REDUCING VARIANCE USING ITERATED CONTROL VARIATES

SYLVAIN MAIRE

Abstract. In this paper we describe a new variance reduction method for Monte Carlo integration based on an iterated computation of L^2 approximations using control variates. This computation leads to non linear unbiased estimates for each of the coefficients of the truncated L^2 expansion. We give estimations of the variance of these estimates without further hypotheses on the approximation basis. We study especially the convergence of our algorithm in the case of a polynomial decay of these coefficients. As an application, regular monodimensional functions will be approximated using a Fourier basis on periodised functions, Legendre and Tchebychef polynomial L^2 approximations. The order of our method will appear to be almost optimal in this case. Numerical examples will be given as a comparison with standard Monte Carlo estimates.

Keywords: Monte Carlo integration, Variance reduction, L^2 approximations, High order estimates, Iterative method.

Introduction

Monte Carlo methods constitute a very interesting alternative to deterministic methods in the numerical solution of a wide range of problems linked to applied mathematics. They offer numerous advantages in relation to these methods which go from very little sensitivity to the dimension and the geometry of the problem to the easiness of its programming passing by their straightforward parallelisation. They also allow the numerical computation of the pointwise solution of some partial differential equations without any kind of discretisation. The main drawback of these methods is a very slow rate of convergence which cannot give a result with an accuracy better than two or three digits without a lot of computation. Numerous methods, known as variance reduction methods, try to get rid of this drawback. We are going to speak here of a new variance reduction method for one of the main fields where Monte Carlo methods are used: numerical integration.

In the first part, we will describe standard variance reduction methods and especially the control variates method for Monte Carlo integration. This will show that these methods are inefficient unless we can find a reasonable approximation for the integrand. We will then present an iterative method which enables a numerical L^2 approximation for any orthonormal approximation basis. It will be based on a Monte Carlo computation of a correction on this approximation at each step of the algorithm. We will then study its intrisic properties by giving estimations on the variance of each of the estimates of the coefficients of the L^2 approximation.

In the second part, we will study our algorithm especially for monodimensional regular functions. We will describe the quality of the approximation in the case of a polynomial decay of the coefficients of the L^2 approximation. We will first apply our algorithm to the Fourier expansion on periodised functions for which such a property holds. Then we will compute approximations using both Legendre and Tchebychef polynomials which satisfy the same property. We will show that some of our estimates are not far from being optimal for C^k functions in a sense that their order are $0(N^{\frac{1}{2}-k-\varepsilon})$ compared to the optimal order which is $0(N^{-\frac{1}{2}-k})$. Numerical results will be given and compared to standard Monte Carlo integration.

1. A global study of the iterated method

1.1. A few words on variance reduction.

1.1.1. *Principle.* We wish to compute the integral

$$I = \int_D f(x)dx$$

where $D = [0,1]^Q$ with $Q \in N^*$ and where $f \in L^2(D)$. To give its Monte Carlo approximation, we write $I = E(f(X))$ where X is a Q-dimensional random variable whose componants are uniform and independent on $[0,1]$. I is approximated by

$$I \approx \frac{1}{N}\sum_{i=1}^{N} f(X_i)$$

where X_i are N independent copies of X. The rate of convergence of this approximation given by the central limit theorem is $\frac{\sigma_f}{\sqrt{N}}$ where σ_f^2 is the variance of the random variable $f(X)$ that is

$$\sigma_f^2 = \int_D f^2(x)dx - (\int_D f(x)dx)^2.$$

A natural way to increase the rate of convergence of this method is to find a function $\widetilde{f}$ such that $\int_D \widetilde{f}(x)dx = \int_D f(x)dx$ and for which $\sigma_{\widetilde{f}}^2 < \sigma_f^2$ holds. This can be done by using a wide number of methods, called variance reduction methods, among which the most important are importance sampling, antithetic variables, stratification method or the one we shall use in the sequel, control variates. For more details about these methods, one can check Davis-Rabinowitz [5], Kalos-Whitlock [9], Krommer-Ueberhuber [10], Lapeyre [11] or Press [15].

1.1.2. *Control variates.* We describe the control variates method as follows: a function h is given such that its integral is known exactly, one can then take as a function with smaller variance $\widetilde{f}(x) = f(x) - h(x)$. We can compute I by

$$I = \int_D (f(x) - h(x))dx + \int_D h(x)dx$$

where the first integral is computed using Monte Carlo and the second is known exactly. Control variates appear to be an interesting method but it has nevertheless some drawbacks. First, variance-reduction has to be big enough to compensate the additional numerical cost due to the computation of the $h(X_i)$. Secondly, the main drawback is that the function h is generally chosen without using the information given by the values of the functions f at the X_i . We are going to try to make a better use of this information to built the function h. To do so, we will compute iteratively the coefficients of a L^2 approximation of f using only this information.

This will give us not only a tool for variance-reduction but also a L^2 approximation of f. We can notice that this idea is very similar to those used for importance sampling and stratification methods. One can use these methods to obtain Monte Carlo methods with increased convergence rate. This is for instance done and described in Davis-Rabinowitz [5], Haber [6], Lepage [12], Philippe [13] and Press-Farrar [14]. Atanassov-Dimov [1] in a recent work have built a Monte Carlo method with an optimal rate of convergence for regular multidimensional functions

1.2. **Description of the approximation algorithm.** We wish to find an approximation h of f with the first p terms of an orthonormal basis e_k in $L^2(D)$. We have

$$f(x) = \sum_{k=1}^{p} a_k e_k(x) + r(x)$$

with $a_k =< f, e_k >$ and $< r, e_k >= 0$, which give for its approximation

$$h(x) = \sum_{k=1}^{p} a_k e_k(x).$$

We will assume furthermore that the e_k are bounded on D. The principle of our method is to compute a Monte carlo approximation $a_k^{(j)}$ of the coefficients a_k at step j of the algorithm with the help of a correction on the residues $a_k^{(j-1)} - a_k$ at step $j-1$ of the l'algorithm. First we will decribe only two steps of our algorithm and check its performances.

The random variables used for these two steps will be called respectively X_i and Y_i and their sizes are N. We compute an approximation $a_k^{(1)}$ of a_k with the X_i by

$$a_k^{(1)} = \frac{1}{N} \sum_{i=1}^{N} f(X_i) e_k(X_i)$$

and also an approximation $f^{(1)}$ of f by

$$f^{(1)}(x) = \sum_{k=1}^{p} a_k^{(1)} e_k(x).$$

We consider now the function $v(x) = f(x) - f^{(1)}(x)$ and we approximate the coefficients $b_k =< v, e_k >$ by $b_k^{(1)}$ with the Y_i, which are independent of the X_i. We have

$$b_k^{(1)} = \frac{1}{N} \sum_{i=1}^{N} (f(Y_i) - f^{(1)}(Y_i)) e_k(Y_i)$$

and a new approximation of a_k by

$$a_k^{(2)} = a_k^{(1)} + b_k^{(1)}.$$

Since each of the componants of the expansion of v is close to zero and from the orthonormality of the e_k, we can expect under some assumptions on r, p and n that estimating a_k by $a_k^{(2)}$ is better than using a crude Monte Carlo approximation with $2N$ sample values. Particularly, we can also expect that the more coefficients we will have to compute, the more we will need sample values to calculate them. Furthermore, it is also crucial to have an accurate approximation for the variance to be strongly reduced. So, our method will not be efficient unless we have a good approximation with only a few terms. We will give in the following, different ways to obtain such an approximation for monodimensional regular functions.

1.3. **Properties of the algorithm.** To estimate the performances of our algorithm, we have to give the expression of $a_k^{(2)}$ with the different random variables, to compute its expectation and try to estimate its variance. This will give us what we can expect from this algorithm without further hypotheses on the approximation basis.

