain

$$(\mathbf{E}|\xi((X_i)_{i=1}^\infty, h, f - g)|^2)^{1/2} \leq \sqrt{\frac{p}{(p - \tilde{\gamma}^2)(1 - p)}} \|f - g\|_0.$$

The estimate with respect to the kernel is

$$\begin{aligned} &\mathbf{E}|\xi((X_i)_{i=1}^\infty, k, f) - \xi((X_i)_{i=1}^\infty, h, f)|^2 \\ &= \sum_{n=1}^\infty \mathbf{E}(|\xi((X_i)_{i=1}^\infty, k, f) - \xi((X_i)_{i=1}^\infty, h, f)|^2 \cdot \chi_{\{n\}}(\tau)) \\ &= \sum_{n=2}^\infty \frac{1}{(1 - p)p^{n-1}} \underbrace{\int \cdots \int}_{n\text{-times}} \left(\prod_{i=1}^{n-1} k(x_i, x_{i+1}) - \prod_{i=1}^{n-1} h(x_i, x_{i+1}) \right)^2 \\ &\quad \cdot f^2(x_n) dx_n \cdots dx_2 d\mu_0(x_1) \end{aligned}$$

$$(15) \quad \leq \frac{\|f\|_0^2}{1 - p} \sum_{n=2}^\infty \left(\frac{1}{p}\right)^{n-1} (n - 1)^2 \|k - h\|_0^2 (\tilde{\gamma}^2)^{n-2}$$

$$\leq \frac{\|f\|_0^2 \|k - h\|_0^2}{(1 - p)\tilde{\gamma}^2} \sum_{n=2}^\infty (n - 1)^2 \left(\frac{\tilde{\gamma}^2}{p}\right)^{n-1}$$

$$\begin{aligned} (16) \quad &= \frac{\|f\|_0^2 \|k - h\|_0^2}{(1 - p)\tilde{\gamma}^2} \frac{(\tilde{\gamma}^2/p)(1 + \tilde{\gamma}^2/p)}{(1 - \tilde{\gamma}^2/p)^3} \\ &= \frac{\|f\|_0^2 \|k - h\|_0^2 p(p + \tilde{\gamma}^2)}{(1 - p)(p - \tilde{\gamma}^2)^3} \leq \frac{2p\tilde{\beta}^2}{(1 - p)(p - \tilde{\gamma}^2)^3} \|k - h\|_0^2. \end{aligned}$$

The derivation of (15) uses the easily verified relation

$$\left| \prod_{i=1}^n a_i - \prod_{i=1}^n b_i \right| \leq n \left(\max_{1 \leq i \leq n} (|a_i|, |b_i|) \right)^{n-1} \max_{1 \leq i \leq n} |a_i - b_i|.$$

For the value of the sum that gave (16) see [4, 5.2.25]. Hence,

$$\begin{aligned}
 & (\mathbf{E}|\xi((X_i)_{i=1}^\infty, k, f) - \xi((X_i)_{i=1}^\infty, h, g)|^2)^{1/2} \\
 & \leq (\mathbf{E}|\xi((X_i)_{i=1}^\infty, k, f) - \xi((X_i)_{i=1}^\infty, h, f)|^2)^{1/2} \\
 & \quad + (\mathbf{E}|\xi((X_i)_{i=1}^\infty, h, f - g)|^2)^{1/2} \\
 & \leq \sqrt{\frac{2p\tilde{\beta}^2}{(1-p)(p-\tilde{\gamma}^2)^3}}\|k - h\|_0 + \sqrt{\frac{p}{(1-p)(p-\tilde{\gamma}^2)}}\|f - g\|_0 \\
 & \leq \sqrt{\frac{p}{(1-p)(p-\tilde{\gamma}^2)}} \max\left\{1, \frac{\sqrt{2}\tilde{\beta}}{p-\tilde{\gamma}^2}\right\} (\|k - h\|_0 + \|f - g\|_0),
 \end{aligned}$$

which finishes the proof of assertion 3 and of the proposition. \square

So far, we assembled the facts about the standard Monte Carlo approach. Next we turn to the deterministic setting before we shall combine both to obtain the final, optimal Monte Carlo method. First recall a known fact from approximation theory (see, e.g., [5, Chapters 2.2 and 3.1]).

