Optimal Stochastic Quadrature Formulas For Convex Functions

Erich Novak and Knut Petras

TR-93-013

March 15, 1993

Abstract

We study optimal stochastic (or Monte Carlo) quadrature formulas for convex functions. While nonadaptive Monte Carlo methods are not better than deterministic methods we prove that adaptive Monte Carlo methods are much better.

Optimal Stochastic Quadrature Formulas For Convex Functions

Erich Novak and Knut Petras

Abstract. We study optimal stochastic (or Monte Carlo) quadrature formulas for convex functions. While nonadaptive Monte Carlo methods are not better than deterministic methods we prove that adaptive Monte Carlo methods are much better.

1. Introduction and Result

For each (finite) quadrature formula, the error in the class of convex functions on, e.g., $[0,1]$ is not uniformly bounded. For the study of optimal quadrature formulas we therefore have to restrict the class of convex functions somehow. The classes

$$F_{uv} = \{f : [0,1] \to \mathbf{R} \mid f \text{ convex}, \ f'_+(0) \geq u, \ f'_-(1) \leq v\},$$

where $v > u$, were studied by Glinkin (1984), Zwick (1988), and Novak (1993). The following is known for these classes.

Fact 1. Let $n \in \mathbf{N}$ and $t_i = (2i-1)/(2n)$. Then the affine and nonadaptive formula

$$Q_n(f) = \frac{1}{16n^2} \cdot (v-u) + \frac{1}{n} \sum_{i=1}^{n} f(t_i) \tag{1}$$

is optimal even in the class of adaptive formulas of the form

$$Q_n(f) = \phi(f(t_1), \ldots, f(t_n)).$$

By adaptive we mean that the knot t_i may depend on the 'already known' function values, i.e., $t_i = t_i(f(t_1), \ldots, f(t_{i-1}))$. The maximal error of Q_n on the class F_{uv} is given by

$$\Delta_{\max}(Q_n) = \frac{v-u}{16n^2}.$$

□

This result has two drawbacks:

1) The optimal formula (1) depends on the class which might be unknown to us. It would be better to have an 'almost optimal' formula which is independent of the parameters u and v. It is easy to show that the linear formula $Q_n(f) = \frac{1}{n}\sum_{i=1}^{n} f(t_i)$ is almost optimal for all u and v.

2) There are convex functions whose one-sided derivatives are not bounded, hence we would prefer larger classes without smoothness assumptions.

Both these drawbacks are resolved by results of Braß (1982) who studied the classes

$$F_1 = \{f : [0,1] \to \mathbf{R} \mid f \text{ convex}, \|f\|_\infty \leq 1\}$$

and

$$\widetilde{F}_1 = \{f : [0,1] \to \mathbf{R} \mid f \text{ convex}, \ \ \max\{|f(0)|,\ |f(1/2)|,\ |f(1)|\} \leq 1\}$$

and proved the following.

Fact 2. Let c_n be defined by

$$c_n = \begin{cases} 2(n^2+2n+1)^{-1} & \text{if } n \text{ is odd and} \\ 2(n^2+2n+2)^{-1} & \text{if } n \text{ is even.} \end{cases}$$

Then the linear formula

(2)

$$Q_n(f) = \begin{cases} c_n \cdot \left(\sum_{i=1}^{\frac{n-1}{2}} 2if(i^2 c_n) + 2if(1-i^2 c_n)\right) + \frac{n}{2} f(1/2) & \text{if } n \text{ is odd and} \\ c_n \cdot \left(\sum_{i=1}^{\frac{n}{2}} 2if(i^2 c_n) + 2if(1-i^2 c_n)\right) & \text{if } n \text{ is even} \end{cases}$$

is optimal for F_1 and also for $\widetilde{F}_1$ in the class of all nonadaptive quadrature formulas. In both cases the maximal error of the optimal formula is given by

$$\Delta_{\max}(Q_n) = c_n. \tag{3}$$

□

We want to make some comments concerning these results:

1) Though adaptive formulas might be slightly better than nonadaptive formulas for F_1 or $\widetilde{F}_1$ it follows from the proofs for the classes F_{uv} that n^{-2} is also the optimal order of convergence for adaptive formulas.

