



# Monte Carlo integration of C<sup>r</sup> functions with adaptive variance reduction: an asymptotic analysis

Leszek Plaskota<sup>1</sup> · Paweł Przybyłowicz<sup>2</sup> · Łukasz Stępień<sup>2</sup>

Received: 24 November 2022 / Accepted: 16 April 2023 / Published online: 22 May 2023 © The Author(s) 2023

## Abstract

The theme of the present paper is numerical integration of  $C^r$  functions using randomized methods. We consider variance reduction methods that consist in two steps. First the initial interval is partitioned into subintervals and the integrand is approximated by a piecewise polynomial interpolant that is based on the obtained partition. Then a randomized approximation is applied on the difference of the integrand and its interpolant. The final approximation of the integral is the sum of both. The optimal convergence rate is already achieved by uniform (nonadaptive) partition plus the crude Monte Carlo; however, special adaptive techniques can substantially lower the asymptotic factor depending on the integrand. The improvement can be huge in comparison to the nonadaptive method, especially for functions with rapidly varying *r*th derivatives, which has serious implications for practical computations. In addition, the proposed adaptive methods are easily implementable and can be well used for automatic integration.

Keywords Monte Carlo · Adaption · Variance reduction · Asymptotic constants

Mathematics Subject Classification 65C05 · 65D30

Communicated by Stefano De Marchi.

Leszek Plaskota leszekp@mimuw.edu.pl

> Paweł Przybyłowicz pprzybyl@agh.edu.pl

Łukasz Stępień lstepie@agh.edu.pl

<sup>1</sup> Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, ul. S. Banacha 2, 02-097 Warsaw, Poland

<sup>2</sup> Faculty of Applied Mathematics, AGH University of Science and Technology, Al. A. Mickiewicza 30, 30-059 Kraków, Poland

🖉 Springer

## **1** Introduction

Adaption is a useful tool to improve performance of algorithms. The problems of numerical integration and related to it  $L^1$  approximation are not exceptions, see, e.g., [9] for a survey of theoretical results on the subject. If an underlying function possesses some singularities and is otherwise smooth, then using adaption is necessary to localize the singular points and restore the convergence rate typical for smooth functions, see, e.g., [6, 12-14]. For functions that are smooth in the whole domain, adaptive algorithms usually do not offer a better convergence rate than nonadaptive algorithms; however, they can essentially lower asymptotic constants. This is why adaptive quadratures are widely used for numerical integration, see, e.g., [1, 3, 7]. Their superiority over nonadaptive quadratures is rather obvious, but precise answers to the quantitative question of "how much adaption helps" are usually missing. This gap was partially filled by recent results of [2, 10, 11], where best asymptotic constants of deterministic algorithms that use piecewise polynomial interpolation were determined for integration and  $L^1$  approximation of r-times continuously differentiable functions  $f:[a,b] \to \mathbb{R}$ . In this case, adaption relies on adjusting the partition of the interval [a, b] to the underlying function. While the convergence rate is of order  $N^{-r}$ , where N is the number of function evaluations used, it turns out that the asymptotic constant depends on f via the factor of  $(b-a)^r \|f^{(r)}\|_{L^1}$  for uniform partition, and  $\|f^{(r)}\|_{L^{1/(r+1)}}$  for best adaptive partition.

In the current paper, we present the line of thinking similar to that of the aforementioned papers. The difference is that now we want to carry out the analysis and obtain asymptotic constants for randomized algorithms.

Our goal is the numerical approximation of the integral

$$Sf = \int_{a}^{b} f(x) \,\mathrm{d}x. \tag{1.1}$$

It is well known that for  $f \in L^2(a, b)$  the *crude Monte Carlo*,

$$M_N f = \frac{b-a}{N} \sum_{i=1}^{N} f(t_i), \quad \text{where} \quad t_i \stackrel{iid}{\sim} U(a,b), \tag{1.2}$$

returns an approximation with expectation  $\mathbb{E}(M_N f) = Sf$  and error (standard deviation)

$$\sqrt{\mathbb{E}(Sf - M_N f)^2} = \frac{\sigma(f)}{\sqrt{N}}, \quad \text{where} \quad \sigma(f)^2 = (b - a)S(f^2) - (Sf)^2. \tag{1.3}$$

Suppose that the function enjoys more smoothness, i.e.,

$$f \in C^r([a, b]).$$

approximation of the form

$$\overline{M}_{N,r}(f) = S(L_{m,r}f) + M_n(f - L_{m,r}f), \qquad (1.4)$$

where  $L_{m,r}$  is the piecewise polynomial interpolation of f of degree r - 1 using a partition of the interval [a, b] into m subintervals,  $M_n$  is a Monte Carlo type algorithm using n random samples of f, and N is the total number of function evaluations used (for arguments chosen either deterministically or randomly). The optimal rate is already achieved for uniform (nonadaptive) partition and crude Monte Carlo. Then, see Theorem 2.1 with  $\beta = 0$ , the error asymptotically equals

$$c (b-a)^{r+1/2} \| f^{(r)} \|_{L^2(a,b)} N^{-(r+1/2)},$$

where c depends only on the choice of the interpolation points within subintervals. The main result of this paper relies on showing that with the help of adaption the asymptotic error of the methods (1.4) can be reduced to

$$c \left\| f^{(r)} \right\|_{L^{1/(r+1)}(a,b)} N^{-(r+1/2)}, \tag{1.5}$$

see Theorems 3.1 and 4.1. Observe that the gain can be significant, especially when the derivative  $f^{(r)}$  drastically changes. For instance, for r = 4, [a, b] = [0, 1], and f(x) = 1/(x + d), adaption is asymptotically roughly  $5.7 \times 10^{12}$  times better than nonadaption if  $d = 10^{-4}$ , and  $1.8 \times 10^{29}$  times if  $d = 10^{-8}$ .

We construct two randomized algorithms, denoted  $\overline{M}_{N,r}^*$  and  $\overline{M}_{N,r}^{**}$ , that achieve the error (1.5). Although they use different initial approaches; namely, stratification versus importance sampling, in the limit they both reach essentially the same partition, such that the  $L^1$  errors of Lagrange interpolation in all subintervals are equalized. However, numerical tests of Sect. 5 show that the algorithm  $\overline{M}_{N,r}^*$  achieves the error (1.5) with some delay, which makes  $\overline{M}_{N,r}^{**}$  worth recommending rather than  $\overline{M}_{N,r}^*$  in practical computations.

Other advantages of  $\overline{M}_{N,r}^{**}$  are that it is easily implementable and, as shown in Sect. 6, it can be successfully used for automatic Monte Carlo integration.

Our analysis has been so far restricted to one-dimensional integrals only. In a future work it will be extended and corresponding adaptive Monte Carlo algorithms will be constructed for multivariate integration, where randomization finds its major application. The current paper is the first step in this direction.

In the sequel, we use the following notation. For two functions of N we write  $g_1(N) \leq g_2(N)$  iff  $\limsup_{N \to \infty} g_1(N)/g_2(N) \leq 1$ , and we write  $g_1(N) \approx g_2(N)$  iff  $\lim_{N \to \infty} g_1(N)/g_2(N) = 1$ . Similarly, for functions of  $\varepsilon$  we write  $h_1(\varepsilon) \leq h_2(\varepsilon)$  iff  $\limsup_{\varepsilon \to 0^+} h_1(\varepsilon)/h_2(\varepsilon) \leq 1$ , and  $h_1(\varepsilon) \approx h_2(\varepsilon)$  iff  $\lim_{\varepsilon \to 0^+} h_1(\varepsilon)/h_2(\varepsilon) = 1$ .