Proposition 2. *There exist constants $\kappa, C_6, C_7 > 0$, a function $q: \mathbf{N} \rightarrow \mathbf{N}$ with $q(l) \leq \kappa \cdot l$, a family $((t_{i,l})_{i=1}^{q(l)})_{l=1}^\infty$ of elements in I^m , and a family of linearly independent continuous functions $((\varphi_{i,l})_{i=1}^{q(l)})_{l=1}^\infty \subset C(I^m)$ such that*

$$(17) \quad \sup_{i,l} \|\varphi_{i,l}\|_0 < \infty,$$

$$(18) \quad \max_{i,l} \text{card}\{j : \text{supp}(\varphi_{i,l}) \cap \text{supp}(\varphi_{j,l}) \neq \emptyset\} < \infty.$$

The mappings $P_l: C(I^m) \rightarrow C(I^m)$ and $Q_l: C(I^m \times I^m) \rightarrow C(I^m \times I^m)$, defined by

$$(19) \quad (P_l f)(s) := \sum_{i=1}^{q(l)} f(t_{i,l}) \varphi_{i,l}(s)$$

and

$$(20) \quad (Q_l k)(s, t) = \sum_{i=1}^{q(l)} \sum_{j=1}^{q(l)} k(t_{i,l}, t_{j,l}) \varphi_{i,l}(s) \varphi_{j,l}(t),$$

satisfy for $f \in C^r(I^m)$ and $k \in C^r(I^m \times I^m)$

$$(21) \quad \|f - P_l f\|_0 \leq C_6 l^{-r/m} \|f\|_r$$

and

$$(22) \quad \|k - Q_l k\|_0 \leq C_7 l^{-r/m} \|k\|_r.$$

Furthermore, (17) and (18) imply

$$(23) \quad \sup_l \|P_l\|_{C(I^m) \rightarrow C(I^m)} < \infty$$

and

$$(24) \quad \sup_l \|Q_l\|_{C(I^m \times I^m) \rightarrow C(I^m \times I^m)} < \infty.$$

The following fact is well known (see, e.g., [6; 10, Chapter 16]) and expresses the existence of deterministic methods of a certain approximation rate (in fact, the optimal rate for the worst-case setting).

Proposition 3. *There exists a constant $C_8 > 0$ such that for all $l \in \mathbf{N}$ there is an $N_l \in \mathcal{N}_{\text{st}}^{l^2}(X, \mathbf{R}^{l^2})$ and a mapping $\psi_l: N_l(X) \rightarrow C(I^m)$ such that with $v_l = \psi_l \circ N_l$ and $R(k, f) := (\text{Id} - T_k)^{-1}f$, $(k, f) \in X_0$,*

$$(25) \quad \sup_{(k, f) \in X} \|v_l(k, f) - R(k, f)\|_0 \leq C_8 l^{-r/m}.$$

Given $(k, f) \in X$, and using both Proposition 2 and Proposition 3, we can find $(h, g) \in X_0$ near (k, f) such that the solution $R(h, g)$ can be given exactly. For this purpose, we define

$$(26) \quad w_l(k, f) := (h, g) := (Q_l k, (\text{Id} - T_{Q_l k})P_l v_l(k, f)).$$

Lemma 5. *There exist $C_9 > 0$ and $l_0 \in \mathbf{N}$ such that for all $l \geq l_0$ the transformation w_l maps X into X_0 ,*

$$(27) \quad \sup_{(k, f) \in X} d((k, f), w_l(k, f)) \leq C_9 l^{-r/m}, \quad l \geq l_0,$$

and

$$(28) \quad S_{\mu_0}(w_l(k, f)) = \langle P_l v_l(k, f), \mu_0 \rangle.$$

Proof. Let l_1 be such that for all $k \in \mathcal{K}^r(\alpha, \gamma)$

$$\|k - h\|_0 \leq C_7 l^{-r/m} \|k\|_r \leq C_7 \alpha l^{-r/m} \leq \tilde{\gamma} - \gamma, \quad l \geq l_1.$$

