Interner Bericht

The Multilevel Method of Dependent Tests

Stefan Heinrich

300/99

FACHBEREICH
INFORMATIK

UNIVERSITÄT
KAISERSLAUTERN

Postfach 3049 · D-67653 Kaiserslautern

The Multilevel Method of Dependent Tests

Stefan Heinrich

300/99

Universität Kaiserslautern
Fachbereich Informatik
Postfach 30 49
D-67653 Kaiserslautern
Germany
email: heinrich@informatik.uni-kl.de

Januar 1999

(submitted to the Proceedings of the 3rd St. Petersburg Workshop on Simulation 1998, Birkhäuser, Basel)

The Multilevel Method of Dependent Tests

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

Abstract

Approximation properties of the underlying estimator are used to improve the efficiency of the method of dependent tests. A multilevel approximation procedure is developed such that in each level the number of samples is balanced with the level-dependent variance, resulting in a considerable reduction of the overall computational cost. The new technique is applied to the Monte Carlo estimation of integrals depending on a parameter.

1 Introduction

The method of dependent tests is a basic way of using Monte Carlo estimates for the approximation of whole functions (as opposed to the approximation of a single function value or a weighted integral as in the classical Monte Carlo approach). The method was developed and studied by Frolov and Chentsov (1962), Sobol (1962, 1973), Ermakov and Mikhailov (1982), Mikhailov (1991), Voytishek (1996, 1997), Prigarin (1995), and others. The aim of the present paper is to propose a multilevel version of this method, based on the ideas developed in Heinrich (1998a). We exploit the approximability of the underlying estimator to decompose it into levels. The number of samples used for each level can be tuned to the variance of the contribution from this level, so that an overall reduction of computational cost is reached. The new method is presented in a general framework, and later on studied in detail for the computation of integrals depending on a parameter. While Heinrich (1998a) and Heinrich and Sindambiwe (1999) consider the class C^r of r-times continuously differentiable functions on the unit cube and provide, besides the algorithm, also a complexity analysis including lower bounds, in this paper we mostly concentrate on the study of the algorithm and give a more general convergence analysis. We consider arbitrary domains instead of the cube and relaxed smoothness assumptions requiring the function to be in a Sobolev class W_p^r with $1 < p < \infty$, and this only with respect to the parameter variable. We study the expected norm error in L_p for $1 < p < \infty$ and we are able to give quantitative results on the convergence. This is new even for the standard, one-level method of dependent tests. For that one, such an approach has so far only been carried out for Hilbert spaces, see, e.g., Mikhailov (1991), Voytishek (1996), Prigarin (1995). For L_p spaces ($p > 2$) only asymptotic results were obtained on the basis of weak convergence, with no information on the speed of convergence (to the Gaussian limit), see, e.g., Frolov and Chentsov (1962), Ermakov and Mikhailov (1982), Prigarin (1995), Voytishek (1997). Finally, the case $1 < p < 2$ is often left out because the usual tools do not work — the involved functions have infinite variance. We are able to study this case too and determine the convergence rates for both one- and multilevel methods. In the end, a few remarks on lower bounds and a comparison between one- and multilevel methods are made. The present paper is an extended version of a note which appeared in the abstract volume of this conference (see Heinrich, 1998b). A multilevel approach with the possibility of independent sampling is developed in Heinrich (1998c).

2 The Standard Method of Dependent Tests

Let X be a Banach space. A random variable with values in X is a Borel measurable mapping $\eta : \Omega \to X$ on some probability space (Ω, Σ, μ) such that the values of η are almost surely contained in a separable subspace of X. For $1 \le p < \infty$ we denote by $L_p(X) = L_p(\Omega, \Sigma, \mu, X)$ the space of all X-valued random variables η on (Ω, Σ, μ) satisfying

$$\mathbb{E}\|\eta\|^p = \int_\Omega \|\eta(\omega)\|^p \, d\mu(\omega) < \infty$$

(see Ledoux and Talagrand, 1991, for details).

Now let $\eta \in L_p(X)$ for some p with $1 \le p < \infty$. We seek to approximate the expectation

$$u = \mathbb{E}\eta \in X.$$

Usually, X is an infinite dimensional function space, which makes it, in general, impossible to compute u itself. Instead, an estimate for Pu is constructed, where P is some continuous linear finite rank operator (an interpolation or approximation operator, for example), acting from X to another Banach space Y (as a rule, either X itself or a larger function space, compare section 4). We shall assume that X is continuously embedded into Y, that is, there is a continuous injection $J : X \to Y$. In the sequel we shall identify X with $J(X) \subseteq Y$ as sets. The norms will be distinguished by $\| \|_X$ and $\| \|_Y$. Let