#### 2 Variance reduction using Lagrange interpolation

We first derive some general error estimates for the variance reduction algorithms of the form (1.4), where the standard Monte Carlo is applied for the error of piecewise Lagrange interpolation. Specifically, we divide the interval [a, b] into m subintervals using a partition  $a = x_0 < x_1 < \cdots < x_m = b$ , and on each subinterval  $[x_{j-1}, x_j]$  we approximate f using Lagrange interpolation of degree r - 1 with the interpolation points

$$x_{j,s} = x_{j-1} + z_s(x_j - x_{j-1}), \quad 1 \le s \le r,$$

where

$$0 \le z_1 < z_2 < \dots < z_r \le 1 \tag{2.1}$$

are fixed (independent of the partition). Denote such an approximation by  $L_{m,r}f$ . Then  $f = L_{m,r}f + R_{m,r}f$  with  $R_{m,r}f = f - L_{m,r}f$ . The integral Sf is finally approximated by

$$M_{m,n,r}f = S(L_{m,r}f) + M_n(R_{m,r}f),$$

where  $M_n$  is the crude Monte Carlo (1.2). We obviously have  $\mathbb{E}(\overline{M}_{m,n,r}f) = Sf$ . Since

$$Sf - \overline{M}_{m,n,r}f = Sf - S(L_{m,r}f) - M_n(R_{m,r}f) = S(R_{m,r}f) - M_n(R_{m,r}f),$$

by (1.3) we have

$$\mathbb{E}\left(Sf - \overline{M}_{m,n,r}f\right)^2 = \frac{1}{n}\left((b-a)S\left((R_{m,r}f)^2\right) - \left(S(R_{m,r}f)\right)^2\right).$$

Note that

$$S((R_{m,r}f)^2) = \int_a^b (f - L_{m,r}f)^2(x) \, \mathrm{d}x = \|f - L_{m,r}f\|_{L^2(a,b)}^2$$

is the squared  $L^2$ -error of the applied (piecewise) polynomial interpolation, while

$$S(R_{m,r}f) = \int_{a}^{b} (f - L_{m,r}f)(x) \, \mathrm{d}x = S(f) - S(L_{m,r}f)$$

is the error of the quadrature  $\overline{Q}_{m,r}f = S(L_{m,r}f)$ .

From now on we assume that f is not a polynomial of degree smaller than or equal to r-1, since otherwise  $\overline{M}_{m,n,r}f = Sf$ . Define the polynomial

$$P(z) = (z - z_1)(z - z_2) \cdots (z - z_r).$$
(2.2)

We first consider the interpolation error  $||f - L_{m,r}f||_{L^2(a,b)}$ . Let

$$\alpha = \|P\|_{L^2(0,1)} = \left(\int_0^1 |P(z)|^2 dz\right)^{1/2}.$$
(2.3)

For each j, the local interpolation error equals

$$\|f - L_{m,r}f\|_{L^{2}(x_{j-1},x_{j})} = \left(\int_{x_{j-1}}^{x_{j}} |(x - x_{j,1}) \cdots (x - x_{j,r})f[x_{j,1}, \dots, x_{j,r}, x]|^{2} dx\right)^{1/2}$$
$$= \alpha h_{j}^{r+1/2} \frac{|f^{(r)}(\xi_{j})|}{r!}, \qquad \xi_{j} \in [x_{j-1}, x_{j}].$$

Hence

$$||f - L_{m,r}f||_{L^{2}(a,b)} = \frac{\alpha}{r!} \left( \sum_{j=1}^{m} h_{j}^{2r+1} |f^{(r)}(\xi_{j})|^{2} \right)^{1/2}.$$

In particular, for the equispaced partition, in which case  $h_j = (b - a)/m$ , we have

$$\|f - L_{m,r}f\|_{L^{2}(a,b)} = \frac{\alpha}{r!} \left(\frac{b-a}{m}\right)^{r} \left(\frac{b-a}{m} \sum_{j=1}^{m} |f^{(r)}(\xi_{j})|^{2}\right)^{1/2}$$
$$\approx \frac{\alpha}{r!} \left(\frac{b-a}{m}\right)^{r} \|f^{(r)}\|_{L^{2}(a,b)} \quad \text{as} \quad m \to +\infty.$$

Now, we consider the quadrature error  $Sf - \overline{Q}_{m,r}f$ . Let

$$\beta = \int_0^1 P(z) \,\mathrm{d}z. \tag{2.4}$$

The local integration errors equal

$$\int_{x_{j-1}}^{x_j} (f - L_{m,r}f)(x) \, \mathrm{d}x = \int_{x_{j-1}}^{x_j} (x - x_{j,1}) \cdots (x - x_{j,r}) f[x_{j,1}, \dots, x_{j,r}, x] \, \mathrm{d}x$$
$$= \frac{1}{r!} \int_{x_{j-1}}^{x_j} (x - x_{j,1}) \cdots (x - x_{j,r}) f^{(r)}(\xi_j(x)) \, \mathrm{d}x, \quad \xi_j(x) \in [x_{j-1}, x_j].$$

Choose arbitrary  $\zeta_j \in [x_{j-1}, x_j]$  for  $1 \le j \le m$ . Then

$$\left| \frac{1}{r!} \int_{x_{j-1}}^{x_j} (x - x_{j,1}) \cdots (x - x_{j,r}) f^{(r)}(\xi_j(x)) \, \mathrm{d}x \right|$$
$$- \frac{f^{(r)}(\zeta_j)}{r!} \int_{x_{j-1}}^{x_j} (x - x_{j,1}) \cdots (x - x_{j,r}) \, \mathrm{d}x \right|$$

D Springer

$$= \frac{1}{r!} \left| \int_{x_{j-1}}^{x_j} (x - x_{j,1}) \cdots (x - x_{j,r}) \left( f^{(r)}(\xi_j(x)) - f^{(r)}(\zeta_j) \right) dx \right|$$
  
$$\leq \omega(h_j) \frac{h_j^{r+1}}{r!} \|P\|_{L^1(0,1)},$$

where  $\omega$  is the modulus of continuity of  $f^{(r)}$ . We also have

$$\frac{f^{(r)}(\zeta_j)}{r!} \int_{x_{j-1}}^{x_j} (x - x_{j,1}) \cdots (x - x_{j,r}) \, \mathrm{d}x = \frac{\beta}{r!} h_j^{r+1} f^{(r)}(\zeta_j).$$

Hence  $Sf - \overline{Q}_{m,r}f = X_m + Y_m$ , where

$$X_m = \frac{\beta}{r!} \sum_{j=1}^m h_j^{r+1} f^{(r)}(\zeta_j) \quad \text{and} \quad |Y_m| \le \frac{\|P\|_{L^1(0,1)}}{r!} \sum_{j=1}^m \omega(h_j) h_j^{r+1}.$$

In particular, for the equispaced partition,

$$X_m = \frac{\beta}{r!} (b-a)^r \left( \sum_{j=1}^m \frac{b-a}{m} f^{(r)}(\zeta_j) \right) m^{-r},$$
  
$$|Y_m| \le \frac{\|P\|_{L^1(0,1)}}{r!} \omega \left( \frac{b-a}{m} \right) (b-a)^{r+1} m^{-r}.$$

Suppose that  $\beta \neq 0$  and  $\int_a^b f^{(r)}(x) dx \neq 0$ . Then  $X_m \approx \frac{\beta}{r!}(b-a)^r \left(\int_a^b f^{(r)}(x) dx\right) m^{-r}$ . Since  $\omega(h)$  goes to zero as  $h \to 0^+$ , the component  $X_m$  dominates  $Y_m$  as  $m \to +\infty$ . Hence

$$Sf - \overline{Q}_{m,r}f \approx \frac{\beta}{r!} \left(\frac{b-a}{m}\right)^r \left(\int_a^b f^{(r)}(x) \,\mathrm{d}x\right) \quad \text{as} \quad m \to +\infty.$$

On the other hand, if  $\beta = 0$  or  $\int_a^b f^{(r)}(x) dx = 0$  then the quadrature error converges to zero faster than  $m^{-r}$ , i.e.

$$\lim_{m \to +\infty} \left( Sf - \overline{Q}_{m,r} f \right) m^r = 0.$$

Note that  $\beta = 0$  if and only if the quadrature  $\overline{Q}_{m,r}$  has the degree of exactness at least r, i.e., it is exact for all polynomials of degree r or less. Obviously, the maximal degree of exactness equals 2r - 1.

We see that for the equidistant partition of the interval [a, b] the error  $(\mathbb{E}(Sf - \overline{M}_{m,n,r}f)^2)^{1/2}$  is asymptotically proportional to

$$\phi(m,n) = n^{-1/2}m^{-r},$$

regardless of the choice of points  $z_i$  in (2.1). Let us minimize  $\phi(m, n)$  assuming the total number of points used is at most N. We have two cases depending on whether both endpoints of each subinterval are used in interpolation. If so, i.e., if  $z_1 = 0$  and  $z_r = 1$  (in this case  $r \ge 2$ ) then N = (r - 1)m + 1 + n. The optimal values are

$$m^* = \frac{2r(N-1)}{(r-1)(2r+1)}, \qquad n^* = \frac{N-1}{2r+1},$$
 (2.5)

for which

$$\phi(m^*, n^*) = \sqrt{2} \left(1 - \frac{1}{r}\right)^r \left(\frac{r+1/2}{N}\right)^{r+1/2}$$

Otherwise we have N = rm + n. The optimal values are

$$m^* = \frac{2N}{2r+1}, \qquad n^* = \frac{N}{2r+1},$$
 (2.6)

for which

$$\phi(m^*, n^*) = \sqrt{2} \left(\frac{r+1/2}{N}\right)^{r+1/2}$$

Denote by  $\overline{M}_{N,r}$  the corresponding algorithm with the equidistant partition, where for given N the values of n and m equal correspondingly  $\lfloor n^* \rfloor$  and  $\lfloor m^* \rfloor$ . Our analysis is summarized in the following theorem.