This yields $\|h\|_0 \leq \tilde{\gamma}$ for $l \geq l_1$. For $(k, f) \in X$ we have

$$\begin{aligned} & \|f - (\text{Id} - T_h)P_l v_l(k, f)\|_0 \\ & \leq \|(\text{Id} - T_k)R(k, f) - (\text{Id} - T_h)R(k, f)\|_0 \\ & \quad + \|(\text{Id} - T_h)R(k, f) - (\text{Id} - T_h)P_l R(k, f)\|_0 \\ & \quad + \|(\text{Id} - T_h)P_l R(k, f) - (\text{Id} - T_h)P_l v_l(k, f)\|_0 \\ & \leq C_{10} l^{-r/m}. \end{aligned}$$

Here we made use of (21), (22), (23), and (24), together with the fact that $(\text{Id} - T_k)^{-1}$ is uniformly bounded on $C^r(I^m)$ for $k \in \mathcal{K}^r(\alpha, \gamma)$. Now we choose l_2 in such a way that for $l \geq l_2$

$$\|f - (\text{Id} - T_h)P_l v_l(k, f)\|_0 \leq \hat{\beta} - \beta,$$

and put $l_0 = \max\{l_1, l_2\}$. It follows that for $l \geq l_0$, the transformation w_l maps X into X_0 and (27) holds. The representation (28) is an immediate consequence of the definitions. \square

Let us put for $(k, f) \in X_0$

$$(29) \quad \hat{u}_l(k, f) = S_{\mu_0}(w_l(k, f)) = \langle P_l v_l(k, f), \mu_0 \rangle.$$

We are now in a position to describe the Monte Carlo method providing the upper estimate in (3). For each $l \geq l_0$ and $M \in \mathbf{N}$ we shall define a Monte Carlo method $P_{l, M}$ on $[\mathcal{M}(X, \mathbf{R}), \mathcal{F}(X, \mathbf{R})]$ as the distribution of a random variable u_{ω_M} on

$$\left[(D^{\mathbf{N}})^M, (\mathcal{D}^{\mathbf{N}})^M, \bigotimes_{j=1}^M P_0 \right]$$

with values $\mathcal{M}(X, \mathbf{R})$. So let

$$\omega_M = (\omega^1, \dots, \omega^M) = ((X_i^1)_{i=1}^\infty, \dots, (X_i^M)_{i=1}^\infty) \in (D^{\mathbf{N}})^M$$

be such that $1 \leq \tau(\omega^j) < \infty$ for $j = 1, \dots, M$ (which holds for almost all ω_M). Define for $(k, f) \in X$

$$(30) \quad u_{\omega_M}(k, f) = \tilde{u}_l(k, f) + \frac{1}{M} \sum_{j=1}^M (\xi(\omega^j, k, f) - \xi(\omega^j, w_l(k, f))).$$

The components of u_{ω_M} are defined by (29), Proposition 3, (26), and (13). A look at these relations shows that u_{ω_M} “uses” the following information:

$$N_{\omega_M} = (V_l, N_l, N_{\omega^1}, \dots, N_{\omega^M}).$$

Here, $V_l \in \mathcal{N}_{\text{st}}^{q(l)^2}(X, \mathbf{R}^{q(l)^2})$ is defined as

$$V_l(k, f) = (k(t_{i,l}, t_{j,l}))_{i,j=1}^{q(l)}$$

(used for $Q_l k$ in the definition of $w_l(k, f)$). The information $N_l \in \mathcal{N}_{\text{st}}^{l^2}(X, \mathbf{R}^{l^2})$ comes from Proposition 3 (and is used for $v_l(k, f)$). For $j = 1, \dots, M$, and $\omega^j = (X_i^j)_{i=1}^\infty$ fixed above, with $1 \leq \tau(\omega^j) < \infty$, the information operator $N_{\omega^j} \in \mathcal{N}_{\text{st}}^{\tau(\omega^j)}(X, \mathbf{R}^{\tau(\omega^j)})$ is defined as

$$N_{\omega^j}(k, f) = \begin{cases} f(X_1) & \text{if } \tau(\omega^j) = 1, \\ ((k(X_i, X_{i+1}))_{i=1}^{\tau(\omega^j)-1}, f(X_{\tau(\omega^j)})) & \text{if } \tau(\omega^j) > 1. \end{cases}$$