$$Px = \sum_{i=1}^{n} \langle x, x_i^* \rangle \, y_i \qquad (x \in X) \tag{1}$$

be a representation of P, where $x_i^* \in X^*$ (the dual of X) and $y_i \in Y$. The standard method of dependent tests consists of the estimate

$$Pu \approx \theta = \frac{1}{N} \sum_{j=1}^{N} P\eta_j, \tag{2}$$

where $(\eta_j)_{j=1}^N$ are independent realizations of η. (We assume that all random variables considered in this paper are defined on the same basic probability space (Ω, Σ, μ).) Combined with (1), this gives

$$\sum_{i=1}^{n} \langle u, x_i^* \rangle \, y_i \approx \theta = \sum_{i=1}^{n} \left(\frac{1}{N} \sum_{j=1}^{N} \langle \eta_j, x_i^* \rangle \right) y_i \tag{3}$$

(which makes it clear that we use the same N samples for the estimation of the whole family of functionals $\langle u, x_i^* \rangle$). As an illustration, we consider integrals depending on a parameter, which will be studied in detail later on. For the moment we do this on an informal level — precise assumptions follow in section 4.

Let $G_1 \subset \mathbb{R}^{d_1}, G_2 \subset \mathbb{R}^{d_2}$ and f be a function on $G_1 \times G_2$. We want to approximate

$$u(s) = \int_{G_2} f(s,t) \, dt$$

as a function of the parameter $s \in G_1$. Let X and Y with $X \subseteq Y$ be some spaces of functions on G_1 and let P be an operator acting on $g \in X$ as

$$(Pg)(s) = \sum_{i=1}^{n} g(s_i)\varphi_i(s)$$

(one can think, e.g., of piecewise linear interpolation). Let ξ be a uniformly distributed on G_2 random variable. We set

$$\eta(\omega) = |G_2| f(\,\cdot\,, \xi(\omega)).$$

Then

$$\mathbb{E}\eta = \int_{G_2} f(\,\cdot\,, t)\, dt = u.$$

Now the method of dependent tests approximates

$$Pu \approx \sum_{i=1}^{n} \left(\frac{|G_2|}{N} \sum_{j=1}^{N} f(s_i, \xi_j) \right) \varphi_i$$

with $(\xi_j)_{j=1}^{N}$ being independent realizations of ξ.

3 The Multilevel Approach

Assume that we are given a sequence of continuous linear finite rank operators $(P_\ell)_{\ell=1}^{m}$ from X to Y with $P_m = P$ instead of P alone (usually, the approximation operators P belong to such scales in a natural way). Let

$$P_\ell x = \sum_{i=1}^{n_\ell} \langle x, x_{\ell i}^* \rangle\, y_{\ell i} \qquad (x \in X)$$

$(\ell = 1, \ldots, m)$ be the respective representations. Choose positive integers $(N_\ell)_{\ell=1}^{m}$ and estimate

$$P_m u = \sum_{\ell=1}^{m} (P_\ell - P_{\ell-1}) u \tag{4}$$

(with $P_0 = 0$) by

$$\begin{aligned}\zeta &= \sum_{\ell=1}^{m} \frac{1}{N_\ell} \sum_{j=1}^{N_\ell} (P_\ell - P_{\ell-1})\eta_{\ell j} && (5)\\ &= \sum_{\ell=1}^{m} \left[\sum_{i=1}^{n_\ell} \left(\frac{1}{N_\ell} \sum_{j=1}^{N_\ell} \langle \eta_{\ell j}, x_{\ell i}^* \rangle \right) y_{\ell i} - \sum_{i=1}^{n_{\ell-1}} \left(\frac{1}{N_\ell} \sum_{j=1}^{N_\ell} \left\langle \eta_{\ell j}, x_{\ell-1,i}^* \right\rangle \right) y_{\ell-1,i} \right]\end{aligned}$$

where $(\eta_{\ell j} : j = 1, \dots, N_\ell,\ \ell = 1, \dots, m)$ are independent realizations of η. We set $n_0 = 0$, so for $\ell = 1$ the second term of the last line of (5) is to be understood as zero. Observe that the standard (one-level) method corresponds to the case $m = 1$ and $N_1 = N$. For parametric integration the concrete form of (5) is given later on — see relation (15). Now we shall analyze the error. For $1 \le p < \infty$ we define the p-th expected norm error of the estimate ζ as

$$e_p(\zeta) = (\mathbb{E}\|u - \zeta\|_Y^p)^{1/p}.$$

By the triangle inequality, $e_p(\zeta)$ can be bounded by a deterministic and a stochastic component:

$$\begin{aligned}e_p(\zeta) &= (\mathbb{E}\|u - P_m u + P_m u - \zeta\|_Y^p)^{1/p}\\ &\le \|u - P_m u\|_Y + (\mathbb{E}\|P_m u - \zeta\|_Y^p)^{1/p}. && (6)\end{aligned}$$