**Theorem 2.1** We have as  $N \to +\infty$  that

$$\sqrt{\mathbb{E}(Sf-\overline{M}_{N,r}f)^2} \approx c_r (b-a)^r C(P,f) N^{-(r+1/2)}$$

where

$$C(P, f) = \sqrt{\alpha^2 (b - a) \left( \int_a^b |f^{(r)}(x)|^2 dx \right) - \beta^2 \left( \int_a^b f^{(r)}(x) dx \right)^2},$$

 $\alpha$  and  $\beta$  are given by (2.3) and (2.4), and

$$c_r = \begin{cases} \sqrt{2} \left(1 - \frac{1}{r}\right)^r \frac{(r+1/2)^{r+1/2}}{r!}, & \text{if } r \ge 2, \ z_1 = 0, \ z_r = 1, \\ \sqrt{2} \frac{(r+1/2)^{r+1/2}}{r!}, & \text{otherwise.} \end{cases}$$
(2.7)

We add that the algorithm  $\overline{M}_{N,r}$  is fully implementable since we assume that we have access to function evaluations at points from [a, b].

🖄 Springer

## 3 First adaptive algorithm

Now we add a stratification strategy to our algorithm of Theorem 2.1 to obtain an adaptive algorithm with a much better asymptotic constant. That is, we divide the initial interval [a, b] into k equal length subintervals  $I_i$ ,  $1 \le i \le k$ , and on each subinterval we apply the approximation of Theorem 2.1 with some  $N_i$ , where

$$\sum_{i=1}^{k} N_i \le N. \tag{3.1}$$

Denote such an approximation by  $\overline{M}_{N,k,r}$ . (Note that  $\overline{M}_{N,r} = \overline{M}_{N,1,r}$ .) Then, by Theorem 2.1, for fixed k we have as all  $N_i \to +\infty$  that

$$\sqrt{\mathbb{E}(Sf-\overline{M}_{N,k,r}f)^2} \approx c_r h^r \left(\sum_{i=1}^k \frac{C_i^2}{N_i^{2r+1}}\right)^{1/2},$$

where

$$C_{i} = C_{i}(P, f) = \sqrt{\alpha^{2} h \int_{I_{i}} |f^{(r)}(x)|^{2} dx - \beta^{2} \left( \int_{I_{i}} f^{(r)}(x) dx \right)^{2}}, \qquad h = \frac{b - a}{k}$$

Minimization of  $\psi(N_1, \dots, N_k) = \left(\sum_{i=1}^k C_i^2 N_i^{-(2r+1)}\right)^{1/2}$  with respect to (3.1) gives

$$N_i^* = \frac{C_i^{1/(r+1)}}{\sum_{j=1}^k C_j^{1/(r+1)}} N, \qquad 1 \le i \le k,$$

and then

$$\psi(N_1^*,\ldots,N_k^*) = \left(\sum_{i=1}^k C_i^{1/(r+1)}\right)^{r+1} N^{-(r+1/2)}.$$

Let  $\xi_i, \eta_i \in I_i$  satisfy  $\int_{I_i} |f^{(r)}(x)|^2 dx = h |f^{(r)}(\xi_i)|^2$  and  $\int_{I_i} f^{(r)}(x) dx = h f^{(r)}(\eta_i)$ . Then

$$C_i = h \sqrt{\alpha^2 |f^{(r)}(\xi_i)|^2 - \beta^2 |f^{(r)}(\eta_i)|^2}$$

and we have as  $k \to +\infty$  that

$$\left(\sum_{i=1}^{k} C_{i}^{1/(r+1)}\right)^{r+1} = h\left(\sum_{i=1}^{k} \left(\alpha^{2} |f^{(r)}(\xi_{i})|^{2} - \beta^{2} |f^{(r)}(\eta_{i})|^{2}\right)^{\frac{1}{2(r+1)}}\right)^{r+1}$$

$$\approx h (\alpha^{2} - \beta^{2})^{1/2} \left( \sum_{i=1}^{k} \left| f^{(r)}(\xi_{i}) \right|^{1/(r+1)} \right)^{r+1}$$
$$\approx h^{-r} (\alpha^{2} - \beta^{2})^{1/2} \left( \sum_{i=1}^{k} h \left| f^{(r)}(\xi_{i}) \right|^{1/(r+1)} \right)^{r+1}$$
$$\approx h^{-r} (\alpha^{2} - \beta^{2})^{1/2} \left( \int_{a}^{b} \left| f^{(r)}(x) \right|^{1/(r+1)} dx \right)^{r+1}$$

It is clear that we have to take  $N_i$  to be an integer and at least r, for instance

$$N_i = \left\lfloor N_i^* \left( 1 - \frac{kr}{N} \right) + r \right\rfloor, \quad 1 \le i \le k.$$

Then the corresponding number  $m_i$  of subintervals and number  $n_i$  of random points in  $I_i$  can be chosen as

$$m_i = \max\left(\lfloor m_i^* \rfloor, 1\right), \quad n_i = \lfloor n_i^* \rfloor,$$

where  $m_i^*$  and  $n_i^*$  are given by (2.5) and (2.6) with N replaced by  $N_i$ .

Denote by  $\overline{M}_{N,r}^*$  the above constructed approximation  $\overline{M}_{N,k_N,r}$  with  $k_N$  such that  $k_N \to +\infty$  and  $k_N/N \to 0$  as  $N \to +\infty$ . For instance,  $k_N = N^{\kappa}$  with  $0 < \kappa < 1$ . Our analysis gives the following result.

**Theorem 3.1** *We have as*  $N \to +\infty$  *that* 

$$\sqrt{\mathbb{E}\left(Sf-\overline{M}_{N,r}^{*}f\right)^{2}} \approx c_{r}\sqrt{\alpha^{2}-\beta^{2}}\left(\int_{a}^{b}\left|f^{(r)}(x)\right|^{1/(r+1)}\mathrm{d}x\right)^{r+1}N^{-(r+1/2)}.$$

The asymptotic constant of the approximation  $\overline{M}_{N,r}^*$  of Theorem 3.1 is never worse than that of  $\overline{M}_{N,r}$  of Theorem 2.1. Indeed, comparing both constants we have

$$c_{r}(b-a)^{r}\sqrt{\alpha^{2}(b-a)\left(\int_{a}^{b}|f^{(r)}(x)|^{2}dx\right) - \beta^{2}\left(\int_{a}^{b}f^{(r)}(x)dx\right)^{2}}$$
  

$$\geq c_{r}\sqrt{\alpha^{2} - \beta^{2}}(b-a)^{r+1/2}\left(\int_{a}^{b}|f^{(r)}(x)|^{2}\right)^{1/2}dx$$
  

$$\geq c_{r}\sqrt{\alpha^{2} - \beta^{2}}\left(\int_{a}^{b}|f^{(r)}(x)|^{1/(r+1)}dx\right)^{r+1},$$

where the first inequality follows from the Schwarz inequality and the second one from Hölder's inequality for integrals. As shown in the introduction, the gain can be significant, especially when the derivative  $f^{(r)}$  drastically changes.

The approximation  $\overline{M}_{N,r}^*$  possesses good asymptotic properties, but is not feasible since we do not have a direct access to the  $C_i$ s. In a feasible implementation one can

approximate  $C_i$  using divided differences, i.e.,

$$\widetilde{C}_i = h \sqrt{\alpha^2 - \beta^2 |d_i| r!}$$
 where  $d_i = f[x_{i,0}, x_{i,1}, \dots, x_{i,r}]$  (3.2)

and  $x_{i,j}$  are arbitrary points from  $I_i$ . Then

$$N_i^* = \frac{|d_i|^{1/(r+1)}}{\sum_{j=1}^{k_N} |d_j|^{1/(r+1)}} N.$$

This works well for functions f for which the rth derivative does not nullify at any point in [a, b]. Indeed, then  $f^{(r)}$  does not change its sign and, moreover, it is separated away from zero. This means that

$$\lim_{N\to\infty} \max_{1\le i\le k_N} C_i/\widetilde{C}_i = 1,$$

which is enough for the asymptotic equality of Theorem 3.1 to hold true.