Hence,

$$N_{\omega_M} \in \mathcal{N}_{\text{st}}^{\eta(\omega_M, l)}(X, \mathbf{R}^{\eta(\omega_M, l)}),$$

where $\eta(\omega_M, l) = l^2 + q(l)^2 + \sum_{j=1}^M \tau(\omega^j)$. In other words, for each ω_M as above, there is a mapping $\varphi_{\omega_M} \in \Phi(N_{\omega_M}, \mathbf{R})$ such that $u_{\omega_M} = \varphi_{\omega_M} \circ N_{\omega_M}$. In order to show that u_{ω_M} is a random variable, we have to verify that the mapping $\omega_M \rightarrow u_{\omega_M}$ is $[(D^{\mathbf{N}})^M, \mathcal{F}(X, \mathbf{R})]$ -measurable. But this is easily derived from the structure of u_{ω_M} . Consequently, the distribution $P_{l, M}$ of u_{ω_M} is a probability on $[\mathcal{M}(X, \mathbf{R}), \mathcal{F}(X, \mathbf{R})]$, hence a Monte Carlo method. Moreover,

$$\begin{aligned} \text{MC-card}(P_{l, M}) &\leq \mathbf{E}\eta(\omega_M, l) = l^2 + q(l)^2 + \sum_{j=1}^M \mathbf{E}\tau(\omega^j) \\ &\leq l^2 + q(l)^2 + M \frac{1}{1-p} \\ &\leq C_{11}(l^2 + M), \quad l \geq l_0, \quad M \in \mathbf{N}. \end{aligned}$$

Using Proposition 1, we can see that $\mathbf{E}\xi((X_i)_{i=1}^\infty, w_l(k, f)) = S_{\mu_0}(w_l(k, f)) = \tilde{u}_l(k, f)$ and that the family $(\xi_{(k, f)}(k, f))_{(k, f) \in X_0}$ is Lipschitz in mean with constant

C_5 . Inequalities (11) in Lemma 4 and (27) in Lemma 5 provide the estimate

$$\begin{aligned} & \int |S_{\mu_0}(k, f) - u(k, f)| dP_{l, M}(u) \\ &= \int \left| S_{\mu_0}(k, f) - \left\{ \tilde{u}(k, f) + \frac{1}{M} \sum_{j=1}^M (\xi(\omega^j, k, f) - \xi(\omega^j, w_l(k, f))) \right\} \right| dP_0^M(\omega_M) \\ &\leq \left(\int \left| S_{\mu_0}(k, f) - \left\{ S_{\mu_0}(w_l(k, f)) + \frac{1}{M} \sum_{j=1}^M (\xi(\omega^j, k, f) - \xi(\omega^j, w_l(k, f))) \right\} \right|^2 dP_0^M(\omega_M) \right)^{1/2} \\ &\leq C_5 \cdot \frac{d((k, f), w_l(k, f))}{\sqrt{M}} \leq C_5 C_9 \cdot \frac{l^{-r/m}}{\sqrt{M}}, \quad (k, f) \in X. \end{aligned}$$

We conclude

$$e^{mc}(S_{\mu_0}, X, P_{l, M}) \leq \frac{C_{12}}{\sqrt{M}} l^{-r/m}, \quad l \geq l_0, \quad M \in \mathbf{N}.$$

Setting $M = l^2$, we conclude

$$e_{2C_{11}l^2}^{mc}(S_{\mu_0}, X) \leq C_{12} l^{-r/m-1}, \quad l \geq l_0,$$

from which the upper estimate in (3) follows. The proof of Theorem 1 is finished. \square

5. COMPLEXITY ANALYSIS

Besides the order of approximation for a given class of Monte Carlo methods, as expressed by the quantity $e_n^{mc}(S_\mu, X)$, another function is of interest, the MC-complexity

$$\varepsilon \rightarrow \text{MC-comp}(S_\mu, X, \varepsilon), \quad \varepsilon > 0,$$

describing the average cost of arithmetic operations needed to approximate S_μ with given Monte Carlo error $\leq \varepsilon$. We shall follow the outline in [20, p. 417]. Suppose we are given a function

$$\theta : \bigcup \{ \Phi(N, \mathbf{R}), N \in \mathcal{N}_{\text{st}}^n(X, \mathbf{R}^n), n \in \mathbf{N} \} \rightarrow \mathbf{R}^+ \cup \{+\infty\}$$