1.3.1. *Computation of the coefficients of the quadratic approximation.* We have

$$a_k^{(1)} = \sum_{j=1}^{p} \alpha_{k,j} a_j + R_k^{(1)}$$

with

$$\alpha_{k,j} = \frac{1}{N} \sum_{i=1}^{N} e_j(X_i) e_k(X_i)$$

and

$$R_k^{(1)} = \frac{1}{N} \sum_{i=1}^{N} r(X_i) e_k(X_i).$$

Furthermore,

$$v(x) = \sum_{j=1}^{p}(a_j - a_j^{(1)})e_j(x) + r(x),$$

$$b_k^{(1)} = \sum_{j=1}^{p}(a_j - a_j^{(1)})\beta_{k,j} + R_k^{(2)}$$

with

$$\beta_{k,j} = \frac{1}{N}\sum_{i=1}^{N} e_j(Y_i)e_k(Y_i)$$

and

$$R_k^{(2)} = \frac{1}{N}\sum_{i=1}^{N} r(Y_i)e_k(Y_i).$$

Hence, as $a_k^{(2)} = a_k^{(1)} + b_k^{(1)}$, we obtain

$$a_k^{(2)} = \sum_{j=1}^{p} Q_{k,j}^{(2)}a_j + T_k^{(2)}$$

with

$$Q_{k,j}^{(2)} = \alpha_{k,j} + \beta_{k,j} - \sum_{s=1}^{p}\alpha_{s,j}\beta_{k,s}$$

and

$$T_k^{(2)} = R_k^{(1)}(1 - \beta_{k,k}) - \sum_{j\neq k} R_j^{(1)}\beta_{k,j} + R_k^{(2)}.$$

Two terms occur in the expression of $a_k^{(2)}$. The term $T_k^{(2)}$ represents the cut-off linked part in the approximation of a_k by $a_k^{(2)}$. The term

$$\sum_{j=1}^{p} Q_{k,j}^{(2)}a_j$$

represents the estimation of a_k if $r(x) \equiv 0$. We will call it the iteration term and the term $T_k^{(2)}$ the cut-off term. We will first study the properties of the iteration term.

1.3.2. *Study of the iteration term.* We will first give its expectation in the following lemma.

Lemma 1.1. *We have* $E(Q_{k,j}^{(2)}) = \delta_{k,j}$ *and also* $E(\sum_{j=1}^{p} Q_{k,j}^{(2)}a_j) = a_k$.

Proof. We already have

$$E(\alpha_{k,j}) = E(\beta_{k,j}) = \int_D e_k(x)e_j(x)dx = \delta_{k,j}.$$

Then,

$$E(Q^{(2)}_{k,j}) = E(\alpha_{k,j}) + E(\beta_{k,j}) - \sum_{s=1}^{p} E(\alpha_{s,j})E(\beta_{k,s})$$

because of the independence between X_i and Y_i. This implies that

$$E(Q^{(2)}_{k,j}) = \delta_{k,j} + \delta_{k,j} - \sum_{s=1}^{p} \delta_{s,j}\delta_{k,s}$$

and so that $E(Q^{(2)}_{k,j}) = \delta_{k,j}$. The proof of the second assumption is straightforward. □

We now have to give an estimation of its variance. We will first give an estimation of the variance of the $Q^{(2)}_{k,j}$ in the following theorem.

Theorem 1.2. *There is a constant $C(p)$ and a constant $K(p) \leq p^2$, depending only on the approximation basis, such that*

$$Var(Q^{(2)}_{k,j}) \leq K(p)\frac{C(p)^2}{N^2}.$$

If we call furthermore $Q^{(M)}_{k,j}$ the coefficient of a_j in the approximation of a_k by $a^{(M)}_k$ at the Mth step of the algorithm, we have

$$E(Q^{(M)}_{k,j}) = \delta_{k,j}$$

and

$$Var(Q^{(M)}_{k,j}) \leq \frac{K(p)^{M-1}}{N^M}C(p)^M.$$

Proof. See section 3.1 □

1.3.3. *Study of the cut-off term.* We are now going to study the term $T^{(M)}_k$ in the definition of $a^{(M)}_k$. The nature of this term is quite different from the others. We cannot describe its behaviour without knowing $r(x)$ precisely. For the moment, we will give a control on this term depending just on $r(x)$ in the following lemma.

Lemma 1.3. *There are constant $\gamma(p)$ and $\gamma_1(p)$ such that*

$$Var(T_k^{(2)}) \leq 2(\frac{\gamma(p)}{N} + \frac{p^2}{N^2}\gamma_1(p)) \int_D r^2(x)dx.$$

Proof. See section 3.2 □

2. THE MONODIMENSIONAL CASE

We will now give the performances of our algorithm assuming that the coefficients a_k decrease as $\frac{C}{L^k}$ where L and C are positive constants. This may seem to be a very restrictive assumption, but it is in fact a quite natural choice for regular functions in dimension one. We will first use the Fourier basis on periodised functions and then Legendre and Tchebychef polynomial L^2 approximations for which such an assumption holds. We will show that each of these basis has its own interest and that the relative algorithm is incomparably better than a standard Monte Carlo method.

2.1. Polynomial decay of the a_k.

We will now describe, in the following theorem, the accuracy of our estimates for each of the coefficients of the approximation basis and for the approximation itself when the a_k have a polynomial decay.

Theorem 2.1. *Assuming that*

(i)

$$\frac{p}{N} < 1,$$

(ii) there are constants C_1 and $L > 1$ such that for every $k \geq 1$

$$|a_k| \leq \frac{C_1}{k^L},$$

(iii)

$$\tau(p) = \max(\sup_{1\leq k\leq p} \int_D (1 - e_k^2(x))^2 dx, \sup_{1\leq j,k\leq p, j\neq k} \int_D e_k^2(x)e_j^2(x)dx) \leq \frac{N}{4}.$$

Putting

$$f(x) = \sum_{k=1}^{p} a_k e_k(x) + r(x)$$

where

$$r(x) = \sum_{k=p+1}^{\infty} a_k e_k(x).$$

Let

$$f^{(M)}(x) = \sum_{k=1}^{p} a_k^{(M)} e_k(x)$$

be the approximation of f at the Mth step of the algorithm. We have

$$E(a_k^{(M)}) = a_k,$$

furthermore

$$Var(a_k^{(M)}) \le 2(\mu(p)\frac{1}{p^{2L-1}} + \mu_1 \frac{K(p)^{M-1}}{N^M} C(p)^M)$$

and also

$$E(\int_D (f(x) - f^{(M)}(x))^2 dx) \le 2p(\mu(p)\frac{1}{p^{2L-1}} + \mu_1 \frac{K(p)^{M-1}}{N^M} C(p)^M) + \mu_2 \frac{1}{p^{2L-1}}$$

where $\mu(p)$, μ_1, μ_2, $K(p)$ and $C(p)$ are positive constants.

Proof. See section 3.3 □

Remark 2.2. This theorem shows that if

$$\frac{K(p)C(p)}{N} < 1,$$

which is always true for N large enough, then we have

$$\limsup_{M\to+\infty} Var(a_k^{(M)}) \le 2\mu(p)\frac{1}{p^{2L-1}},$$

and

$$\limsup_{M\to+\infty} E(\int_D (f(x) - f^{(M)}(x))^2 dx) \le 2\mu(p)\frac{1}{p^{2L-2}} + \mu_2 \frac{1}{p^{2L-1}}.$$

This also shows that when p increases, it is better to use $a_k^{(M)}$ to approximate a_k than using control variates on $f - f^{(M)}$.

2.2. **Application to the Fourier expansion on a periodised function.**

2.2.1. *The periodisation method.* This method enables this kind of decay for Fourier coefficients with the help of a change of variables for regular functions on the unit interval with possible singularities at the boundaries. One can build a lot of quadrature formulas using this tool by using different types of changes of variables which are efficient for both regular or singular functions. For their description, one can check Davis-Rabinowitz [5], Krommer-Ueberhuber [10], Iri [8], Beckers-Haegmans [2], Sag-Szekeres [16] or Helluy [7] for their use as a unique quadrature in the numerical solution of the Helmoltz integral equation. We will assume that f is either C^∞ or that $f(x) = \frac{h(x)}{x^\alpha}$ with $0 < \alpha < 1$ or that $f(x) = h(x)\ln(x)$ with $h \in C^\infty([0,1])$ and $h(0) \neq 0$. We will describe the case $f \in C^\infty([0,1])$. A lot of changes of variables are used in practice but we will only describe polynomial ones. We now consider the polynomial P such that $P(0) = 0$,$P(1) = 1$ and $P^{(i)}(0) = P^{(i)}(1) = 0$ if $1 \leq i \leq L$.