Next we shall give an upper bound for the stochastic component. For this purpose we let $1 \le p \le 2$ and recall that a Banach space Z is said to be of type p if there is a constant $c > 0$ such that for all $n \in \mathbb{N}$ and $(z_i)_{i=1}^n \subset Z$,

$$(\mathbb{E}\|\sum_{i=1}^{n} \varepsilon_i z_i\|^p)^{1/p} \le c \left(\sum_{i=1}^{n} \|z_i\|^p \right)^{1/p}, \tag{7}$$

where $(\varepsilon_i)_{i=1}^n$ is a sequence of independent Bernoulli variables with $\mu\{\varepsilon_i = 1\} = \mu\{\varepsilon_i = -1\} = \frac{1}{2}$. We refer to ch. 9.2 of Ledoux and Talagrand (1991) for this definition and background. The smallest possible constant in (7) is called the type p constant of Z, denoted by $T_p(Z)$. Let us mention that every Banach space is of type 1 (triangle inequality), and that type p implies type q for $1 \le q < p$. Each finite dimensional space is of type 2, and for $1 \le p < \infty$ the spaces $L_p(\nu)$ (with ν an arbitrary measure) are of type $\min(p, 2)$. Clearly, all subspaces U of a type p space Z are of type p themselves, with $T_p(U) \le T_p(Z)$. Now we can present a bound of the stochastic part of the error.

Proposition 1 *Let $1 < p \leq 2$ and assume $\eta \in L_p(X)$. Then*

$$(\mathbb{E}\|P_m u - \zeta\|_Y^p)^{1/p} \leq 2T_p(Y_m) \left(\sum_{\ell=1}^{m} N_\ell^{1-p} \mathbb{E}\|(P_\ell - P_{\ell-1})(u - \eta)\|_Y^p \right)^{1/p},$$

where $Y_m = \mathrm{span}\,(\bigcup_{\ell=1}^m P_\ell(X)) \subset Y$.

PROOF. By (4) and the first part of (5) we have

$$P_m u - \zeta = \sum_{\ell=1}^{m} \frac{1}{N_\ell} \sum_{j=1}^{N_\ell} (P_\ell - P_{\ell-1})(u - \eta_{\ell j}). \tag{8}$$

Now put

$$\rho_{\ell j} = N_\ell^{-1}(P_\ell - P_{\ell-1})(u - \eta_{\ell j})$$

$(\ell = 1, \dots, m,\ j = 1, \dots, N_\ell)$. These are independent Y_m-valued mean zero random variables with finite p-th moment

$$\mathbb{E}\|\rho_{\ell j}\|_Y^p = N_\ell^{-p} \mathbb{E}\|(P_\ell - P_{\ell-1})(u - \eta)\|_Y^p. \tag{9}$$

Proposition 9.11 of Ledoux and Talagrand (1991) states that

$$\mathbb{E}\|\sum_{\ell=1}^{m} \sum_{j=1}^{N_\ell} \rho_{\ell j}\|_Y^p \leq (2T_p(Y_m))^p \sum_{\ell=1}^{m} \sum_{j=1}^{N_\ell} \mathbb{E}\|\rho_{\ell j}\|_Y^p.$$

Combining this with (8) and (9) yields the result. ■

Corollary 2 *Let $1 < p \leq 2$ and assume that $\eta \in L_p(X)$. Then*

$$(\mathbb{E}\|P_m u - \zeta\|_Y^p)^{1/p} \leq 2T_p(Y_m)(\mathbb{E}\|u - \eta\|_X^p \sum_{\ell=1}^{m} N_\ell^{1-p} \|P_\ell - P_{\ell-1} : X \to Y\|^p)^{1/p}.$$

PROOF. This follows directly from

$$\mathbb{E}\|(P_\ell - P_{\ell-1})(u - \eta)\|_Y^p \leq \|P_\ell - P_{\ell-1} : X \to Y\|^p \mathbb{E}\|u - \eta\|_X^p.$$