(the cost function for the algorithms). Let $c^* > 0$ be a fixed constant, and define the cost of a deterministic method $u \in \mathcal{M}(X, \mathbf{R})$ by

$$\text{cost}(u) := \inf \{ c^* n + \theta(\varphi), \varphi \in \Phi(N, \mathbf{R}), N \in \mathcal{N}_{\text{st}}^n(X, \mathbf{R}), n \in \mathbf{N}, u = \varphi \circ N \}.$$

Hence, we assume that we have to pay for the information N (where each component of N is supposed to cost c^*) and for the execution of the algorithm φ , both independent of $(k, f) \in X$. It follows that $\text{cost}(u) \geq c^* \text{card}(u)$.

We have to make one further assumption. Suppose we can produce a concrete “program” which uses the arithmetic operations $+$, $-$, $*$, $/$ and which computes $\varphi \in \Phi(N, \mathbf{R})$. To be more precise, suppose φ can be computed by a straight-line program as defined in [1, 11.2]. Then $\theta(\varphi)$ is assumed to be bounded from above by the number of steps executed by this program. For further details on models of computation and the respective cost functions, we refer to [1, 20]. To define the cost of a Monte Carlo method, it would be appropriate to average over the cost of the deterministic methods involved. However, we cannot guarantee the $[\mathcal{F}(X, \mathbf{R}), \mathcal{B}(\mathbf{R})]$ -measurability. But since cost is positive, we can define the MC-cost of a Monte Carlo method P via an upper integral

$$\text{MC-cost}(P) := \int_{\mathcal{M}(X, \mathbf{R})}^* \text{cost}(u) dP(u),$$

as done for the MC-cardinality in §2. The MC-complexity of the problem S_μ on the class X of problem elements is defined by

$$(31) \quad \text{MC-comp}(S_\mu, X, \varepsilon) := \inf\{\text{MC-cost}(P), e^{mc}(S_\mu, X, P) \leq \varepsilon\}.$$

Remark. The given definition of MC-complexity of random methods does not express how hard it is to realize the random process necessary to find $u \in \mathcal{M}(X, \mathbf{R})$. This may be justified if we think of a precomputational process, which provides realizations of all random variables needed to choose u . In particular, we assume that we can sample from any given distribution, e.g., μ_0 , the initial distribution of P_0 (compare also the remark below).

The above definition (31) of $\text{MC-comp}(S_\mu, X, \varepsilon)$, together with the assumptions made on cost , immediately imply

$$\begin{aligned} \text{MC-comp}(S_\mu, X, \varepsilon) &\geq C_{13} \inf\{\text{MC-card}(P), e^{mc}(S_\mu, X, P) \leq \varepsilon\} \\ &\geq C_{13} \inf\{n-1, e_n^{mc}(S_\mu, X) \leq \varepsilon\}, \end{aligned}$$

proving the lower bound in (4).

To prove the upper estimate, we analyze the given algorithm described in §4. Therefore, we shall briefly outline a verbal description of the stochastic method defined by (30). While the approximative aspects of this method have been handled there, we shall see below that all steps can be executed within the number of operations as claimed in Corollary 1.

First, we have to fix $l \in N$ and put $M := l^2$. Next, points $(t_{i,l})_{i=1}^{q(l)}$ in I^m , a family $(\varphi_{i,l})_{i=1}^{q(l)}$ of continuous functions, and mappings P_l, Q_l from Proposition 2 have to be chosen. Also, M realizations $\omega^1, \dots, \omega^M$ of the Markov chain $(X_i)_{i=1}^\infty$ can be precomputed, resulting in corresponding sample points.

Now, given $(k, f) \in X$, the following steps have to be performed:

1. Find $v_l(k, f)$ according to Proposition 3 in all points $(t_{i,l})_{i=1}^{q(l)}$. To do this within the complexity bound, a multilevel method, as, e.g., proposed by Emelyanov and Ilin [6], is required. Thus, we are able to compute $\tilde{u}_l(k, f) := \langle P_l v_l(k, f), \mu_0 \rangle$.