One can easily see that P is increasing, that the function $g(t) = f(P(t))P^{'}(t)$ is such that

$$\int_0^1 g(t)dt = \int_0^1 f(t)dt$$

and that $g^{(i)}(0) = g^{(i)}(1) = 0$ if $0 \leq i \leq L$. We can now check by integrating by parts that the Fourier coefficients are decreasing like $\frac{C_1}{k^L}$. Our algorithm can be used on the function g in this case.

2.2.2. *Convergence of the algorithm.* Our main task is to study how the constants which are used in the theorem (1.4) depend on p. We choose the functions in the approximation basis to be $\sqrt{2}\cos(2k\pi x)$, $\sqrt{2}\sin(2k\pi x)$ for $1 \leq k \leq q$ and the function identically equal to 1 which we will call $e_p(x)$. We have $p = 2q+1$ terms in our approximation basis and it is obvious that $\mu(p)$ does not depend on p because the basis functions are uniformly bounded. We will show furthermore in section 3.4 that $C(p)$ is equal to 1 and that $K(p) = O(p)$.

2.2.3. *Computation of the order of the method.* The estimate of the variance $a_k^{(M)}$ is given by

$$Var(a_k^{(M)}) \leq 2(\mu(p)\frac{1}{p^{2L-1}} + \mu_1\frac{K(p)^{M-1}}{N^M}C(p)^M).$$

From what we have just seen, we have

$$Var(a_k^{(M)}) \leq 2(\mu\frac{1}{p^{2L-1}} + \mu_1\frac{(Ap)^{M-1}}{N^M}).$$

If we now choose $N = 2Ap$, we get

$$Var(a_k^{(M)}) \leq 2(\mu\frac{1}{N^{2L-1}} + \mu_1(\frac{1}{2})^M)$$

with $Q = MN$ sample values. The optimal choice for M and N is achieved when

$$\frac{1}{N^{2L-1}} = \beta(\frac{1}{2})^M$$

where β is a positive constant, which gives asymptotically

$$M = \frac{\log(N)}{(2L-1)\log(2)}.$$

Hence, we get using $Q = N\log(N)$ sample values the estimate

$$Var(a_k^{(M)}) \leq C_1\frac{1}{N^{2L-1}}.$$

This shows that using N sample values, we have $\forall\varepsilon > 0$

$$Var(a_k^{(M)}) \leq C_2\frac{1}{N^{2L-1-\varepsilon}}$$

and so that the order of our method is $\forall\varepsilon > 0$

$$\frac{1}{N^{L-\frac{1}{2}-\varepsilon}}$$

Remark 2.3. We can compare this result to the optimal order we can obtain by Monte Carlo method for a L times differentiable function which is

$$\frac{1}{N^{L+\frac{1}{2}}}$$

and which can be achieved using various methods and in particular in Atanassov-Dimov [1]. This shows that our method is not quite optimal, but it gives in addition an approximation of the function with few terms.

2.2.4. *Numerical results.* We are now going to give numerical results for different functions using a polynomial of degree $deg(p)$ in the change of variable and where the coefficients of the approximation basis will be computed with M steps and N sample values by step for a total number of sample values equal to $N_t = NM$. We will compare the exact value of the integral $I(f) = \int_0^1 f(x)dx$ and its approximation given by $\widetilde{I}(f) =< 1, g^{(M)} >$ at the end of the M steps. This will be done for the functions $f_1(x) = \exp(x)$, $f_2(x) = \ln(x)$ and $f_3(x) = \frac{1}{\sqrt{x}}$ on $[0,1]$.

$deg(p)$	N	M	N_t	q	p	$\left\|I(f_1) - \widetilde{I}(f_1)\right\|$	$\left\|I(f_2) - \widetilde{I}(f_2)\right\|$	$\left\|I(f_3) - \widetilde{I}(f_3)\right\|$
21	50	10	500	5	11	$1.6\ 10^{-4}$	$3.8\ 10^{-4}$	$4\ 10^{-5}$
21	50	20	1000	10	21	$2.9\ 10^{-6}$	$7.6\ 10^{-6}$	$1.4\ 10^{-5}$
21	100	20	2000	15	31	$8.0\ 10^{-7}$	$2.2\ 10^{-7}$	$6.0\ 10^{-8}$
37	100	30	3000	15	31	$6.2\ 10^{-9}$	$7.4\ 10^{-10}$	$6.4\ 10^{-9}$

We obtain a very good accuracy compared to a standard Monte Carlo method. We can see that for example, we obtain an accuracy of 9 digits with only 3000 evaluations of the functions g_i. This would require 10^{18} evaluations for the standard method. Of course, we have to evaluate other functions to obtain such a good result, but the number of these evaluations for the previous examples is about 10^5 which is clearly smaller than 10^{18} . The algorithm always converges when N is greater than $2p$.

2.3. **Application to the Legendre polynomial approximation basis.** An even more natural basis to approximate regular functions is the Legendre polynomial basis. The polynomial decay of the coefficients of the approximation is achieved with no need of periodisation. We will prove this on $[-1,1]$ using classical results on these polynomials. Most of the results achieved in this part and in the following part concerning Tchebychef polynomials can be found in Bernardi-Maday [3]. One can also find some more results and applications on this subject in Canuto-Hussaini [4].

2.3.1. *Basic properties of Legendre polynomials.* The Legendre polynomials L_n are the orthogonal polynomials on $[-1,1]$ with respect to the inner product $< P, Q >= \int_{-1}^{1} P(x)Q(x)dx$ such that $L_0(x) = 1$ and $L_n(1) = 1$. Their norms are given by

$$(\int_{-1}^{1} L_n^2(x)dx)^{\frac{1}{2}} = (\frac{1}{n+\frac{1}{2}})^{\frac{1}{2}}.$$

and they verify the differential equation

$$\frac{d}{dx}((1-x^2)L_n') + n(n+1)L_n = 0,$$

which will be the main tool used to study the quality of the approximation on the Legendre polynomial

2.3.2. *Quality of the approximation.* We define by $P_N([-1,1])$ the linear space of polynomials with maximal degree N. The projection $\pi_1^{(N)}(f)$ on this space is given by

$$\pi_1^{(N)}(f) = \sum_{n=0}^{N} \alpha_n L_n$$

where α_n are equal to

$$\frac{1}{||L_n||_{L^2}^2} \int_{-1}^{1} f(x)L_n(x)dx.$$

We will from now on assume that f belongs to $C^{2m}([-1,1])$. The speed of decay of the α_n is then given by the following theorem.

Theorem 2.4. *There is a constant C_1 depending only on f such that*

$$|\alpha_n| \leq \frac{C_1}{n^{2m}}.$$

Proof. See section 3.5 □

We deduce immediately the following corollary in which we compute the rate of decay of the coefficients a_n from the expansion of f this time on the orthonormal polynomials $\widetilde{L_n}$.

Corollary 2.5. *There is a constant C_1, depending only on f, such that*

$$|a_n| \leq \frac{C_1}{n^{2m}}.$$

The same kind of result will be valid on $[0,1]$. We can now use the results from theorem (2.1) with the Legendre polynomial basis.

2.3.3. *Numerical implementation of the method.* The situation concerning the constants $\tau(p)$, $\mu(p)$ but mainly $C(p)$ and $K(p)$ does not seem to be as favorable as with the use of the Fourier basis. We can nevertheless say that $K(p) \leq \frac{p^2}{2}$. Indeed, Legendre polynomials on $[-1,1]$ are alternately even and odd which shows that half of the integrals

$$\int_{-1}^{1} e_k^2(x) e_s(x) e_{s_1}(x) dx$$

are equal to zero. The result is also true on D because of the symmetry in relation to $\frac{1}{2}$. The integrals in which one of the functions has a degree higher than the sum of the other three degrees are also equal to zero, but they are on the other hand very few. Hence, we do not have $K(p) = O(p)$ here. The situation is even worse for $C(p)$. For relatively small values of p, $C(p)$ is admittedly lower than 2, but it grows quickly when p increases. This leads to two remarks. First the number of sample values used at each step of the algorithm will have to be greater to ensure its convergence. Secondly, we could not expect to give an order of our method except maybe in using a piecewise approximation, because of the lack of control on $C(p)$.