■

4 Integrals Depending on a Parameter

Let d_1 and d_2 be positive integers and $G_1 \subset \mathbb{R}^{d_1}$ and $G_2 \subset \mathbb{R}^{d_2}$ be bounded open sets with Lipschitz boundary. Let $1 \le q < \infty$, let r be a positive integer with $r/d_1 > 1/q$, and let $W_r^{q,0}(G_1 \times G_2)$ be the space of all $f \in L_q(G_1 \times G_2)$ such that for each multiindex $\alpha = (\alpha_1, \dots, \alpha_{d_1})$ with $|\alpha| = \alpha_1 + \ldots + \alpha_{d_1} \le r$ the generalized derivative $D_1^\alpha f$ with respect to the G_1 coordinates exists and belongs to $L_q(G_1 \times G_2)$. Hence, somewhat loosely speaking, we consider functions $f(s,t)$ with smoothness $W_q^r(G_1)$ in the first variable $s \in G_1$ and summability $L_q(G_2)$ in the second variable $t \in G_2$. The norm on $W_q^{r,0}(G_1 \times G_2)$ is defined as

$$\|f\|_{W_q^{r,0}} = \left(\sum_{|\alpha| \le r} \|D_1^\alpha f\|^q_{L_q(G_1 \times G_2)} \right)^{1/q} .$$

(For all notation concerning Sobolev spaces we refer to Adams, 1975). We study the estimation of

$$u(s) = \int_{G_2} f(s,t) \, dt \tag{10}$$

in $L_q(G_1)$, that is, integration over G_2 with parameter domain G_1 and the error measured in the norm of $L_q(G_1)$. To put this into the framework of sections 2 and 3 we set $X = W_q^r(G_1)$, $Y = L_q(G_1)$ and $p = \min(2,q)$. We let $\xi = \xi(\omega)$ be a uniformly distributed on G_2 random variable on (Ω, Σ, μ) and we define $\eta = \eta(\omega)$ by

$$\eta : \omega \to |G_2| f(\,\cdot\,, \xi(\omega)).$$

Lemma 3 *The function η is a random variable with values in $X = W_q^r(G_1)$, belongs to $L_q(X)$, $\mathbb{E}\eta = u$ and*

$$(\mathbb{E}\|\eta\|_X^q)^{1/q} \le \|f\|_{W_q^{r,0}}. \tag{11}$$

PROOF. We first verify that the values of η almost surely belong to $W_q^r(G_1)$ and that η is Borel measurable as a mapping into $W_q^r(G_1)$ (note that $W_q^r(G_1)$ is a separable Banach space). Let us denote by f_t the function given by $f_t(s) = f(s,t)$. Then $f_t \in L_q(G_1)$ for almost all t, by Fubini's theorem. Using elementary facts from distribution theory, it is readily checked that for all α with $|\alpha| \le r$ the weak derivative $(D_1^\alpha f)(\,\cdot\,, t)$ coincides with $D^\alpha f_t$ for almost all t. This implies $f_t \in W_q^r(G_1)$ for almost all t. Moreover, since

$W_q^r(G_1)$ is isometric to the subspace of $\oplus_q \sum_{|\alpha|\le r} L_q(G_1)$ of those $(f_\alpha)_{|\alpha|\le r}$ with $f_\alpha = D^\alpha f_0$, we can use 6.2.12 of Pietsch (1987) to prove that $t \to (D^\alpha f_t)_{|\alpha|\le r}$ is Borel measurable as a mapping into the direct sum above, and hence $t \to f_t$ is Borel measurable as a mapping into $W_q^r(G_1)$. We have

$$(\mathbb{E}\|\eta\|_X^q)^{1/q} = \left(\int_{G_2} \|f_t\|_{W_q^r(G_1)}^q \, dt\right)^{1/q} = \|f\|_{W_q^{r,0}} \le 1.$$

It follows that $\mathbb{E}\eta$ is well-defined, and because of (10), equals u. ■

Now we have to define suitable approximation tools in $W_q^r(G_1)$. There is a vast literature on this subject and a variety of possibilities. Since a review of these tools is not the subject of this paper, we restrict ourselves to formulating the requirements on the approximating operators needed for our purposes and make a few comments on how to satisfy them. Let $P_\ell : W_q^r(G_1) \to L_q(G_1)$ $(\ell = 1, 2 \dots)$ be a sequence of operators of the form

$$P_\ell f = \sum_{i=1}^{n_\ell} f(s_{\ell i}) \varphi_{\ell i} \tag{12}$$

with $s_{\ell i} \in \overline{G}_1$ (the closure of G_1) and $\varphi_{\ell i} \in L_q(G_1)$. The Sobolev embedding theorem guarantees that the point evaluations are well-defined. We assume that there are constants $c_1, c_2, c_3 > 0$ such that for all ℓ

$$c_1 2^{d_1 \ell} \le n_\ell \le c_2 2^{d_1 \ell} \tag{13}$$

and, if $I_{r,q}$ denotes the identical embedding of $W_q^r(G)$ into $L_q(G_1)$,

$$\|I_{r,q} - P_\ell : W_q^r(G_1) \to L_q(G_1)\| \le c_3 2^{-r\ell}. \tag{14}$$

Such sequences can be constructed for many domains, e.g. by using triangular, rectangular or isoparametric finite elements of suitable order. We refer to Ciarlet (1978) for details. For the unit cube and arbitrary r, piecewise multivariate Lagrange interpolation will do (among many others), as described in Heinrich (1998a) and Heinrich and Sindambiwe (1999). For polyhedral domains and $r = 2$ piecewise linear interpolation on successively finer triangulations is a standard approach, the $s_{\ell i}$ being the vertices of the triangles and the $\varphi_{\ell i}$ being the corresponding hat functions.