If  $f^{(r)}$  is not separated away from zero then we may have problems with proper approximations of  $C_i$  in the intervals  $I_i$  where  $|f^{(r)}|$  assumes extremely small values or even zeros. A possible and simple remedy is to choose 'small'  $\Delta > 0$  and modify  $\tilde{C}_i$  as follows:

$$\widetilde{C}_{i} = \begin{cases} h \sqrt{\alpha^{2} - \beta^{2}} |d_{i}| r! & \text{for } |d_{i}| r! \ge \Delta, \\ h \alpha \Delta r! & \text{for } |d_{i}| r! < \Delta. \end{cases}$$
(3.3)

Then, letting  $A_1 = \{a \le x \le b : |f^{(r)}(x)| \ge \Delta\}$  and  $A_2 = [a, b] \setminus A_1$ , we have as  $k \to +\infty$  that

$$\begin{split} \left(\sum_{i=1}^{k} C_{i}^{1/(r+1)}\right)^{r+1} &\lesssim \left(\sum_{i=1}^{k} \widetilde{C}_{i}^{1/(r+1)}\right)^{r+1} \\ &\approx h^{-r} \left(\alpha^{2} - \beta^{2}\right)^{1/2} \left(\int_{A_{1}} \left|f^{(r)}(x)\right|^{\frac{1}{r+1}} \mathrm{d}x + |A_{2}| \left(\sqrt{\frac{\alpha^{2}}{\alpha^{2} - \beta^{2}}} \Delta\right)^{\frac{1}{r+1}}\right)^{r+1}. \end{split}$$

Hence, the approximation of  $C_i$  by (3.3) results in an algorithm whose error is approximately upper bounded by

$$c_r \sqrt{\alpha^2 - \beta^2} \left( \int_a^b \left| f_{\Delta}^{(r)}(x) \right|^{1/(r+1)} \mathrm{d}x \right)^{r+1} N^{-(r+1/2)},$$

where

$$\left|f_{\Delta}^{(r)}(x)\right| = \max\left(\left|f^{(r)}(x)\right|, \sqrt{\frac{\alpha^2}{\alpha^2 - \beta^2}}\,\Delta\right).\tag{3.4}$$

We obviously have  $\lim_{\Delta \to 0^+} \int_a^b |f_{\Delta}^{(r)}(x)|^{1/(r+1)} dx = \int_a^b |f^{(r)}(x)|^{1/(r+1)} dx.$ 

A closer look at the deterministic part of  $\overline{M}_{N,r}^*$  shows that the final partition of the interval [a, b] tends to equalize the  $L^1$  errors in all of the *m* subintervals. As shown in [11], such a partition is optimal in the sense that it minimizes the asymptotic constant in the error  $||f - L_{m,r}f||_{L^1(a,b)}$  among all possible piecewise Langrange interpolations  $L_{m,r}$ . A disadvantage of the optimal partition is that it is not nested. This makes it necessary to start the computations from scratch when N is updated to a higher value. Also, a proper choice of the sequence  $k_N = N^{\kappa}$  is problematic, especially when N is still relatively small. On one hand, the larger  $\kappa$  the better the approximation of  $C_i$  by  $\tilde{C}_i$ , but also the more far away the partition to the optimal one, but also the worse the approximation of  $C_i$ . This trade–off significantly affects the actual behavior of the algorithm, which can be seen in numerical experiments of Sect. 5.

In the following section, we propose another adaptive approach leading to an easily implementable algorithm that produces nested partitions close to optimal and possesses asymptotic properties similar to that of  $\overline{M}_{N,r}^*$ . As we shall see in Sect. 6, nested partitions are vital for automatic Monte Carlo integration.

#### 4 Second adaptive algorithm

Consider a  $\rho$ -weighted integration problem

$$S_{\rho}f = \int_{a}^{b} f(x)\rho(x) \,\mathrm{d}x,$$

where the function  $\rho : [a, b] \to \mathbb{R}$  is integrable and positive a.e. and  $\int_a^b \rho(x) dx = 1$ . The corresponding Monte Carlo algorithm is

$$M_{n,\rho}f = \frac{1}{n}\sum_{i=1}^{n}f(t_i), \quad t_i \stackrel{iid}{\sim} \mu_{\rho},$$

where  $\mu_{\rho}$  is the probability distribution on [a, b] with density  $\rho$ . Then

$$\mathbb{E}(S_{\rho}f - M_{n,\rho}f)^{2} = \frac{1}{n} (S_{\rho}(f^{2}) - (S_{\rho}f)^{2}).$$

Now, the non-weighted integral (1.1) can be written as

$$S(f) = \int_{a}^{b} h(x)\rho(x) \, \mathrm{d}x = S_{\rho}(h), \text{ where } h(x) = \frac{f(x)}{\rho(x)}$$

Then

$$\mathbb{E}(Sf - M_{n,\rho}h)^2 = \mathbb{E}(S_{\rho}h - M_{n,\rho}h)^2 = \frac{1}{n} (S_{\rho}(h^2) - (S_{\rho}h)^2) = \frac{1}{n} (S_{\rho}(f/\rho)^2 - (Sf)^2).$$

Let's go further on and apply a variance reduction,

$$\overline{M}_{n,\rho}f = S(Lf) + M_{n,\rho}\left(\frac{f-Lf}{\rho}\right),\tag{4.1}$$

where Lf is an approximation to f. Then

$$\mathbb{E}\left(Sf - \overline{M}_{n,\rho}f\right)^2 = \frac{1}{n} \left( \int_a^b \frac{(f - Lf)^2(x)}{\rho(x)} \, \mathrm{d}x - \left( \int_a^b (f - Lf)(x) \, \mathrm{d}x \right)^2 \right).$$

The question is how to choose L and  $\rho$  to make the quantity

$$\int_a^b \frac{(f-Lf)^2(x)}{\rho(x)} \,\mathrm{d}x - \left(\int_a^b (f-Lf)(x) \,\mathrm{d}x\right)^2$$

as small as possible.

Observe that if

$$\rho(x) = \frac{|(f - Lf)(x)|}{\|f - Lf\|_{L^1(a,b)}}$$

then

$$\mathbb{E}\left(Sf - \overline{M}_{n,\rho}f\right)^2 = \frac{1}{n} \left( \|f - Lf\|_{L^1(a,b)}^2 - \left(\int_a^b (f - Lf)(x) \mathrm{d}x\right)^2 \right)$$

and this error is even zero if (f - Lf)(x) does not change its sign. This suggests the following algorithm.

Suppose that  $Lf = L_{m,r}f$  is based on a partition of [a, b] such that the  $L^1$  errors in all *m* subintervals  $I_i$  have the same value, i.e.,

$$\|f - L_{m,r}f\|_{L^{1}(I_{i})} = \frac{1}{m} \|f - L_{m,r}f\|_{L^{1}(a,b)}, \quad 1 \le i \le m.$$
(4.2)

Then we apply the variance reduction (4.1) with density

$$\rho(x) = \frac{1}{mh_i}, \qquad x \in I_i, \quad 1 \le i \le m, \tag{4.3}$$

where  $h_i$  is the length of  $I_i$ . That is, for the corresponding probability measure  $\mu_{\rho}$  we have  $\mu_{\rho}(I_i) = \frac{1}{m}$  and the conditional distribution  $\mu_{\rho}(\cdot|I_i)$  is uniform on  $I_i$ . We now derive an error formula for such an approximation. Let  $\gamma = ||P||_{L^1(a,b)} =$ 

We now derive an error formula for such an approximation. Let  $\gamma = ||P||_{L^1(a,b)} = \int_0^1 |P(z)| dz$ . (Recall that *P* is given by (2.2).) We have

$$\|f - L_{m,r}f\|_{L^{1}(I_{i})} = \frac{\gamma}{r!} h_{i}^{r+1} |f^{(r)}(\xi_{i})| \text{ and } \|f - L_{m,r}f\|_{L^{2}(I_{i})} = \frac{\alpha}{r!} h_{i}^{r+1/2} |f^{(r)}(\zeta_{i})|$$