2.3.4. *Numerical results.* We can only use the exponential function in the comparison with the previous results because the other two functions were not regular and the function $\frac{1}{\sqrt{x}}$ did not even belong to $L^2(D)$. We will however test our algorithm on power functions which offer various degrees of regularity. This will be done by computing the p coefficients of the approximation basis at the Mth step and using

N sample values by step for a number $N_t = NM$ of sample values. We will compare the exact value $I(f) = \int_0^1 f(x)dx$ and its approximation given by $\widetilde{I}(f)$, for the functions $f_1(x) = \exp(x)$, $f_2(x) = x^{\frac{7}{2}}$, $f_3(x) = x^{\frac{3}{2}}$ on $[0, 1]$.

N	M	N_t	p	$\left\|I(f_1) - \widetilde{I}(f_1)\right\|$	$\left\|I(f_2) - \widetilde{I}(f_2)\right\|$	$\left\|I(f_3) - \widetilde{I}(f_3)\right\|$
10	40	400	2	$5.0\ 10^{-5}$	$3.0\ 10^{-4}$	$5.4\ 10^{-4}$
15	40	600	3	$1.4\ 10^{-5}$	$1.5\ 10^{-5}$	$1.2\ 10^{-4}$
20	40	800	5	$7.0\ 10^{-8}$	$2.4\ 10^{-6}$	$1.3\ 10^{-5}$
40	40	1600	7	$1.1\ 10^{-10}$	$1.6\ 10^{-7}$	$3.1\ 10^{-6}$
50	40	2000	10	$3.6\ 10^{-14}$	$1.4\ 10^{-8}$	$1.4\ 10^{-6}$

We have obtained a much better approximation for the exponential function than with the Fourier basis. This can be explained by the artificial increase of the constants appearing in the decay of the coefficients due to the periodisation method. We can also notice that the accuracy decreases with the degree of regularity of the functions which was predictable. The only drawback of this new approach is the increase of the number of sample values at each step of the algorithm. For example, the algorithm failed to converge if N was lower than $3p$ for even relatively small values of p.

2.4. **Application to the Tchebychef polynomial approximation basis.** Both of the two previous methods have given very satisfying results and each of them possesses interesting properties. The algorithm based on the Fourier basis needs less sample values to converge because $C(p)$ is bounded by 1 and $K(p) = O(p)$. The algorithm based on the Legendre polynomials offers a better approximation for regular functions because it avoids the bad effects of periodisation. Because of their trigonometric expressions, we can expect that Tchebychef polynomials reconcile these properties. This will be confirmed in the following.

2.4.1. *Basic properties of Tchebychef polynomials.* Tchebychef polynomials are the orthogonal polynomials on $[-1, 1]$ with respect to the inner product $< P, Q >= \int_{-1}^1 \frac{P(x)Q(x)}{\sqrt{1-x^2}}dx$. Their expressions are given by $T_n(x) = \cos(n \arccos(x))$. This can be easily checked by induction because

$$\int_{-1}^{1} \frac{T_n(x)T_m(x)}{\sqrt{(1-x^2)}}dx = \int_0^{\pi} \cos(n\theta)\cos(m\theta)d\theta.$$

It also shows that $\|T_0\|^2_{L^2_\omega} = \pi$ and $\|T_n\|^2_{L^2_\omega} = \frac{\pi}{2}$ if $n \geq 1$. They also verify the following differential equation

$$\frac{d}{dx}(\sqrt{1-x^2}T_n^{'}(x)) + n^2\frac{T_n(x)}{\sqrt{1-x^2}} = 0.$$

2.4.2. *Quality of the approximation.* We define by $P_N([-1,1])$ the linear space of polynomials with maximal degree N. The projection $\pi_2^{(N)}(f)$ on this space is given by

$$\pi_2^{(N)}(f) = \sum_{n=0}^{N} \beta_n L_n$$

where β_n are equal to

$$\frac{1}{\|T_n\|^2_{L^2_\omega}} \int_{-1}^{1} \frac{f(x)T_n(x)}{\sqrt{1-x^2}}dx.$$

We still assume that f belongs to $C^{2m}([-1,1])$. The speed of decay of the β_n is then given by the following theorem whose proof relies on the previous differential equation.

Theorem 2.6. *There is a constant C_1 depending only on f such that*

$$|\beta_n| \leq \frac{C_1}{n^{2m}}.$$

We deduce immediately the following corollary in which we compute the rate of decay of the coefficients a_n from the expansion of f on the orthonormal polynomials $\widetilde{T_n}$.

Corollary 2.7. *There is a constant C_2, depending only on f, such that*

$$|a_n| \leq \frac{C_2}{n^{2m}}.$$

2.4.3. *Implementation of the method.* We now approximate f on $D = [-1, 1]$. We could attempt to compute

$$< f, \widetilde{T_n} >= \int_{-1}^{1} \frac{f(x)\widetilde{T_n}(x)}{\sqrt{1-x^2}} dx$$

by Monte Carlo in writing

$$< f, \widetilde{T_n} >= E(\frac{2f(Y)\widetilde{T_n}(Y)}{\sqrt{1-Y^2}})$$

where Y is uniform on $[-1, 1]$. This is not a judicious choice because the random variable

$$Z = \frac{2f(Y)\widetilde{T_n}(Y)}{\sqrt{1-Y^2}}$$

usually has an infinite variance because of the divergence of the integral

$$E(Z^2) = \int_{-1}^{1} \frac{2f^2(x)\widetilde{T_n^2}(x)}{1-x^2} dx.$$

To get rid of this drawback, we will compute $< f, \widetilde{T_n} >= E(\pi f(V)\widetilde{T_n}(V))$ where the density of V is

$$\frac{1}{\pi\sqrt{1-v^2}} 1_{[-1,1]}(v).$$

The variance of the random variable $W = \pi f(V)\widetilde{T_n}(V)$ is now finite because

$$E(W^2) = \int_{-1}^{1} \frac{\pi f^2(x)\widetilde{T_n^2}(x)}{\sqrt{1-x^2}} dx < \infty.$$

We now take the approximation algorithm of paragraph (1.2) using random drawings with the same law as V. The only modifications are the definitions of $\alpha_{i,j}$, $\beta_{i,j}$, $R_k^{(1)}$ and $R_k^{(2)}$ which are obtained by

$$\alpha_{k,j} = \frac{\pi}{N} \sum_{i=1}^{N} e_j(X_i) e_k(X_i),$$

$$\beta_{k,j} = \frac{\pi}{N} \sum_{i=1}^{N} e_j(Y_i) e_k(Y_i),$$

$$R_k^{(1)} = \frac{\pi}{N}\sum_{i=1}^{N} r(X_i)e_k(X_i)$$

and

$$R_k^{(2)} = \frac{\pi}{N}\sum_{i=1}^{N} r(Y_i)e_k(Y_i).$$

It is easy to see that the results from lemma (1.1) and from theorem (1.2) are still valid and that we can obtain $A_{k,k}$, $B_{k,k}$ and $C_{k,k}$ by

$$A_{k,k} = E((1-\frac{\pi}{N}\sum_{i=1}^{N} e_k^2(Y_i))^2) = \int_{-1}^{1}\frac{(1-\pi e_k^2(x))^2}{\pi\sqrt{1-x^2}}dx,$$

$$B_{k,k} = \sup_{s\neq k}\left|\int_{-1}^{1}\frac{(1-\pi e_k^2(x))e_k(x)e_s(x)}{\sqrt{1-x^2}}dx\right|,$$

$$E_{k,k} = \sup_{s,s_1\neq k}\left|\int_{-1}^{1}\frac{\pi e_k^2(x)e_s(x)e_{s_1}(x)}{\sqrt{1-x^2}}dx\right|.$$