The restriction to point evaluations of f in (12) was just made for notational simplicity. One could also admit values of derivatives $(D^\alpha f)(s)$ with

$|\alpha| < r/d_1 - 1/q$. Now the multilevel method of dependent test fixes an m and approximates u according to (5) by

$$\zeta = \sum_{\ell=1}^{m} \left[\sum_{i=1}^{n_\ell} \left(\frac{|G_2|}{N_\ell} \sum_{j=1}^{N_\ell} f(s_{\ell i}, \xi_{\ell j}) \right) \varphi_{\ell i} - \sum_{i=1}^{n_{\ell-1}} \left(\frac{|G_2|}{N_\ell} \sum_{j=1}^{N_\ell} f(s_{\ell-1,i}, \xi_{\ell j}) \right) \varphi_{\ell-1,i} \right] \tag{15}$$

where the N_ℓ ($\ell = 1, \dots, m$) are positive integers, $\xi_{\ell j}$ ($\ell = 1, \dots, m$, $j = 1, \dots, N_\ell$) are independent realizations of the uniformly distributed on G_2 random variable ξ, and for $\ell = 1$ we set $n_0 = 0$, so in this case the second term is zero. Next we provide a bound for the stochastic part of the error of the multilevel method.

Proposition 4 *Let $1 < q < \infty$, $p = \min(2, q)$ and let $(P_\ell)_{\ell=1}^{\infty}$ satisfy (13) and (14). Then there is a constant $c > 0$ such that for all $f \in W_q^{r,0}(G_1 \times G_2)$ with $\|f\|_{W_q^{r,0}} \le 1$, for all $m \in \mathbb{N}$ and $N_\ell \in \mathbb{N}$ ($\ell = 1, \dots, m$) the multilevel estimate ζ defined above satisfies.*

$$(\mathbb{E}\|P_m u - \zeta\|_{L_q(G_1)}^p)^{1/p} \le c \left(\sum_{\ell=1}^{m} N_\ell^{1-p} 2^{-rp\ell} \right)^{1/p}. \tag{16}$$

PROOF. From Lemma 3 we get $\eta \in L_q(X) \subseteq L_p(X)$ and

$$\begin{aligned} (\mathbb{E}\|u - \eta\|_X^p)^{1/p} &\le \|u\|_X + (\mathbb{E}\|\eta\|_X^p)^{1/p} \\ &= \|\mathbb{E}\eta\|_X + (\mathbb{E}\|\eta\|_X^p)^{1/p} \\ &\le 2(\mathbb{E}\|\eta\|_X^q)^{1/q} \le 2\|f\|_{W_q^{r,0}} \le 2. \end{aligned}$$

Moreover, by (14)

$$\|P_\ell - P_{\ell-1} : X \to Y\| \le c 2^{-r\ell}.$$

Finally, since $L_q(G_1)$ is of type p, all of its subspaces have a type p constant not exceeding that of $L_q(G_1)$. Now Corollary 2 yields the result. ■

Remark. For the one level method (2) we obtain under the assumptions of Proposition 4 with $P = P_k$ for some $k \ge 1$,

$$(\mathbb{E}\|Pu - \theta\|_{L_q(G_1)}^p)^{1/p} \le c N^{1/p-1}. \tag{17}$$

For $q = 2$ $(= p)$ this is well-known (see, e.g.,Voytishek, 1996).

In the following we shall choose the N_ℓ and balance deterministic and stochastic error in such a way that we obtain minimal error at fixed computational cost. In the theorem below 'cost of computing ξ' means the total number of arithmetic operations, random number calls and function evaluations required for the computation of the coefficients of all $\varphi_{\ell i}$ in (15).