🖄 Springer

for some  $\xi_i, \zeta_i \in I_i$ . Denoting

$$A = h_i^{r+1} \left| f^{(r)}(\xi_i) \right|$$
(4.4)

(which is independent of *i*) we have as  $m \to +\infty$  that

$$\left(\int_{a}^{b} \frac{(f - L_{m,r}f)^{2}(x)}{\rho(x)} dx\right)^{1/2} = \left(m \sum_{i=1}^{m} h_{i} \int_{I_{i}} (f - L_{m,r}f)^{2}(x) dx\right)^{1/2}$$
  
$$= \frac{\alpha}{r!} \left(m \sum_{i=1}^{m} h_{i}^{2r+2} |f^{(r)}(\zeta_{i})|^{2}\right)^{1/2} \approx \frac{\alpha}{r!} \left(m \sum_{i=1}^{m} h_{i}^{2r+2} |f^{(r)}(\xi_{i})|^{2}\right)^{1/2}$$
  
$$= \frac{\alpha}{r!} \left(m \sum_{i=1}^{m} A^{2}\right)^{1/2} = \frac{\alpha}{r!} mA = \frac{\alpha}{r!} (mA^{1/(r+1)})^{r+1}m^{-r}$$
  
$$= \frac{\alpha}{r!} \left(\sum_{i=1}^{m} h_{i} |f^{(r)}(\xi_{i})|^{1/(r+1)}\right)^{r+1} m^{-r} \approx \frac{\alpha}{r!} \left(\int_{a}^{b} |f^{(r)}(x)|^{1/(r+1)} dx\right)^{r+1} m^{-r}.$$

To get an asymptotic formula for  $\int_a^b (f - L_{m,r}f)(x) dx$  we use the analysis done in Sect. 2. If  $\beta = 0$  then the integral decreases faster than  $m^{-r}$ . Let  $\beta \neq 0$ . Then

$$\int_{a}^{b} (f - L_{m,r}f)(x) \, \mathrm{d}x \approx \frac{\beta}{r!} \sum_{i=1}^{m} h_{i}^{r+1} f^{(r)}(\xi_{i}) = \frac{\beta}{r!} (m_{+} - m_{-}) A,$$

where  $\xi_i$ s are as in (4.4), and  $m_+$  and  $m_-$  are the numbers of indexes *i* for which  $f^{(r)}(\xi_i) \ge 0$  and  $f^{(r)}(\xi_i) < 0$ , respectively. Let

$$D_+ = \{x \in [a, b] : f^{(r)}(x) \ge 0\}, \qquad D_- = \{x \in [a, b] : f^{(r)}(x) < 0\}.$$

Since  $A \approx \|f^{(r)}\|_{L^{1/(r+1)}(a,b)} m^{-(r+1)}$  and  $m_+ A^{1/(r+1)} \approx \int_{D_+} |f^{(r)}(x)|^{1/(r+1)} dx$ , we have

$$m_{+} \approx \frac{\int_{D_{+}} |f^{(r)}(x)|^{1/(r+1)} \mathrm{d}x}{\int_{D_{+}\cup D_{-}} |f^{(r)}(x)|^{1/(r+1)} \mathrm{d}x}, \qquad m_{-} \approx \frac{\int_{D_{-}} |f^{(r)}(x)|^{1/(r+1)} \mathrm{d}x}{\int_{D_{+}\cup D_{-}} |f^{(r)}(x)|^{1/(r+1)} \mathrm{d}x}.$$

Thus

$$\int_{a}^{b} (f - L_{m,r}f)(x) \, \mathrm{d}x \approx \frac{\beta}{r!} \left( \frac{\int_{a}^{b} |f^{(r)}(x)|^{1/(r+1)} \mathrm{sgn} f^{(r)}(x) \, \mathrm{d}x}{\int_{a}^{b} |f^{(r)}(x)|^{1/(r+1)} \mathrm{d}x} \right) \|f^{(r)}\|_{L^{1/(r+1)}(a,b)} \, m^{-r}$$

provided  $\int_a^b |f^{(r)}(x)|^{1/(r+1)} \operatorname{sgn} f^{(r)}(x) dx \neq 0$ , and otherwise the convergence is faster than  $m^{-r}$ .

Our analysis above shows that if *m* and *n* are chosen as in (2.5) and (2.6) then the error of the described algorithm asymptotically equals (as  $N \to +\infty$ )

$$c_r \sqrt{\alpha^2 - \beta^2 \left(\frac{\int_a^b |f^{(r)}(x)|^{1/(r+1)} \operatorname{sgn} f^{(r)}(x) \, \mathrm{d}x}{\int_a^b |f^{(r)}(x)|^{1/(r+1)} \, \mathrm{d}x}\right)^2 \left(\int_a^b |f^{(r)}(x)|^{1/(r+1)} \, \mathrm{d}x\right)^{r+1} N^{-(r+1/2)}.$$
(4.5)

The factor at  $\beta^2$  in (4.5) can be easily replaced by 1 with the help of stratified sampling. Indeed, instead of randomly sampling *n* times with density (4.3) on the whole interval [a, b], one can apply the same sampling strategy independently on *k* groups  $G_j$  of subintervals. Each group consists of s = m/k subintervals,

$$G_j = \bigcup_{\ell=1}^s I_{(j-1)s+\ell}, \qquad 1 \le j \le k,$$

and the number of samples for each  $G_j$  equals n/k. As in the algorithm  $\overline{M}_{N,r}^*$ , we combine  $k = k_N$  and N in such a way that  $k_N \to +\infty$  and  $k_N/N \to 0$  as  $N \to \infty$ . Then the total number of points used in each  $G_j$  is  $N_j = N/k$ . Denoting

$$C_j = \left(\int_{G_j} \left|f^{(r)}(x)\right|^{1/(r+1)} \mathrm{d}x\right)^{r+1} = \left(\frac{1}{k}\int_a^b \left|f^{(r)}(x)\right|^{1/(r+1)} \mathrm{d}x\right)^{r+1}$$

and using the fact that the factor at  $\beta^2$  equals 1 if  $f^{(r)}$  does not change its sign, the error of such an approximation asymptotically equals

$$c_r \sqrt{\alpha^2 - \beta^2} \left( \sum_{j=1}^k \frac{C_j^2}{N_j^{2r+1}} \right)^{1/2} = c_r \sqrt{\alpha^2 - \beta^2} \left( \int_a^b \left| f^{(r)}(x) \right|^{1/(r+1)} \mathrm{d}x \right)^{r+1} N^{-(r+1/2)},$$

as claimed. (Note that  $N_j = N/k$  minimize the sum  $\sum_{j=1}^k C_j^2 N_j^{-(2r+1)}$  with respect to  $\sum_{j=1}^k N_j = N$ ; compare with the analysis in Sect. 3.)

Thus we obtained exactly the same error formula as in Theorem 3.1 for  $\overline{M}_{Nr}^*$ .

It remains to show a feasible construction of a nested partition that is close to the one satisfying (4.2). To that end, we utilize the iterative method presented in [11], where the  $L^p$  error of piecewise Lagrange interpolation is examined.

We first consider the case when

$$f^{(r)} > 0 \quad \text{or} \quad f^{(r)} < 0.$$
 (4.6)

In the following construction, we use a priority queue  $\mathscr{S}$  whose elements are subintervals. For each subinterval  $I_i$  of length  $h_i$ , its priority is given as

$$p_f(I_i) = h_i^{r+1} |d_i|,$$

where  $d_i$  is the divided difference (3.2). In the following pseudocode, insert( $\mathscr{S}$ , I) and  $I := \text{extract}_max(\mathscr{S})$  implement correspondingly the actions of inserting an interval to  $\mathscr{S}$ , and extracting from  $\mathscr{S}$  an interval with the highest priority.

```
algorithm PARTITION
```

```
\begin{aligned} \mathscr{S} &= \emptyset; \text{ insert}(\mathscr{S}, [a, b]); \\ \text{for } k &= 2:m \\ [l, r] &= \text{extract}_{\max}(\mathscr{S}); \\ c &= (l+r)/2; \\ \text{insert}(\mathscr{S}, [l, c]); \text{ insert}(\mathscr{S}, [c, r]); \\ \text{endfor} \end{aligned}
```

After execution, the elements of  $\mathscr{S}$  form a partition into *m* subintervals  $I_i$ . Note that if the priority queue is implemented through a *heap* then the running time of PARTITION is proportional to *m* log *m*.