2. Compute $\xi(\omega^j, k, f)$ and $\xi(\omega^j, h, g)$, $j = 1, \dots, M$, where ξ is the standard Monte Carlo method defined in (13), and calculate the final approxi-

mation as

$$u(k, f) = \tilde{u}_l(k, f) + \frac{1}{M} \sum_{j=1}^M (\xi(\omega^j, k, f) - \xi(\omega^j, h, g))$$

according to (30). The computation of $\xi(\omega^j, k, f)$, $j = 1, \dots, M$, requires the evaluation of k and f at the sample points. To compute $\xi(\omega^j, h, g)$, $j = 1, \dots, M$, we have to evaluate $h := Q_l k$ and $g := P_l v_l(k, f) - T_h P_l v_l(k, f)$ in all points corresponding to $\omega^1, \dots, \omega^M$. (Observe that $P_l v_l(k, f)$ is the exact solution of the Fredholm equation $y(s) = \int_{I^m} h(s, t)y(t) dt + g(s)$, $s \in I^m$.)

The considerations below will prove that both steps above can be carried out. Step 1 is the deterministic part of the algorithm, while Step 2 forms the stochastic part.

To proceed, we shall count the required arithmetic operations. A look at Step 1 shows that we have to compute $\tilde{u}_l(k, f)$. By (29) and (19) we have

$$\tilde{u}_l(k, f) = \langle P_l v_l(k, f), \mu_0 \rangle = \sum_{i=1}^{q(l)} v(t_{i,l}) \langle \varphi_{i,l}, \mu_0 \rangle,$$

where v stands for $v_l(k, f)$. Since μ_0 as well as the functions $\varphi_{i,l}$ are known beforehand, the scalar products $\langle \varphi_{i,l}, \mu_0 \rangle$ can be precomputed. Moreover, it is known that there are $C_{14} > 0$ and v_l satisfying Proposition 3, for which the values $v_l(k, f)(t_{i,l})$ ($i = 1, \dots, q(l)$) can be computed in $C_{14}l^2$ arithmetic operations (see again [2, 6, 10]).

We turn to the randomized part, Step 2. It requires the computation of M realizations of $\xi((X_i)_{i=1}^\infty, k, f)$ and $\xi((X_i)_{i=1}^\infty, w_l(k, f))$, respectively. Suppose we are given a realization $(x_i)_{i=1}^\infty$ of $(X_i)_{i=1}^\infty$ with length $\mathbf{T} = \tau((x_i)_{i=1}^\infty)$. The representation (13) shows that the calculation of $\xi((x_i)_{i=1}^\infty, k, f)$ needs no more than $2\mathbf{T}$ arithmetic operations. To evaluate $\xi((x_i)_{i=1}^\infty, w_l(k, f))$, we need to know $h = Q_l k$ at $\mathbf{T} - 1$ points and $g = P_l v_l(k, f) - T_{Q_l k} P_l v_l(k, f)$ at another point. The representation (20) shows that the evaluation of $Q_l k$ at $\mathbf{T} - 1$ points requires the values of $\varphi_{i,l}(s)$ and $\varphi_{j,l}(t)$ at $\mathbf{T} - 1$ points. Since any given point (s, t) meets only a constant number of supports (see [18]), we only have to carry out a constant number of evaluations of $\varphi_{i,l}$ at s and $\varphi_{j,l}$ at t . In concrete cases, the functions $\varphi_{i,l}(s)$ and $\varphi_{j,l}(t)$ are piecewise polynomials of fixed degree, the pieces and the supports being sets of simple structure (see [5]). So it is clear that the evaluation of $\varphi_{i,l}(s)$ at one point is proportional to the degree, independent of l . This implies that the number of arithmetic operations needed to know $Q_l k$ at $\mathbf{T} - 1$ points requires arithmetic operations proportional to $\mathbf{T} - 1$.