Theorem 5 *Let $1 < q < \infty$, $p = \min(2, q)$ and let $(P_\ell)_{\ell=1}^\infty$ satisfy (13) and (14). Then there exist constants $c_1, c_2 > 0$ such that for each integer $M > 1$ there is a choice of parameters $m, (N_\ell)_{\ell=1}^m$ such that the cost of computing ζ is bounded by $c_1 M$ and for each $f \in W_q^{r,0}(G_1 \times G_2)$ with $\|f\|_{W_q^{r,0}} \le 1$ the p-th expected norm error (with respect to the norm of $L_q(G_1)$) satisfies*

$$\begin{array}{lcll} e_p(\zeta) & \le & c_2 M^{-r/d_1} & \text{if} \quad r/d_1 < 1 - 1/p, \\ e_p(\zeta) & \le & c_2 M^{1/p-1} \log M & \text{if} \quad r/d_1 = 1 - 1/p, \\ e_p(\zeta) & \le & c_2 M^{1/p-1} & \text{if} \quad r/d_1 > 1 - 1/p. \end{array}$$

PROOF. Throughout the proof and in the sequel the same symbol c, c_1, or c_2 is used for possibly different positive constants, not depending on m, M and f. The cost of computing ζ is obviously bounded by

$$c \sum_{\ell=1}^{m} 2^{d_1 \ell} N_\ell. \tag{18}$$

The line of the subsequent proof is the following. For the moment we fix m to be any positive integer with

$$2^{d_1(m-1)} \le M. \tag{19}$$

First we choose the N_ℓ for this fixed m and estimate the stochastic part of the error. Later on we select the final m so that deterministic and stochastic part of the error are in balance.
So let

$$N_\ell = \left\lceil 2^{-(r+d_1/p)\ell - ((1-1/p)d_1 - r)m} M \right\rceil \tag{20}$$

if $r/d_1 < 1 - 1/p$,

$$N_\ell = \left\lceil m^{-1} 2^{-d_1 \ell} M \right\rceil \tag{21}$$

if $r/d_1 = 1 - 1/p$, and

$$N_\ell = \left\lceil 2^{-(r+d_1/p)\ell} M \right\rceil \tag{22}$$

if $r/d_1 > 1 - 1/p$. Although this choice looks complicated, it has an obvious source — this is (up to constants) what we get when minimizing the bound from Proposition 4,

$$\sum_{\ell=1}^{m} N_\ell^{1-p} 2^{-rp\ell}$$

(the variance, for $p = 2$), subject to the condition

$$\sum_{\ell=1}^{m} 2^{d_1 \ell} N_\ell \leq M$$

(the cost). Since this aspect is not relevant for the proof, we omit the standard calculation. It is readily checked that from (19) and the choices (20), (21) or (22) it follows that

$$\sum_{\ell=1}^{m} 2^{d_1 \ell} N_\ell \leq cM.$$

The deterministic part of the error (see (6)) satisfies, by (14) and (11),

$$\|u - P_m u\|_{L_q(G_1)} \leq c 2^{-rm} \|u\|_{W_q^r(G_1)} \leq c 2^{-rm}. \tag{23}$$

Next we compute the bounds on the stochastic part of the error in Proposition 4. First we treat the case $r/d_1 < 1 - 1/p$. We have

$$\left(\sum_{\ell=1}^{m} N_\ell^{1-p} 2^{-rp\ell} \right)^{1/p} \leq c M^{1/p-1} 2^{((1-1/p)d_1 - r)m}, \tag{24}$$

which is a consequence of (20) and the following calculation of exponents

$$\begin{aligned}
&(1-p)[-(r + d_1/p)\ell - ((1-1/p)d_1 - r)m] - rp\ell \\
&\quad = ((p-1)(r + d_1/p) - rp)\ell + (p-1)((1-1/p)d_1 - r)m \\
&\quad = ((1-1/p)d_1 - r)\ell + (p-1)((1-1/p)d_1 - r)m \\
&\quad = p((1-1/p)d_1 - r)m + ((1-1/p)d_1 - r)(\ell - m).
\end{aligned}$$

Now we choose m in such a way that

$$c_1 2^{-rm} \leq M^{1/p-1} 2^{((1-1/p)d_1 - r)m} \leq c_2 2^{-rm}, \tag{25}$$

which means that, up to constants, we equalize the bounds for deterministic and stochastic part of the error, that is, the right hand sides of (23) and (24). Clearly, (25) is equivalent to

$$c_1 2^{d_1 m} \leq M \leq c_2 2^{d_1 m}$$

(different constants!), and it suffices to take m to be the largest integer satisfying

$$2^{d_1(m-1)} \leq M.$$

(23), (24), (25) together with (16) and (6) yield

$$e_p(\zeta) \leq cM^{-r/d_1}.$$

For $r/d_1 = 1 - 1/p$ we use (21) and argue similarly:

$$(1-p)(-d_1\ell) - rp\ell = p((1-1/p)d_1 - r)\ell = 0$$

and hence

$$\left(\sum_{\ell=1}^{m} N_\ell^{1-p} 2^{-rp\ell}\right)^{1/p} \leq cM^{1/p-1}m.$$