Denote by  $\overline{M}_{N,r}^{**}$  the corresponding algorithm that uses the above nested partition and density (4.3), and N is related to the number *m* of subintervals and the number *n* of random samples as in (2.5) and (2.6). We want to see how much worse is this algorithm than that using the (not nested) partition (4.2).

Let  $A = (A_1, A_2, ..., A_m)$  with

$$A_i = p_f(I_i) r! = h_i^{r+1} \left| f^{(r)}(\omega_i) \right|, \qquad \omega_i \in I_i,$$

and  $||A||_p = \left(\sum_{i=1}^m A_i^p\right)^{1/p}$ . For the corresponding piecewise Lagrange approximation  $L_{m,r}f$  and density  $\rho$  given by (4.3) we have

$$\int_{a}^{b} \frac{(f - L_{m,r}f)^{2}(x)}{\rho(x)} dx - \left(\int_{a}^{b} (f - L_{m,r}f)(x) dx\right)^{2}$$
  
$$\approx \frac{1}{(r!)^{2}} \left(\alpha^{2}m \sum_{i=1}^{m} A_{i}^{2} - \beta^{2} \left(\sum_{i=1}^{m} A_{i}\right)^{2}\right) = \frac{1}{(r!)^{2}} \left(\alpha^{2}m \|A\|_{2}^{2} - \beta^{2} \|A\|_{1}^{2}\right).$$

We also have  $\left(\int_{a}^{b} |f^{(r)}(x)|^{1/(r+1)} dx\right)^{r+1} \approx \left(\sum_{i=1}^{m} A_{i}^{1/(r+1)}\right)^{r+1} = ||A||_{\frac{1}{r+1}}.$ Hence

$$\begin{split} \sqrt{\mathbb{E}(Sf - \overline{M}_{N,r}^{**}f)^2} &\approx \left(\alpha^2 m \|A\|_2^2 - \beta^2 \|A\|_1^2\right)^{1/2} n^{-1/2} m^{-r} \\ &\approx K_{m,r}(A) \, c_r \, \sqrt{\alpha^2 - \beta^2} \left(\int_a^b \left|f^{(r)}(x)\right|^{1/(r+1)}\right)^{r+1} N^{-(r+1/2)}, \end{split}$$

where

$$K_{m,r}(A) = \frac{\sqrt{\kappa_{\alpha}^2 m \|A\|_2^2 - \kappa_{\beta}^2 \|A\|_1^2}}{\|A\|_{\frac{1}{r+1}}} m^r, \qquad \kappa_{\alpha} = \frac{\alpha}{\sqrt{\alpha^2 - \beta^2}}, \quad \kappa_{\beta} = \frac{\beta}{\sqrt{\alpha^2 - \beta^2}}.$$

🖄 Springer

Observe that halving an interval results in two subintervals whose priorities are asymptotically (as  $m \to +\infty$ )  $2^{r+1}$  times smaller than the priority of the original interval. This means that  $K_{m,r}(A)$  is asymptotically not larger than

$$K^{*}(r) = \limsup_{m \to \infty} \max\left\{ K_{m,r}(A) : A = (A_{1}, \dots, A_{m}), \max_{1 \le i, j \le m} \frac{A_{i}}{A_{j}} \le 2^{r+1} \right\}.$$
(4.7)

Thus we obtained the following result.

**Theorem 4.1** If the function f satisfies (4.6) then we have as  $N \to +\infty$  that

$$\sqrt{\mathbb{E}(Sf - \overline{M}_{N,r}^{**}f)^2} \lesssim K^*(r) c_r \sqrt{\alpha^2 - \beta^2} \left( \int_a^b \left| f^{(r)}(x) \right|^{1/(r+1)} \mathrm{d}x \right)^{r+1} N^{-(r+1/2)},$$

where  $K^*(r)$  is given by (4.7).

We numerically calculated  $K^*(r)$  in some special cases. For instance, if the points  $z_i$  in (2.1) are equispaced,  $z_i = (i - 1)/(r - 1)$ ,  $1 \le i \le r$ , then for r = 2, 3, 4, 5, 6 we correspondingly have

$$K^*(r) = 4.250, 3.587, 7.077, 11.463, 23.130,$$

while for any  $z_i$ s satisfying  $\beta = \int_0^1 (z - z_1) \cdots (z - z_r) dz = 0$  we have

$$K^*(r) = 2.138, 3.587, 6.323, 11.463, 21.140.$$

If f does not satisfy (4.6) then the algorithm  $\overline{M}_{N,r}^{**}f$  may fail. Indeed, it may happen that  $p_f(I_i) = 0$  while  $f^{(r)} \neq 0$  in  $I_i$ . Then this subinterval may never be further subdivided. In this case, we can repeat the same construction, but with the modified priority

$$p_f(I_i) = h_i^{r+1} \max\left(|d_i|, \Delta/r!\right),$$

where  $\Delta > 0$ . Then the error is asymptotically upper bounded by

$$K^{*}(r) c_{r} \sqrt{\alpha^{2} - \beta^{2}} \left( \int_{a}^{b} \left| f_{\Delta}^{(r)}(x) \right|^{1/(r+1)} \mathrm{d}x \right)^{r+1} N^{-(r+1/2)},$$

where  $|f_{\Delta}^{(r)}(x)|$  is given by (3.4).

## **5 Numerical experiments**

In this section, we present results of two numerical experiments that illustrate the performance of the nonadaptive Monte Carlo algorithm  $\overline{M}_{N,r}$  and the adaptive algorithms



Fig. 1 Comparison of nonadaptive and adaptive Monte Carlo algorithms together with related asymptotic constants (AC) for r = 2

 $\overline{M}_{N,r}^*$  and  $\overline{M}_{N,r}^{**}$ . Our test integral is

$$\int_0^1 \frac{1}{x+10^{-4}} \, \mathrm{d}x.$$

Since for  $r \in \mathbb{N}$  we have  $(-1)^r f^{(r)} > 0$ , the parameter  $\Delta$  is set to zero.

The three algorithms are verified for r = 2 and r = 4. In both cases, the interpolation nodes are equispaced, i.e., in (2.1) we take

$$z_i = \frac{i-1}{r-1}, \qquad 1 \le i \le r.$$

In addition, for the first adaptive algorithm  $\overline{M}_{N,r}^*$  we take  $k_N = N^{\kappa}$  with  $\kappa = 0.8$ . This exponent was chosen to ensure a trade–off as per our discussion in Sect. 3, and some empirical results. Also, for a fixed N we plot a single output instead of the expected value estimator. Therefore the error fluctuations are visible. For completeness, we also show the asymptotes corresponding to the theoretical errors from Theorems 2.1 and 3.1, and the upper bound from Theorem 4.1. The scale is logarithmic,  $-\log_{10}(\text{error})$  versus  $\log_{10} N$ .

The results for r = 2 are presented in Fig. 1.

As it can be observed, both adaptive algorithms significantly outperform the nonadaptive MC; however, the right asymptotic behaviour of the first adaptive algorithm is visible only for large N.

Similar conclusions can be inferred from validation performed for r = 4, with all other parameters unchanged. We add that the results for N larger than  $10^{4.8}$  are not illustrative any more, since the process is disturbed by a serious reduction of significant digits when calculating divided differences in the partition part (Fig. 2).



Fig. 2 Comparison of nonadaptive and adaptive Monte Carlo algorithms together with related asymptotic constants (AC) for r = 4

Notably, both adaptive algorithms attain their asymptotic errors, but this is not the case for nonadaptive MC for which the output is not stable. Initially, the first adaptive algorithm does not leverage additional sampling since for all intervals  $I_i$  we have  $N_i/(2r + 1) < 1$ . The Monte Carlo adjustments are visible only for  $N \ge 10^3$  and the error tends to the theoretical asymptote.

In conclusion, the numerical experiments confirm our theoretical findings and, in particular, superiority of the second adaptive algorithm  $\overline{M}_{N,r}^{**}$ .

#### 6 Automatic Monte Carlo integration

We now use the results of Sect. 4 for automatic Monte Carlo integration. The goal is to construct an algorithm that for given  $\varepsilon > 0$  and  $0 < \delta < 1$  returns an  $\varepsilon$ -approximation to the integral *Sf* with probability at least  $1 - \delta$ , asymptotically as  $\varepsilon \to 0^+$ . To that end, we shall use the approximation  $\overline{M}_{N,r}^{**}f$  with *N* determined automatically depending on  $\varepsilon$  and  $\delta$ .