By replacing the e_k by their expressions, we get

$$A_{k,k} = \frac{1}{\pi}\int_0^{\pi}(1-2\cos^2(kx))^2dx,$$

$$B_{k,k} = \frac{2}{\pi}\int_0^{\pi}(1-2\cos^2(kx))\cos(kx)\cos(sx)dx,$$

$$E_{k,k} = \frac{2}{\pi}\int_0^{\pi}\cos^2(kx)\cos(sx)\cos(s_1x)dx,$$

if we do not take into account $\widetilde{T_0}$. One can now check using the same kind of computation as for the Fourier basis that $K(p) = O(p)$ and that $C(p)$ is equal to 1. The crucial point is that the last integral is a product of only cosine functions. One can also check that $\mu(p)$ is independent of p. This shows furthermore that the order of our method is

$$\frac{1}{N^{L-\frac{1}{2}-\varepsilon}}$$

$\forall\varepsilon > 0$, if f is L times continuously differentiable, for the computation of

$$\int_{-1}^{1}\frac{f(x)}{\sqrt{1-x^2}}dx.$$

2.4.4. *Numerical results.* The approximation of f by $f^{(M)}$ is

$$f^{(M)}(x) = \sum_{k=0}^{p} a_k^{(M)} \widetilde{T_k}(x)$$

where $a_k^{(M)}$ is an approximation of

$$< f, \widetilde{T_k} >= \int_{-1}^{1} \frac{f(x)\widetilde{T_k}(x)}{\sqrt{1-x^2}} dx.$$

This is a quite satisfying approximation but it does not give us a direct approximation of

$$I(f) = \int_{-1}^{1} f(x)dx.$$

There are two ways two obtain such an approximation. One can compute

$$\widetilde{I}(f) = \sum_{k=0}^{p} a_k^{(M)} \int_{-1}^{1} \widetilde{T_k}(x)dx,$$

or one can use control variates with $f^{(M)}$ as an approximation for f. We are going to take once more the numerical examples we have already studied using Legendre polynomials, using the first process, after having linearly transformed these integrals. For example, the computation of

$$\int_0^1 \exp(x)dx$$

will be compared to

$$\int_{-1}^{1} \frac{\exp(\frac{x+1}{2})}{2} dx.$$

We give our results in the following table, for each of the transformations of the functions $f_1(x) = \exp(x)$, $f_2(x) = x^{\frac{7}{2}}$, $f_3(x) = x^{\frac{3}{2}}$, keeping the same notations as for Legendre polynomials.

N	M	N_t	p	$\left\|I(f_1) - \widetilde{I}(f_1)\right\|$	$\left\|I(f_2) - \widetilde{I}(f_2)\right\|$	$\left\|I(f_3) - \widetilde{I}(f_3)\right\|$
6	30	180	3	$1.1\ 10^{-4}$	$6.6\ 10^{-4}$	$5.3\ 10^{-4}$
10	40	400	5	$9.0\ 10^{-8}$	$3.6\ 10^{-6}$	$6.2\ 10^{-5}$
20	50	1000	10	$2.0\ 10^{-15}$	$4.0\ 10^{-9}$	$1.5\ 10^{-6}$

For a given value of p, we get almost the same accuracy as with the use of Legendre polynomials, but this accuracy is obtained with twice fewer sample values. We can nevertheless notice that these results appear to be slightly less accurate than previous results. This can certainly be explained by the fact that errors on the $a_k^{(M)}$ can accumulate in the approximation of $I(f)$ by

$$\sum_{k=0}^{p} a_k^{(M)} \int_{-1}^{1} \widetilde{T_k}(x)dx,$$

while this integral was directly given by one of these coefficients in the previous examples.

Conclusion

We have developed and studied in this paper a global tool for variance reduction in Monte Carlo integration. We have obtained Monte Carlo estimates with increased convergence rate for monodimensional integrals of regular functions using only sample values. These estimates are not quite optimal but our algorithm gives us in addition an accurate L^2 approximation with only few terms. We will show in a forthcoming paper how to extend this algorithm to multidimensional integration in trying to attenuate the dimensional effect by making a good choice of the elements in the approximation basis.

3. Appendice

3.1. **Proof of theorem 1.2.** Assuming $j = k$, we have

$$Var(Q_{k,k}^{(2)}) = E((Q_{k,k}^{(2)} - 1)^2) = E((\beta_{k,k} - 1)(1 - \alpha_{k,k}) - \sum_{s \neq k} \alpha_{s,k}\beta_{k,s})^2).$$

The term under expectation in the last expression is worth

$$(1 - \alpha_{k,k})^2(\beta_{k,k} - 1)^2 - 2\sum_{s \neq k} \alpha_{s,k}(1 - \alpha_{k,k})\beta_{k,s}(\beta_{k,k} - 1) + \sum_{s, s_1 \neq k} \alpha_{s,k}\alpha_{s_1,k}\beta_{k,s}\beta_{k,s_1}.$$

Each of the p^2 terms in this sum can be written

$$(\frac{1}{N}\sum_{i=1}^{N} h_1(X_i))(\frac{1}{N}\sum_{i=1}^{N} h_2(X_i))(\frac{1}{N}\sum_{i=1}^{N} h_3(Y_i))(\frac{1}{N}\sum_{i=1}^{N} h_4(Y_i))$$

with $E(h_j(X_i)) = 0$ for each of the functions h_j. Its expectation is

$$(\frac{1}{N^2}\sum_{i=1}^{N} E(h_1(X_i)h_2(X_i)))(\frac{1}{N^2}\sum_{i=1}^{N} E(h_3(Y_i)h_4(Y_i)))$$

and because X_i and Y_i have the same distribution

$$\frac{1}{N^2}E(h_1(X_1)h_2(X_1))E(h_3(X_1)h_4(X_1)).$$

If we look for instance at $(1-\alpha_{k,k})^2(\beta_{k,k}-1)^2$, we have

$$\frac{1}{N^2}\int_D (1-e_k^2(x))^2 dx \int_D (1-e_k^2(y))^2 dy.$$

At last

$$Var(Q_{k,k}^2) \leq \frac{p^2}{N^2} D_{k,k}^2,$$

where $D_{k,k} = \max(A_{k,k}, B_{k,k}, E_{k,k})$ with

$$A_{k,k} = \int_D (1-e_k^2(x))^2 dx,$$

$$B_{k,k} = \sup_{s\neq k}\left|\int_D (1-e_k^2(x))e_k(x)e_s(x)dx\right|,$$

$$E_{k,k} = \sup_{s,s_1\neq k}\left|\int_D e_k^2(x)e_s(x)e_{s_1}(x)dx\right|.$$

Assuming now $k \neq j$, we have

$$Var(Q_{k,j}^{(2)}) = E(Q_{k,j}^{2(2)}) = E((\alpha_{k,j}(1-\beta_{k,k}) + (1-\alpha_{j,j})\beta_{k,j} - \sum_{s\neq k,j}\alpha_{s,j}\beta_{k,s})^2).$$

Using the same argument for each of the $Var(Q_{k,j}^{(2)})$, there are constants $D_{k,j}$ very similar to the $D_{k,k}$ such that

$$Var(Q_{j,j_1}^{(2)}) \leq \frac{p^2}{N^2} D_{k,j}^2.$$

If we now let

$$C(p) = \sup_{j,j_1} D_{j,j_1},$$

we have

$$Var(Q^{(2)}_{j,j_1}) \leq \frac{p^2}{N^2} C(p)^2.$$

We can do better than the latter by noticing that when one integrates a product of four terms of an orthonormal basis the result is likely to be zero. This is especially true for functions with disjoint domains, trigonometric functions and also for certain kinds of orthogonal polynomials. We will now define $K(p)$ as the maximum number of non zero valued integrals which appear in the definition of each of the $D_{k,j}$. Hence we have

$$Var(Q^{(2)}_{j,j_1}) \leq \frac{K(p)}{N^2} C(p)^2$$

and we will now show by induction on M that

$$E(Q^{(M)}_{j,j_1}) = \delta_{j,j_1}$$

and also that

$$Var(Q^{(M)}_{j,j_1}) \leq \frac{K(p)^{M-1}}{N^M} C(p)^M.$$

The property is true for M equal to two. We now suppose it is true for M. We consider uniform independent random variables W_i independent of all previous random variables and we define

$$\gamma_{k,j} = \frac{1}{N} \sum_{i=1}^{N} e_j(W_i) e_k(W_i)$$

which are used to define

$$Q^{(M+1)}_{k,j} = Q^{(M)}_{k,j} + \gamma_{k,j} - \sum_{s=1}^{p} Q^{(M)}_{s,j} \gamma_{k,s}.$$