We choose m in such a way that

$$c_1 2^{-rm} \leq M^{1/p-1}m \leq c_2 2^{-rm}.$$

This is equivalent to

$$c_1 m^{1/(1-1/p)} 2^{d_1 m} \leq M \leq c_2 m^{1/(1-1/p)} 2^{d_1 m},$$

and we let m be the largest integer satisfying

$$m^{1/(1-1/p)} 2^{d_1(m-1)} \leq M.$$

We obtain

$$e_p(\zeta) \leq cM^{1/p-1} \log M.$$

Finally, for $r/d_1 > 1 - 1/p$ we have

$$(1-p)(-(r + d_1/p)\ell) - rp\ell = ((1-1/p)d_1 - r)\ell$$

and hence

$$\left(\sum_{\ell=1}^{m} N_\ell^{1-p} 2^{-rp\ell}\right)^{1/p} \leq cM^{1/p-1}.$$

Here we choose m so that

$$c_1 2^{-rm} \leq M^{1/p-1} \leq c_2 2^{-rm},$$

or equivalently

$$c_1 2^{r(1-1/p)^{-1}m} \le M \le c_2 2^{r(1-1/p)^{-1}m},$$

and we let m be the largest integer with

$$2^{r(1-1/p)^{-1}(m-1)} \le M.$$

This yields

$$e_p(\zeta) \le cM^{1/p-1}$$

and proves the theorem. ■

Remark. As already mentioned, by 'computation of ζ' we meant the computation of coefficients of the functions $\varphi_{\ell i}$ in (15). Having accomplished this, it is often possible to combine these functions in a computationally favorable way. Usually, the spaces $\operatorname{span}\{\varphi_{\ell i} : i = 1, \ldots, n_\ell\}$ are nested, and one can decompose $\varphi_{\ell i}$ successively into combinations of $\varphi_{\ell+1,i}$ until level m is reached. For standard choices of approximation (as e.g. finite elements, piecewise Lagrange polynomials, piecewise linear functions, mentioned above) such a decomposition can be achieved in $cn_m \le cM$ operations.

Now assume this is done, as well, and we want to compute $\zeta(s)$ for many $s \in G_1$ (e.g., to produce a graph of the approximating function). For each s, this can be carried out in $\le c$ operations, provided the functions $\varphi_{m,i}$ can be computed in $\le c$ operations and the supports of these functions are almost disjoint, which means that

$$\sup_m \max_i |\{j : \operatorname{supp}(\varphi_{m,i}) \cap \operatorname{supp}(\varphi_{m,j}) \ne \emptyset| < \infty.$$

Again, many known approximation scales, including the above mentioned examples, possess this property.

Let us finally consider the one-level method and make comparisons. The sum of deterministic and stochastic error (see (17)) amounts to

$$c(2^{-rk} + N^{1/p-1}),$$

while the cost M is of the order $2^{d_1 k}N$. Equalizing both terms above, we see that at cost M we can reach an error

$$cM^{\frac{1/p-1}{1+(1-1/p)d_1/r}}. \tag{26}$$

This is certainly larger than

$$M^{\max(1/p-1,-rd_1)},$$

which we get (up to the log term) from Theorem 5. The saving by the multilevel method can be seen better if we compare the cost of reaching an error $\varepsilon > 0$. For the one-level method the cost is

$$c\left(\frac{1}{\varepsilon}\right)^{d_1/r+(1-1/p)^{-1}},$$

while for the multilevel method (up to log's)

$$c\left(\frac{1}{\varepsilon}\right)^{\max(d_1/r,(1-1/p)^{-1})}.$$

The results of Theorem 5 are optimal in a very general sense: No randomized algorithm of cost M can do better (except for a constant factor independent of M or, perhaps, a log-term in the case $r/d_1 = 1 - 1/p$). We cannot give the required formal framework for such statements here and refer instead to the literature on information-based complexity theory (see Traub, Wasilkowski, and Woźniakowski, 1988, Novak, 1988, Heinrich, 1994, Heinrich and Sindambiwe, 1999). Nevertheless, a few words on these lower bounds seem appropriate. First of all, we now restrict ourselves to the model case $G_1 = [0,1]^{d_1}$, $G_2 = [0,1]^{d_2}$. For the problem of parametric integration of functions from the class $W_q^{r,0}(G_1 \times G_2)$ lower bounds are, in fact, easily derived from known results (quite in contrast to the situation of $C^r(G_1 \times G_2)$ studied in Heinrich and Sindambiwe, 1999). Indeed, by considering the subclass of $W_q^{r,0}(G_1 \times G_2)$ of functions depending only on the second component, i.e. $f(s,t) \equiv g(t)$, we see that the problem is no easier than stochastic integration of $L_q(G_2)$ functions. For this, the lower bound $M^{1/p-1}$ with $p = \min(2,q)$ is known, see Novak (1988, 2.2.9, Proposition 1, and references). Similarly, the subclass of all functions in $W_q^{r,0}(G_1 \times G_2)$ depending only on the first component $f(s,t) \equiv g(s)$, can be identified with $W_q^r(G_1)$, hence the problem is no easier than approximation of functions of $W_q^r(G_1)$ in $L_q(G_1)$, for which the known lower bound for stochastic methods is M^{-r/d_1}, see Heinrich (1994, Thm. 6.1 and references). Thus