Let  $X_i$  for  $1 \le i \le n$  be independent copies of the random variable

$$X = S(f - L_{m,r}f) - \frac{(f - L_{m,r}f)(t)}{\rho(t)}, \quad t \sim \mu_{\rho},$$

where  $L_{m,r}$ , *n* and  $\rho$  are as in  $\overline{M}_{N,r}^{**} f$ . Then  $\mathbb{E}(X) = 0$  and

$$Sf - \overline{M}_{N,r}^{**}f = \frac{X_1 + X_2 + \dots + X_n}{n}.$$

By Hoeffding's inequality [5] we have

$$\operatorname{Prob}\left(\left|Sf - \overline{M}_{N,r}^{**}f\right| > \varepsilon\right) \leq 2 \exp\left(\frac{-\varepsilon^2 n}{2 B_m^2}\right),$$

where  $B_m = \max_{a \le t \le b} |X(t)|$ . Hence we fail with probability at most  $\delta$  if

$$\frac{\varepsilon^2 n}{2B_m^2} \ge \ln \frac{2}{\delta}.$$
(6.1)

Now we estimate  $B_m$ . Let  $\lambda = ||P||_{L^{\infty}(0,1)} = \max_{0 \le t \le 1} |P(t)|$ , and

$$\mathscr{L}_r(f) = \left(\int_a^b \left|f_{\Delta}^{(r)}(x)\right|^{1/(r+1)} \mathrm{d}x\right)^{r+1},$$

where  $\Delta = 0$  if  $f^{(r)} > 0$  or  $f^{(r)} < 0$ , and  $\Delta > 0$  otherwise. Let  $A = (A_1, A_2, \dots, A_m)$  with

$$A_i = h_i^{r+1} \max_{x \in I_i} |f^{(r)}(x)|, \quad 1 \le i \le m,$$

where, as before,  $\{I_i\}_{i=1}^m$ , is the partition used by  $\overline{M}_{N,r}^{**} f$  and  $h_i$  is the length of  $I_i$ . Since  $||A||_{\frac{1}{r+1}} = \left(\sum_{i=1}^m A_i^{1/(r+1)}\right)^{r+1} \leq \mathscr{L}_r(f)$ , for  $x \in I_i$  we have

$$\frac{\left|f(x)-L_{m,r}f(x)\right|}{\rho(x)} \leq \frac{\lambda}{r!} m A_i \lesssim \frac{\lambda}{r!} \left(\frac{m \|A\|_{\infty}}{\|A\|_{\frac{1}{r+1}}}\right) \mathscr{L}_r(f) \lesssim 2^{r+1} \frac{\lambda}{r!} \mathscr{L}_r(f) m^{-r}.$$

We have the same upper bound for  $S(f - L_{m,r}f)$  since by mean-value theorem

$$S(f - L_{m,r}f) = \int_{a}^{b} \frac{(f - L_{m,r}f)(x)}{\rho(x)} \,\rho(x) \,\mathrm{d}x = \frac{(f - L_{m,r}f)(\xi)}{\rho(\xi)}, \quad \xi \in [a, b].$$

Hence

$$B_m \lesssim 2^{r+2} \frac{\lambda}{r!} \mathscr{L}_r(f) m^{-r}.$$

Using the above inequality and the fact that  $\sqrt{n} m^r \approx N^{r+1/2}/(c_r r!)$  with  $c_r$  given by (2.7), we get

$$\frac{\varepsilon^2 n}{2B_m^2} \gtrsim \left(\frac{\varepsilon N^{r+1/2}}{\hat{c}_r \,\mathscr{L}_r(f)}\right)^2, \quad \text{where} \quad \hat{c}_r = 2^{r+5/2} \lambda c_r.$$

The last inequality and (6.1) imply that we fail to have error  $\varepsilon$  with probability at most  $\delta$  for

$$N \gtrsim \left(\hat{c}_r \,\mathscr{L}_r(f) \,\frac{\sqrt{\ln(2/\delta)}}{\varepsilon}\right)^{\frac{1}{r+1/2}}, \quad \text{as} \quad \varepsilon \to 0^+.$$
(6.2)

Now the question is how to obtain the random approximation  $\overline{M}_{N,r}^{**} f$  for N satisfying (6.2).

One possibility is as follows. We first execute the iteration **for** in the algorithm PARTITION of Sect. 4 for k = 2 : m, where *m* satisfies  $\lim_{\epsilon \to 0^+} m \epsilon^{\frac{1}{r+1/2}} = 0$ , e.g.,

$$m = \left\lfloor \left( \frac{\sqrt{\ln(2/\delta)}}{\varepsilon} \right)^{\frac{1}{r+1}} \right\rfloor.$$

Let  $\{I_i\}_{i=1}^m$  be the obtained partition. Then we replace  $\mathscr{L}_r(f)$  in (6.2) by its asymptotic equivalent

$$\widetilde{\mathscr{L}}_{r}(f) = \left(\sum_{i=1}^{m} p_{f}(I_{i})^{\frac{1}{r+1}}\right)^{r+1},$$
(6.3)

set

$$N_{\varepsilon} = \left\lfloor \left( \hat{c}_r \, \widetilde{\mathscr{L}}_r(f) \, \frac{\sqrt{\ln(2/\delta)}}{\varepsilon} \right)^{\frac{1}{r+1/2}} \right\rfloor,\tag{6.4}$$

and continue the iteration for k = m + 1:  $m_{\varepsilon}$ , where  $m_{\varepsilon}$  is the number of subintervals corresponding to  $N_{\varepsilon}$ . Finally, we complete the algorithm by  $n_{\varepsilon}$  random samples.

Denote the final randomized approximation by  $\mathscr{A}_{\varepsilon,\delta}f$ . Then we have  $\mathscr{A}_{\varepsilon,\delta}f = \overline{M}_{N_{\varepsilon,r}}^{**}f$  and

$$\operatorname{Prob}(|Sf - \mathscr{A}_{\varepsilon,\delta}f| > \varepsilon) \lesssim \delta, \quad \text{as} \quad \varepsilon \to 0^+.$$

A disadvantage of the above algorithm is that it uses a priority queue and therefore its total running time is proportional to  $N \log N$ . It turns out that by using recursion the running time can be reduced to N.

A crucial component of the algorithm with the running time proportional to N is the following recursive procedure, in which  $\mathcal{S}$  is a set of intervals.

**procedure** AUTO (f, a, b, e) **if**  $p_f([a, b]) \le e$ insert( $\mathscr{S}, [a, b]$ ); **else**  c := (a + b)/2;AUTO(f, a, c, e); AUTO(*f*, *c*, *b*, *e*); endif

Similarly to  $\mathscr{A}_{\varepsilon,\delta}$ , the algorithm consists of two steps. First AUTO is run for  $e = \varepsilon'$  satisfying  $\varepsilon' \to 0^+$  and  $\varepsilon/\varepsilon' \to 0^+$  as  $\varepsilon \to 0^+$ , e.g.,

 $\varepsilon' = \varepsilon^{\kappa}$ , where  $0 < \kappa < 1$ .

Then  $\mathscr{L}_r(f)$  in (6.2) is replaced by  $\widetilde{\mathscr{L}}_r(f)$  given by (6.3), and  $N_{\varepsilon}$  found from (6.4). The recursion is resumed with the target value  $e = \varepsilon''$ , where

$$\varepsilon'' = \widetilde{\mathscr{L}}_r(f) \, m_{\varepsilon}^{-(r+1)}.$$

The algorithm is complemented by the corresponding  $n_{\varepsilon}$  random samples.

Observe that the number m'' of subintervals in the final partition is asymptotically at least  $m_{\varepsilon}$ . Indeed, for any function  $g \in C^r([a, b])$  with  $g^{(r)}(x) = |f_{\Delta}^{(r)}(x)|$  we have  $\mathscr{L}_r(g) = \mathscr{L}_r(f)$  and

$$\frac{\gamma}{r!} \widetilde{\mathscr{L}}_{r}(f) (m'')^{-r} \lesssim \|g - L_{m'',r}g\|_{L^{1}(a,b)} \approx \frac{\gamma}{r!} \sum_{i=1}^{m''} (h_{i}'')^{r+1} g^{(r)}(\xi_{i})$$
$$\approx \frac{\gamma}{r!} \sum_{i=1}^{m''} p_{f}(I_{i}'') \lesssim \frac{\gamma}{r!} m'' \varepsilon'',$$

where the first inequality above follows from Proposition 2 of [11]. This implies

$$m'' \gtrsim \left(\frac{\widetilde{\mathscr{L}}_r(f)}{\varepsilon''}\right)^{\frac{1}{r+1}} \approx m_{\varepsilon},$$

as claimed.