We have

$$E(Q^{(M+1)}_{k,j}) = E(Q^{(M)}_{k,j}) + E(\gamma_{k,j}) - \sum_{s=1}^{p} E(Q^{(M)}_{s,j}) E(\gamma_{k,s}),$$

by induction

$$E(Q_{k,j}^{(M+1)}) = \delta_{k,j} + \delta_{k,j} - \sum_{s=1}^{p} \delta_{s,j}\delta_{k,s}$$

and also

$$E(Q_{k,j}^{(M+1)}) = \delta_{k,j}.$$

We now have to study the variance of the $Q_{j,k}^{(M+1)}$. We will first look at $Q_{k,k}^{(M+1)}$. The term under expectation in

$$Var(Q_{k,k}^{(M+1)}) = E((Q_{k,k}^{(M+1)} - 1)^2)$$

is worth, as one can easily check

$$(1 - Q_{k,k}^{(M)})^2(\gamma_{k,k} - 1)^2 - 2\sum_{s\neq k} Q_{s,k}^{(M)}(1 - Q_{k,k}^{(M)})\gamma_{k,s}(\gamma_{k,k} - 1) + \sum_{s,s_1\neq k} Q_{s,k}^{(M)} Q_{s_1,k}^{(M)}\gamma_{k,s}\gamma_{k,s_1}.$$

There are at most $K(p)$ terms with non-zero expectation in this sum. From the induction step, we have

$$E((1 - Q_{k,k}^{(M)})^2) \leq \frac{K(p)^{M-1}}{N^M} C(p)^M.$$

From the Cauchy-Schwartz inequality we have,

$$\left| E(Q_{s,k}^{(M)}(1 - Q_{k,k}^{(M)})) \right| \leq \frac{K(p)^{M-1}}{N^M} C(p)^M$$

and also

$$\left| E(Q_{s,k}^{(M)} Q_{s_1,k}^{(M)}) \right| \leq \frac{K(p)^{M-1}}{N^M} C(p)^M.$$

Furthermore

$$E(\gamma_{k,k} - 1)^2 \leq \frac{C(p)}{N},$$

$$|E(\gamma_{k,s}(\gamma_{k,k} - 1))| \leq \frac{C(p)}{N},$$

$$|E(\gamma_{k,s}\gamma_{k,s_1})| \leq \frac{C(p)}{N}.$$

We have finally

$$Var(Q_{k,k}^{(M+1)}) \leq K(p)\frac{K(p)^{M-1}}{N^M}C(p)^M\frac{C(p)}{N} = \frac{K(p)^M}{N^{M+1}}C(p)^{M+1}.$$

The reasoning is analogous for estimating the variance of the other terms.

3.2. **Proof of lemma 1.3.**

$$T_k^{(2)} = R_k^{(1)}(1-\beta_{k,k}) - \sum_{j\neq k} R_j^{(1)}\beta_{k,j} + R_k^{(2)} = \sum_j R_j^{(1)}Z_j + R_k^{(2)},$$

where $Z_k = 1-\beta_{k,k}$ and $Z_j = -\beta_{k,j}$ if $j \neq k$. Hence, we have

$$Var(T_k^{(2)}) \leq 2(E((R_k^{(2)})^2) + E((\sum_j R_j^{(1)}Z_j)^2)$$

and since

$$E((R_k^{(2)})^2) = \frac{1}{N}\int_D e_k^2(x)r^2(x)dx,$$

we have

$$E((R_k^{(2)})^2) \leq \frac{\gamma(p)}{N}\int_D r^2(x)dx$$

with

$$\gamma(p) = \sup_{1\leq k\leq p}\sup_{x\epsilon D} e_k^2(x).$$

We also have from the independence between Z_j and $R_j^{(1)}$

$$E((\sum_j R_j^{(1)}Z_j)^2) = E(\sum_{j,j_1} R_j^{(1)}Z_jR_{j_1}^{(1)}Z_{j_1}) = \sum_{j,j_1} E(R_j^{(1)}R_{j_1}^{(1)})E(Z_jZ_{j_1}).$$

From the Cauchy-Schwartz inequality,

$$\left|E(R_j^{(1)}R_{j_1}^{(1)})\right| \leq E\left|R_j^{(1)}R_{j_1}^{(1)}\right| \leq (E((R_j^{(1)})^2)E((R_{j_1}^{(1)})^2))^{\frac{1}{2}}$$

with gives from the latter

$$\left|E(R_j^{(1)}R_{j_1}^{(1)})\right| \leq \frac{\gamma(p)}{N}\int_D r^2(x)dx.$$

We also have from Cauchy-Schwartz inequality

$$|E(Z_j Z_{j_1})| \leq E\,|Z_j Z_{j_1}| \leq (E((Z_j)^2)E((Z_{j_1})^2))^{\frac{1}{2}}.$$

Since

$$E(\beta_{k,j}^2) = \frac{1}{N}\int_D e_k^2(x)e_j^2(x)dx$$

and

$$E((1-\beta_{k,k})^2) = \frac{1}{N}\int_D (1-e_k^2(x))^2 dx,$$

it is easy to see that putting

$$\tau(p) = \max(\sup_{1\leq k\leq p}\int_D (1-e_k^2(x))^2 dx, \sup_{j\neq k}\int_D e_k^2(x)e_j^2(x)dx),$$

we have

$$|E(Z_j Z_{j_1})| \leq \frac{\tau(p)}{N}.$$

If we now let $\gamma_1(p) = \tau(p)\gamma(p)$, we have

$$E((\sum_j R_j^{(1)} Z_j)^2) \leq \frac{p^2}{N^2}\gamma_1(p)\int_D r^2(x)dx$$

and finally

$$Var(T_k^{(2)}) \leq 2(\frac{\gamma(p)}{N} + \frac{p^2}{N^2}\gamma_1(p))\int_D r^2(x)dx.$$

3.3. **Proof of theorem 2.1.** To prove this theorem we have to study the expectation and the variance of each of the terms in the definition of $a_k^{(M)}$ by

$$a_k^{(M)} = \sum_{j=1}^{p} Q_{k,j}^{(M)} a_j + T_k^{(M)}.$$

We begin by $T_k^{(M)}$. We already have $E(T_k^{(2)}) = 0$ and also

$$\int_D r^2(x)dx = \sum_{k=p+1}^{\infty} a_k^2 \leq C_1^2 \sum_{k=p+1}^{\infty} \frac{1}{k^{2L}} \leq \mu_2 \frac{1}{p^{2L-1}}.$$

From lemma (1.4) we have

$$Var(T_k^{(2)}) \leq 2\mu_2(\frac{\gamma(p)}{N} + \frac{p^2}{N^2}\gamma_1(p))\frac{1}{p^{2L-1}}$$

and since $\frac{p}{N} < 1$

$$Var(T_k^{(2)}) \leq \mu(p)\frac{1}{p^{2L-1}}$$

where $\mu(p) = 2\mu_2\gamma(p) + \gamma_1(p)$ is a positive constant.