The amount of work needed to evaluate g at one point t can be derived from the representation (26) given above,

$$\begin{aligned} g(t) &= \sum_{i=1}^{q(l)} v_l(t_{i,l}) \varphi_{i,l}(t) - \sum_{i,j,k=1}^{q(l)} k(t_{i,l}, t_{j,l}) v_l(t_{k,l}) \varphi_{i,l}(t) \langle \varphi_{j,l}, \varphi_{k,l} \rangle \\ &= \sum_{i=1}^{q(l)} \left\{ v_l(t_{i,l}) - \sum_{j,k=1}^{q(l)} k(t_{i,l}, t_{j,l}) v_l(t_{k,l}) \langle \varphi_{j,l}, \varphi_{k,l} \rangle \right\} \varphi_{i,l}(t). \end{aligned}$$

The terms in braces can be computed once for all t . Since $\varphi_{i,l}$ are known functions, we can think of $\langle \varphi_{j,l}, \varphi_{k,l} \rangle$ as being given real numbers. The condition (18) imposed on $\varphi_{i,l}$ shows that the sum over $j, k = 1, \dots, q(l)$, reduces to a sum over a number of summands proportional to $q(l)$. Hence, the computation of the terms in braces for $i = 1, \dots, q(l)$ requires a number of arithmetic operations proportional to $q(l)^2$. Once this is done, the computation of g at any given point t requires only a constant number of operations, since the sum over $i = 1, \dots, q(l)$ above reduces to one of constant length. Summarizing the arguments given before, and having the representation (30) in mind, we see that the computation of a realization of $u_{\omega_M}(k, f)$ with lengths $\mathbf{T}_1, \dots, \mathbf{T}_M$, respectively, requires a number of arithmetic operations bounded by a multiple of $l^2 + \sum_{i=1}^M \mathbf{T}_i + q(l)^2 + M$, which is proportional to $l^2 + \sum_{i=1}^M \mathbf{T}_i$. Since we assumed that the algorithmic cost θ can be estimated in this way, we obtain

$$\begin{aligned} \text{MC-cost}(P_{l,M}) &= \int^* \text{cost}(u) dP_{l,M}(u) \\ &\leq C_{15} \left\{ l^2 + \sum_{j=1}^M \int_{\Omega} \tau(\omega^j) dP_0^M(\omega_M) \right\} \\ &\leq C_{15} \{ l^2 + M \mathbf{E}\tau \} \leq C_{16} (l^2 + M). \end{aligned}$$

The choice of $M = l^2$, made to prove the upper estimate in (1), finally yields

$$\text{MC-comp}(S_{\mu_0}, X, \varepsilon) \leq C_{17} \cdot (1/\varepsilon)^{2m/(r+m)}, \quad \varepsilon > 0,$$

proving Corollary 1.

Remark. Although the model of computation for Monte Carlo methods, introduced above, does not deal with the problem of realizing the random variables needed in the computational process, some hints seem to be useful. A different, but from the point of view of numerical simulation more convenient way to introduce $(X_i)_{i=1}^\infty$ is as follows.

Let $(Z_i)_{i=1}^\infty$ be a sequence of independent random variables, each with uniform distribution on $[0, 1]$. Let τ be the random natural number, defined as $\tau = \min\{i, Z_i \geq p\}$, i.e.,

$$P(\tau = n) = p^{n-1}(1 - p), \quad n \geq 1.$$

Further, let $(Y_i)_{i=2}^\infty$ be independent identically distributed with $P_{Y_i} = \lambda$, where λ is the Lebesgue measure on I^m . Putting $X'_1 = X_1$ distributed according to μ_0 , and for $i \geq 2$

$$X'_i = \begin{cases} Y_i & \text{if } i \leq \tau, \\ \Delta & \text{if } i > \tau, \end{cases}$$

we can easily check that $P_{(X'_i)_{i=1}^\infty} = P_{(X_i)_{i=1}^\infty}$, thus providing the possibility of using a pseudo-random number generator for uniform distributions, readily available on most computers.