$$M^{\max(1/p-1,-r/d_1)} \tag{27}$$

is a lower bound, which shows that Theorem 5 yields, in fact, the optimal rate and hence the minimal Monte Carlo error in the sense of information-based complexity theory (up to a possible log factor in the case $r/d_1 = 1 - 1/p$).

When comparing the one- and the multilevel method, it seemed that we compared only upper bounds. Such a discussion would be meaningless since it does not exclude the existence of better estimates for any of the methods under comparison. In our case, however, this is not so. Looking again at functions depending only on the first or the second variable, it is easy to check directly that the one level method cannot be better than

$$c\left(2^{-rk} + N^{1/p-1}\right),$$

and hence, the rate (26) is sharp. Let us finally mention that the lower bound (27) also holds for $q = p = 1$, in which case it turns into a positive constant. This shows that no method can have a nontrivial convergence rate for $q = 1$.

References

1. Adams, R. A. (1975). *Sobolev Spaces*, New York: Academic Press.

2. Ciarlet, P. G. (1978). *The Finite Element Method for Elliptic Problems*, Amsterdam: North-Holland.

3. Ermakov, S. M. and Mikhailov, G. A. (1982). *Statistical Modelling*, Moscow: Nauka (in Russian).

4. Frolov, A. S. and Chentsov, N. N. (1962). On the calculation of certain integrals dependent on a parameter by the Monte Carlo method, *Zh. Vychisl. Mat. Mat. Fiz.*, **2**, 714 – 717 (in Russian).

5. Heinrich, S. (1994). Random approximation in numerical analysis, In: Bierstedt, K. D., Pietsch, A., Ruess, W. M. and Vogt, D., editors, *Functional Analysis*, pp. 123 – 171, New York, Basel, Hong Kong: Marcel Dekker.

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

7. Heinrich, S. (1998b). A multilevel version of the method of dependent tests, In: Ermakov, S. M., Kashtanov, Y. N. and Melas, V. B., editors, *Proceedings of the 3rd St. Petersburg Workshop on Simulation* (collection of extended abstracts), pp. 31 – 35, St. Petersburg: St. Petersburg University Press.

8. Heinrich, S. (1998c). Wavelet Monte Carlo methods for the global solution of integral equations, In: Niederreiter, H. and Spanier, J., editors, *Proceedings of the Third International Conference on Monte Carlo and Quasi-Monte Carlo Methods, Claremont, 1998* (submitted).

9. Heinrich, S. and Sindambiwe, E. (1999). Monte Carlo complexity of parametric integration, *J. Complexity*, (to appear).

10. Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces*, Berlin, Heidelberg, New York: Springer.

11. Mikhailov, G. A. (1991). *Minimization of Computational Costs of Non-Analogue Monte Carlo Methods*, Singapore: World Scientific.

12. Novak, E. (1988). *Deterministic and Stochastic Error Bounds in Numerical Analysis*, Lecture Notes in Mathematics 1349. Berlin, Heidelberg, New York: Springer.

13. Pietsch, A. (1987). *Eigenvalues and s-numbers*, Leipzig: Geest and Portig, and Cambridge: Cambridge University Press.

14. Prigarin, S. M. (1995). Convergence and optimization of functional estimates in statistical modelling in Sobolev's Hilbert spaces, *Russian J. Numer. Anal. Math. Modelling*, **10**, 325 - 346.

15. Sobol, I. M. (1962). The use of ω^2-distribution for error estimation in the calculation of integrals by the Monte Carlo method, *U.S.S.R. Comput. Math. and Math. Phys.*, **2**, 717 - 723.

16. Sobol, I. M. (1973). *Computational Monte Carlo Methods*, Moscow: Nauka (in Russian).

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

18. Voytishek, A. V. (1996). Discrete-stochastic procedures for the global estimation of an integral which depends on a paramter, *Comp. Maths Math. Phys.*, **36**, 997 - 1009.

19. Voytishek, A. V. (1997). Using the Strang-Fix approximation in discrete-stochastic numerical procedures, *Monte Carlo Methods and Appl.*, **3**, 89 - 112.