We will know check by induction that this stays true at each of the M steps of the algorithm. We have

$$T_k^{(M)} = T_k^{(M-1)}(1-\rho_{k,k}) - \sum_{j\neq k} T_j^{(M-1)}\rho_{k,j} + R_k^{(M)} = \sum_j T_j^{(M-1)} Z_j + R_k^{(M)}$$

putting

$$\rho_{k,j} = \frac{1}{N}\sum_{i=1}^{N} e_j(W_i)e_k(W_i),$$

where the W_i are iid and independent of all random variables previously used, where

$$R_k^{(M)} = \frac{1}{N}\sum_{i=1}^{N} r(W_i)e_k(W_i)$$

and where $Z_k = 1-\rho_{k,k}$ and $Z_j = -\rho_{k,j}$ if $j \neq k$. Hence $E(T_k^{(M)}) = 0$ and

$$Var(T_k^{(M)}) \leq 2(E((R_k^{(M)})^2) + E((\sum_j T_j^{(M-1)} Z_j)^2)),$$

hence, as we assume

$$Var(T_k^{(M-1)}) \leq \mu(p)\frac{1}{p^{2L-1}},$$

using the same arguments than for the estimation of $Var(T_k^{(2)})$,

$$Var(T_k^{(M)}) \leq 2(\mu_2\frac{\gamma(p)}{N} + \mu(p)\frac{\tau(p)}{N})\frac{1}{p^{2L-1}}.$$

We only have to check that

$$2(\mu_2\frac{\gamma(p)}{N} + \mu(p)\frac{\tau(p)}{N}) \leq \mu(p)$$

to have

$$Var(T_k^{(M)}) \leq \mu(p)\frac{1}{p^{2L-1}}.$$

This is obviously true since $\frac{\tau(p)}{N} \leq \frac{1}{4}$ and $2\mu_2\frac{\gamma(p)}{N} \leq \frac{\mu(p)}{2}$.

We shall now study the iteration term

$$\sum_{j=1}^{p} Q_{k,j}^{(M)} a_j.$$

Since $E(Q_{j,j_1}^{(M)}) = \delta_{j,j_1}$, we have

$$E(\sum_{j=1}^{p} Q_{k,j}^{(M)} a_j) = a_k$$

and

$$Var(\sum_{j=1}^{p} Q_{k,j}^{(M)} a_j) = E((\sum_{j=1}^{p} Q_{k,j}^{(M)} a_j - a_k)^2)$$

For the sake of simplicity we will write from now on $Q_{k,k}^{(M)}$ instead of $Q_{k,k}^{(M)} - 1$. Using the assumption (ii), we have

$$E((\sum_{j=1}^{p} Q_{k,j}^{(M)} a_j)^2) \leq C_1^2 E((\sum_{j=1}^{p} \left| Q_{k,j}^{(M)} \right| \frac{1}{jL})^2),$$

expanding this expression,

$$E((\sum_{j=1}^{p} Q_{k,j}^{(M)} a_j)^2) \leq C_1^2 \sum_{1 \leq i, i_1 \leq p} \frac{E(\left| Q_{k,i}^{(M)} \right| \left| Q_{k,i_1}^{(M)} \right|)}{(ii_1)^L},$$

from the Cauchy-Schwartz inequality

$$E((\sum_{j=1}^{p} Q_{k,j}^{(M)} a_j)^2) \leq C_1^2 \sum_{1 \leq i, i_1 \leq p} \frac{E(\left| Q_{k,i}^{(M)} \right|^2)^{\frac{1}{2}} E(\left| Q_{k,i_1}^{(M)} \right|^2)^{\frac{1}{2}}}{(ii_1)^L}$$

and from theorem (1.2)

$$E((\sum_{j=1}^{p} Q_{k,j}^{(M)} a_j)^2) \leq C_1^2 \frac{K(p)^{M-1}}{N^M} C(p)^M \sum_{1 \leq i, i_1 \leq p} \frac{1}{(ii_1)^L}.$$

Since $L > 1$, we have

$$E((\sum_{j=1}^{p} Q_{k,j}^{(M)} a_j)^2) \leq \mu_1 \frac{K(p)^{M-1}}{N^M} C(p)^M$$

where

$$\mu_1 = (\sum_{1 \leq i \leq \infty} \frac{1}{i^L})^2.$$

From what we have just seen

$$E(a_k^{(M)}) = E(\sum_{j=1}^{p} Q_{k,j}^{(M)} a_j + T_k^{(M)}) = E(\sum_{j=1}^{p} Q_{k,j}^{(M)} a_j) + E(T_k^{(M)}) = a_k$$

and using the inequality

$$Var(a_k^{(M)}) \leq 2(E(\sum_{j=1}^{p} Q_{k,j}^{(M)} a_j - a_k)^2 + E(T_k^{(M)})^2),$$

$$Var(a_k^{(M)}) \leq 2(\mu(p)\frac{1}{p^{2L-1}} + \mu_1 \frac{K(p)^{M-1}}{N^M} C(p)^M).$$

We can also check the quadratic error between f and its approximation $f^{(M)}$. We have

$$E(\int_D (f(x) - f^{(M)}(x))^2 dx) = E(\sum_{k=1}^{p} (a_k - a_k^{(M)})^2) + \int_D r^2(x) dx \leq$$
$$\sum_{k=1}^{p} Var(a_k^{(M)}) + \int_D r^2(x) dx,$$

which shows that

$$E(\int_D (f(x) - f^{(M)}(x))^2 dx) \leq 2p(\mu(p)\frac{1}{p^{2L-1}} + \mu_1 \frac{K(p)^{M-1}}{N^M} C(p)^M) + \mu_2 \frac{1}{p^{2L-1}}.$$

3.4. **Computation of $K(p)$ and $C(p)$ for the Fourier basis.** We will proove that $C(p)$ is equal to 1 and that $K(p) = O(p)$ by looking precisely at the contribution of $Q_{k,k}^{(2)}$ in the definition of $K(p)$ and $C(p)$. We begin by the terms of the form

$$\int_D e_k^2(x) e_s(x) e_{s_1}(x) dx$$

which have the main contribution in the definition of these constants. We will first consider the case when only cosine functions appear in these terms. Their expressions are given by

$$4\int_0^1 \cos^2(2k\pi x)\cos(2s\pi x)\cos(2s_1\pi x)dx$$

with k, s and s_1 between 1 and q and $s, s_1 \neq k$. Linearizing, we obtain

$$\int_0^1 \cos(2\pi(s-s_1)x)+\cos(2\pi(s+s_1)x)+\frac{1}{2}(\cos(2\pi(-2k-s+s_1)x)+$$
$$\cos(2\pi(2k-s+s_1)x)+\cos(2\pi(-2k+s+s_1)x)+\cos(2\pi(2k+s+s_1)x)dx$$

that is, because $s+s_1>0$ and $2k+s+s_1>0$,

$$\int_0^1 \cos(2\pi(s-s_1)x)+\frac{1}{2}(\cos(2\pi(-2k-s+s_1)x)+$$
$$\cos(2\pi(2k-s+s_1)x)+\cos(2\pi(-2k+s+s_1)x))dx.$$

The value of this expression is 1 if $s = s_1$, because in this case,

$$\int_0^1 \cos(2\pi(s-s_1)x)dx = 1$$

and the other integrals are zero valued. If now $s \neq s_1$, it is easy to check that for given values of s and s_1, one and only one of the last three integrals is likely to be non zero valued and its value is then $\frac{1}{2}$. Furthermore, we can check that for a given value of s, there are at most two values of s_1 which lead to non zero valued integrals. As a conclusion, there are at most $3(q-1)$ non zero valued integrals of the form

$$4\int_0^1 \cos^2(2k\pi x)\cos(2s\pi x)\cos(2s_1\pi x)dx$$

and

$$\left|4\int_0^1 \cos^2(2k\pi x)\cos(2s\pi x)\cos(2s_1\pi x)dx\right| \leq 1$$

when $1 \leq s, s_1 \neq k \leq q$. Using the same kind of computations on

$$4\int_0^1 \cos^2(2k\pi x)\cos(2s\pi x)\sin(2s_1\pi x)dx,$$

$$4\int_0^1 \cos^2(2k\pi x)\sin(2s\pi x)\cos(2s_1\pi x)dx,$$

$$4\int_0^1 \cos^2(2k\pi x)\sin(2s\pi x)\sin(2s_1\pi x)dx,$$

we can notice that

$$\left|4\int_0^1 \cos^2(2k\pi x)\cos(2s\pi x)\sin(2s_1\pi x)dx\right| \leq 1,$$

$$\left|4\int_0^1 \cos^2(2k\pi x)\sin(2s\pi x)\cos(2s_1\pi x)dx\right| \leq 1,$$

$$\left|4\int_0^1 \cos^2(2k\pi x)\sin(2s\pi x)\sin(2s_1\pi x)dx\right| \leq 1.$$

and furthermore that there are at most $12(q-1)$ non zero valued integrals among the $(p-2)^2$ integrals of the form