It seems worthwhile to make some remarks on the deterministic case of the problem under consideration. Define the error of a deterministic method $u \in \mathcal{M}(X, \mathbf{R})$ by

$$e(S_\mu, X, u) := \sup\{|S_\mu(k, f) - u(k, f)|, (k, f) \in X\}$$

and put

$$e_n(S_\mu, X) := \inf\{e(S_\mu, X, u), u \in \mathcal{M}^n(X, \mathbf{R})\}.$$

Then

$$(32) \quad cn^{-r/(2m)} \leq \sup_{\|\mu\| \leq 1} e_n(S_\mu, X) \leq Cn^{-r/(2m)}.$$

This was proved by Emelyanov and Ilin [6]. To draw the final conclusion, we see that for the optimal Monte Carlo method, both rates multiply—the standard Monte Carlo rate for general continuous data, $n^{-1/2}$, and the deterministic rate for r -smooth data, $n^{-r/(2m)}$. Statements analogous to (32), and the conclusions above, hold also for the complexity.

BIBLIOGRAPHY

1. A. V. Aho, J. E. Hopcroft, and J. D. Ullmann, *The design and analysis of computer algorithms*, Addison-Wesley, Reading, MA, 1974.
2. K. E. Atkinson, *A survey of numerical methods for the solution of Fredholm integral equations of the second kind*, SIAM, Philadelphia, 1976.
3. N. S. Bakhvalov, *On approximate computation of multiple integrals*, Vestnik Moskov. Univ. Ser. I Mat. Mekh. **4** (1959), 3–18. (Russian)
4. Yu. A. Brychkov, O. I. Marichev, and A. P. Prudnikov, *Integrals and series, Volume 1: Elementary functions*, Gordon & Breach, New York, London, and Paris, 1986.
5. P. G. Ciarlet, *The finite element method for elliptic problems*, North-Holland, Amsterdam, New York, and Oxford, 1978.
6. 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** (1967), 905–910; English transl. in USSR Comput. Math. and Math. Phys. **7** (1967), 259–267.
7. S. M. Ermakov, *The Monte Carlo method and related problems*, Nauka, Moscow, 1971. (Russian)
8. S. M. Ermakov and G. A. Michailov, *The theory of statistical trials*, 2nd ed., Nauka, Moscow, 1982. (Russian)
9. S. M. Ermakov, W. W. Nekrutkin, and A. S. Sipin, *Random processes for classical equations of mathematical physics*, Math. Appl. (Soviet Ser.), vol. 34, Kluwer, Dordrecht, 1989.
10. W. Hackbusch, *Multi-grid methods and applications*, Springer, Berlin, Heidelberg, and New York, 1985.
11. S. Heinrich, *On the optimal error of degenerate kernel methods*, J. Integral Equations **9** (1985), 251–266.
12. —, *Probabilistic analysis of numerical methods for integral equations*, J. Integral Equations Appl. **3** (1991), 289–319.
13. St. Kaczmarz and H. Steinhaus, *Theory of orthogonal series*, 2nd ed., Chelsea, New York, 1951.
14. M. Loève, *Probability theory*, 4th ed., Graduate Texts in Math., vol. 46, Springer, New York, 1978.
15. P. Mathé, *Random approximation of Sobolev embeddings*, J. Complexity **7** (1991), 261–281.
16. E. Novak, *Deterministic and stochastic error bounds in numerical analysis*, Springer, Berlin, Heidelberg, and New York, 1988.
17. S. V. Pereverzev, *On the optimization of adaptive methods for the approximate solution of integral equations*, Dokl. Akad. Nauk SSSR **267** (1982), 1304–1308; English transl. in Soviet Math. Dokl. **26** (1982), 770–774.
18. —, *On the complexity of the problem of finding solutions of Fredholm equations of the second kind with smooth kernels. I*, Ukrain. Mat. Zh. **40** (1988), 84–91; English transl. in Ukrainian Math. J. **40** (1988), 71–76.

19. —, *On the complexity of the problem of finding solutions of Fredholm equations of the second kind with smooth kernels. II*, Ukrain. Mat. Zh. **41** (1989), 189–193; English transl. in Ukrainian Math. J. **41** (1989), 169–173.
20. J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski, *Information-based complexity*, Academic Press, New York, 1988.

INSTITUT FÜR ANGEWANDTE ANALYSIS UND STOCHASTIK, MOHRENSTRASSE 39, PF 1304, D-O-1086 BERLIN, GERMANY