Denote the resulting approximation by  $\mathscr{A}_{\varepsilon,\delta}^* f$ . Observe that its running time is proportional to  $N_{\varepsilon}$  since recursion can be implemented in linear time.

Theorem 6.1 We have

$$\operatorname{Prob}(|Sf - \mathscr{A}_{\varepsilon,\delta}^* f| > \varepsilon) \lesssim \delta, \quad as \ \varepsilon \to 0^+.$$

Now we present outcomes of the second automatic procedure  $\mathscr{A}^*_{\varepsilon,\delta}$  for the test integral

$$\int_0^1 \cos\left(\frac{100\,x}{x+10^{-4}}\right) \mathrm{d}x.$$
(6.5)

Although the derivatives fluctuate and nullify many times in this case, we take  $\Delta = 0$ . We confront the outcomes for r = 2 and r = 4. In each case, we compute the

Table 1         Performance of the second automatic algorithm for the integral (6.5)		ε	δ	Κ	$N_{\mathcal{E}}$	e <sub>max</sub>
	r = 2	$10^{-3}$	0.05	10,000	3092	$4.5 \cdot 10^{-5}$
	r = 4	$10^{-3}$	0.05	10,000	811	$5.5 \cdot 10^{-6}$

number of breaches (i.e. when the absolute error is greater than  $\varepsilon = 10^{-3}$ ) based on  $K = 10\,000$  independent executions of the code (Table 1). We also take  $\varepsilon' = \varepsilon^{1/2}$ . In our testing, we expect the empirical probability of the breach to be less than  $\delta = 0.05$ . For completeness, we also present the maximum error from all executions together with obtained  $N_{\varepsilon}$ .

Note that in both cases we did not identify any exceptions. The magnitude of the maximum errors indicate a serious overestimation of  $N_{\varepsilon}$ , but the results are satisfactory given the upper bound estimate of Theorem 6.1.

Acknowledgements The work of L. Plaskota and P. Przybyłowicz was partially supported by the National Science Centre, Poland, under project 2017/25/B/ST1/00945.

## Declarations

Conflict of interest All authors declare that they have no conflicts of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

#### Appendix

Below we present a crucial part of the code in the Python programming language, where all the algorithms were implemented. In addition, we provide relevant comments linked to particular fragments of the code.

**Listing 1** Second adaptive algorithm  $\overline{M}_{N,r}^{**}$  - crucial part of the code.

```
def second_adaptive_MC(a, b, N, main_nodes, f, r, node_type = 'uniform'): #(1)
    partial_quad = Decimal('0.0') #(2)

    if node_type == 'uniform': #(3)
        n = int(np.floor((N-1)/(1 + 2 *r)))
        m = int(np.floor((2 * r * (N-1))/(2 * r ** 2 - r - 1)))

    mc_init = MC_samples_nonunif(m, n) #(4)

    #(5)
MnR = Decimal('0.0')
    l = 0

    for i in range(len(main_nodes) - 1):
        h_i = main_nodes[i+1] - main_nodes[i]

        #(6)
```

```
if node type == 'uniform':
        interpol_ix = main_nodes[i] * np.ones(r) + np.multiply(optimalt_equidistant(r),
  h i)
   interpol iy = []
   for s in range(1,r+1):
    interpol_iy.append(f(interpol_ix[s-1]))
   #(7)
   if r == 2:
       SL_mr = Decimal('0.5') * Decimal(h_i) \
                * (Decimal(f(main_nodes[i])) + Decimal(f(main_nodes[i+1])))
   elif r == 4:
       #(8)
   while l < n and mc_init[1] < (i+1):
    mc_point = math.modf(mc_init[1])[0] * h_i + main_nodes[i] #(9)
           = MnR + (Decimal(f(mc_point)) - Decimal(lagrange(interpol_ix, interpol_iy,
        MnR
  mc_point))) \
        Decimal(h i) #(10)
       1 = 1 + 1
   partial_quad = partial_quad + SL_mr #(11)
#(12)
MnR = Decimal(MnR) * Decimal(m) / Decimal(n)
partial_quad = partial_quad + MnR
return partial_quad
```

- The almost optimal partition main\_nodes is derived out of this function in order to save computation time when the trajectories are computed subsequently. Moreover, node\_type argument lets the user insert his own partitions, e.g. those based on Chebyshev polynomials of the second kind.
- (2) In order to minimize errors resulting from (possibly) adding relatively small adjustments to the estimated quadrature value, we use Decimal library. It enables us to increase the precision of intermediate computations, which is now set to 28 digits in decimal system.
- (3) In our case, the interpolating polynomial is based on equidistant mesh including endpoints of a subinterval  $I_i$ . By np we understand the references to NumPy library.
- (4) Initializing the variables which control Monte Carlo adjustments for our quadrature. In particular, 1 stores the number of currently used random points, while we loop through the subintervals.
- (5) The program calculates all interpolation nodes in the interval  $I_i$ . For that reason, the function optimalt\_uniform is executed to provide distinct  $z_1, \ldots, z_r \in [0, 1]$ .
- (6) Depending on the value of *r*, different formulas for (nonadaptive, deterministic) quadrature SL<sub>m,r</sub> are leveraged.
- (7) Below, we calculate the Monte Carlo adjustment on interval  $I_i$ .
- (8) This code yields random points used for. MC\_init function reports them in a from of a number from 0 to m. The integer part points the index i of subinterval, while the fractional part its position within I<sub>i</sub>. Both parameters are sourced by using math.modf function.
- (9) For stability reasons, the coefficients of interpolating polynomial in canonical base are not stored. Therefore, for every point, lagrange function is invoked separately.
- (10) We decided to add  $SL_{m,r}$  for each subinterval and then add the cumulative adjustments. Since latter are usually relatively much smaller than the quadrature values,

this might result in neglecting the actual adjustment values. Please note that Decimal library was also used to address such constraints.

(11) Ultimately, we add Monte Carlo result to the previous approximation.

As it can be observed, the current solution enables the user to insert own interpolation meshes, increase the precision of computations, as well as extend the method to arbitrary regularity  $r \in \mathbb{N}$ .

## References

- 1. Davis, P., Rabinowitz, P.: Methods of Numerical Integration, 2nd edn. Academic Press, New York (1984)
- Goćwin, M.: On optimal adaptive quadratures for automatic integration. BIT Numer. Math. 61, 411– 439 (2021)
- 3. Gonnet, P.: A review of error estimation in adaptive quadrature. ACM Comput. Surv. 44, 1-36 (2012)
- Heinrich, S.: Random approximation in numerical analysis. In: Bierstedt, K.D., et al. (eds.) Proceedings of the Functional Analysis Conference, Essen 1991, pp. 123–171. Marcel Dekker, New York (1993)
- Hoeffding, W.: Probability inequalities for sums of bounded random variables. J. Am. Stat. Assoc. 58, 13–30 (1963)
- Kacewicz, B., Przybyłowicz, P.: Complexity of the derivative-free solution of systems of IVPs with unknown singularity hypersurface. J. Complex. 31, 75–97 (2015)
- Lyness, J.N.: Guidelines for automatic quadrature routines. In: Freeman, C.V. (ed.) Information Processing 71, vol. 2, pp. 1351–1355. North-Holland Publ. (1972)
- Novak, E.: Deterministic and Stochastic Error Bounds in Numerical Analysis. Vol. 1349 of Lecture Notes in Math. Springer, Berlin (1988)
- 9. Novak, E.: On the power of adaption. J. Complex. 12, 199-237 (1996)
- Plaskota, L.: Automatic integration using asymptotically optimal adaptive Simpson quadrature. Numer. Math. 131, 173–198 (2015)
- Plaskota, L., Samoraj, P.: Automatic approximation using asymptotically optimal adaptive interpolation. Numer. Algorithms 89, 277–302 (2022)
- Plaskota, L., Wasilkowski, G.W.: Adaption allows efficient integration of functions with unknown singularities. Numer. Math. 102, 123–144 (2005)
- Plaskota, L., Wasilkowski, G.W., Zhao, Y.: The power of adaption for approximating functions with singularities. Math. Comput. 77, 2309–2338 (2008)
- Przybyłowicz, P.: Adaptive Itô-Taylor algorithm can optimally approximate the Itô integrals of singular functions. J. Comput. Appl. Math. 235, 203–217 (2010)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.