$$\int_0^1 e_k^2(x)e_s(x)e_{s_1}(x)dx$$

where $1 \leq s, s_1 \neq k \leq p-1$. We now only have to study the terms of the previous form in which the function $e_p(x)$ is involved. They can be written, letting $e_{s_1}(x) = e_p(x)$,

$$2\sqrt{2}\int_0^1 \cos^2(2k\pi x)\cos(2s\pi x)dx,$$

or

$$2\sqrt{2}\int_0^1 \cos^2(2k\pi x)\sin(2s\pi x)dx,$$

if $e_s(x) \neq e_p(x)$ and

$$2\int_0^1 \cos^2(2k\pi x)dx = 1$$

otherwise. As

$$\cos^2(2k\pi x)\cos(2s\pi x) = \frac{1}{2}\cos(2s\pi x) + \frac{1}{4}(\cos(2(s+2k)\pi x) + \cos(2(s-2k)\pi x)),$$

all the integrals of the form

$$2\sqrt{2}\int_0^1 \cos^2(2k\pi x)\cos(2s\pi x)dx$$

are equal to zero except when $s = 2k$. There value is then $\frac{\sqrt{2}}{2}$. We have the same kind of results for the term

$$2\sqrt{2}\int_0^1 \cos^2(2k\pi x)\sin(2s\pi x)dx.$$

We have proved that there are at most $O(p)$ non zero valued integrals of the form

$$\int_D e_k^2(x)e_s(x)e_{s_1}(x)dx$$

and that their absolute value is bounded by 1. We only have to study integrals of the form

$$\int_0^1 (1-e_k^2(x))e_k(x)e_s(x)dx$$

and

$$\int_0^1 (1-e_k^2(x))^2 dx$$

which are anyway at most $O(p)$. Noticing that

$$1-(\sqrt{2}\cos(2\pi kx))^2 = -2\sin^2(2\pi kx), 1-(\sqrt{2}\sin(2\pi kx))^2 = -2\cos^2(2\pi kx),$$

we can easily see, using previous computations, that their absolute value is bounded by 1. This shows that the contribution $Q_{k,k}^{(2)}$ in the definition of $K(p)$ and $C(p)$ are as expected. We will obtain the same result on the contribution of $Q_{k,j}^{(2)}$ that is, as a conclusion, $K(p) \leq Ap$ and $C(p) = 1$.

3.5. **Proof of theorem 2.4.** From the differential equation satisfied by the L_n, we have

$$\alpha_n = \frac{1}{\|L_n\|_{L^2}^2}\int_{-1}^1 f(x)L_n(x)dx = \frac{1}{\|L_n\|_{L^2}^2}\frac{1}{n(n+1)}\int_{-1}^1 -f(x)\frac{d}{dx}((1-x^2)L_n^{'}(x))dx.$$

Integrating by parts, we get

$$\int_{-1}^1 -f(x)\frac{d}{dx}((1-x^2)L_n^{'}(x))dx = [-f(x)(1-x^2)L_n^{'}(x)]_{-1}^1 + \int_{-1}^1 -f^{'}(x)(1-x^2)L_n^{'}(x)dx,$$

integrating by parts once more we have

$$\int_{-1}^{1} -f(x)\frac{d}{dx}((1-x^2)L_n^{'}(x))dx = \int_{-1}^{1} -\frac{d}{dx}(f^{'}(x)(1-x^2))L_n(x)dx.$$

Hence we obtain

$$\alpha_n = \frac{1}{\|L_n\|_{L^2}^2}\frac{1}{n(n+1)}\int_{-1}^{1}(Af)(x)L_n(x))dx,$$

letting

$$(Af)(x) = -\frac{d}{dx}(f^{'}(x)(1-x^2)).$$

By induction we have

$$\alpha_n = \frac{1}{\|L_n\|_{L^2}^2}\frac{1}{(n(n+1))^m}\int_{-1}^{1}(A^m f)(x)L_n(x))dx.$$

Furthermore from the Cauchy-Schwartz inequality

$$|\alpha_n| \le \frac{1}{\|L_n\|_{L^2}^2}\frac{1}{(n(n+1))^m}\|A^m f\|_{L^2}\|L_n\|_{L^2},$$

which gives because of the regularity of f

$$|\alpha_n| \le \frac{1}{\|L_n\|_{L^2}}\frac{C_1}{(n(n+1))^m},$$

where C_1 is a positive constant. We obtain for $n \ge 1$,

$$|\alpha_n| \le C_1\frac{\sqrt{n+\frac{1}{2}}}{n^{2m}}.$$

References

[1] E. I. ATANASSOV, I. T. DIMOV, A new optimal Monte Carlo method for calculating integral of smooth functions, Monte Carlo Methods and Appl., Vol. 5, No. 2, pp. 149-167 (1999).

[2] M. BECKERS, A. HAEGMANS, Transformations of integrands for lattice rules, in Numerical Integration-Recent Developpements, Software and applications, T. O. Espelid, A.Genz, Eds., Kluwer, Dordrecht, the Netherlands, 1992, pp. 329-340.

[3] C. BERNARDI, Y. MADAY, Approximations spectrales de problèmes aux limites elliptiques. Springer-Verlag. 1992

[4] C. CANUTO, M. Y. HUSSAINI, A. Quarteroni, T. A. Zang, Spectral methods in fluid dynamics. Springer-Verlag. 1990

[5] P. DAVIS, P. RABINOWITZ, Methods of numerical integration, second edition, Comp. Sci. and applied Math., Academic Press, 1984.

[6] S. HABER, A modified Monte-Carlo quadrature, Math. Comp. 20 (1966), pp. 361-368.

[7] P. HELLUY, S. MAIRE, P. RAVEL, Higher order numeric quadratures for regular or singular functions, applications for the helmoltz integral equations, ASME design engineering technical conferences, 1999. DECT99/VIB-8819.

[8] M. IRI, S. MORIGUTI, Y. TAKAZAWA, On a certain quadrature formula (in japanese), Kokyuroku of Res., Inst. for Math. Sci., Kyoto Univ., Vol. 91, 1970, pp. 82-118.

[9] M. H. KALOS, P. A. WHITLOCK, Monte Carlo methods, John Wiley and sons, 1986.

[10] A. R. KROMMER et C. W. UEBERHUBER, Computational integration. SIAM, 1998.

[11] B. LAPEYRE. E. PARDOUX, R. SENTIS, Méthodes de Monte-Carlo pour les équations de transport et de diffusion. Springer-Verlag, 1998.

[12] G. P. LEPAGE, Journal of Computational Physics, Vol 27, pp. 192-203, 1978.

[13] A . PHILIPPE, Processing simulation output by Riemann sums. J. Statistic. Comput. Simulation 59, 295-314, 1997.

[14] W. H. PRESS, G. R. FARRAR, Computer in Physics, vol. 4, pp. 190-195, 1990.

[15] W. H. PRESS, S. A. TEUKOLSKY, W. T. VETTERLING, B. P. FLANNERY. Numerical recipes in FORTRAN. Cambridge university press, 1996.

[16] T. W. SAG, G. SZEKERES, Numerical evaluation of high-dimensionnal integrals, Math. Comp. 18 (1964), pp. 245-253.

ISITV, Université de Toulon et du Var. Avenue G. Pompidou BP 56 - 83262 La Valette du Var CEDEX FRANCE. e-mail: maire@isitv.univ-tln.fr