2) The class F_1 seems to be a more natural class compared to $\widetilde{F}_1$. It is, however, much simpler to check the membership $f \in \widetilde{F}_1$ than to check $f \in F_1$.

3) F_1 (and also $\widetilde{F}_1$) is indeed a very general class of convex functions. If f is any convex function on $[0,1]$, then we have $f \cdot \|f\|_\infty^{-1} \in F_1$ such that error estimates on F_1 easily can be used. Hence, if $Q_n(\alpha f) = \alpha Q_n(f)$ for all $\alpha \geq 0$, the order $O(\Delta_{\max}(Q_n))$ of the error does not only hold for $f \in F_1$, but may also be extended to all convex functions on $[0,1]$. The algorithm mentioned in our Theorem below has this property.

4) We should stress that the analysis and the results for the classes F_{uv} and F_1 (or $\widetilde{F}_1$) are quite different. In particular, it is easy to see that for the class F_1 we have

$$\Delta(Q_n) \geq \frac{1}{2n}$$

for all methods using the equidistant knots which are optimal for each F_{uv}.

5) The optimal formula of Braß leads to an error of the order n^{-2} even for very smooth functions, i.e., even for $f = 1$. It is known that the Gaussian formula leads to an error $O(n^{-k})$ for each C^k-function. Therefore it would be interesting to know whether the Gaussian formula also gives the optimal $O(n^{-2})$ error for arbitrary, i.e., nonsmooth, convex functions. This problem was posed by Braß in 1981 and solved by Förster and Petras (1990) who even studied the slightly bigger class

$$F_2 = \{f : [0,1] \to \mathbf{R} \mid f \text{ convex}, \ \ f(0) - 2f(1/2) + f(1) \leq 4\}.$$

Here the constant 4 is chosen in order to have

$$F_1 \subset \widetilde{F}_1 \subset F_2.$$

The class F_2 also can be compared with the F_{uv}, we have

$$F_{uv} \subset F_2 \qquad \text{iff} \qquad v - u \leq 8. \tag{4}$$

Fact 3. The maximal error of the n-point Gaussian formula on the class F_2 is less than $\frac{\pi^2}{3n^2}$, see Petras (1993b).

□

We do not mention more of the known results concerning (deterministic) quadrature formulas, the interested reader should consult the survey Petras (1993a).

□

In this paper we mainly are interested in stochastic or Monte Carlo methods. We do not repeat known results about stochastic methods for other function classes, see Novak (1988, 1992) and Traub, Wasilkowski, Woźniakowski (1988). We present results for methods of the form

$$Q_n^\omega(f) = \phi^\omega(f(t_1^\omega), \ldots, f(t_n^\omega)),$$

i.e., the knots t_i^ω are random variables (which can be chosen nonadaptively or adaptively) and also the ϕ^ω is randomly chosen. Then, as usual, the error of Q_n^ω is defined in a 'worst case stochastic sense',

$$\Delta_{\max}(Q_n^\omega) = \max_{f \in F} E(|\int_0^1 f(t)\,dt - Q_n^\omega(f)|).$$

The following is mentioned in Novak (1993) without a full proof of the upper bound. We remark that the proof of our Theorem also contains this upper bound.

Fact 4. Let Q_n^ω be any nonadaptive stochastic method for the integration problem on the class F_{uv}. Then the lower bound

$$\Delta_{\max}(Q_n^\omega) \geq \frac{1}{128n^2}(v-u)$$

holds. Thus the (stochastic) error of nonadaptive stochastic methods can be only slightly smaller than the worst case error of deterministic methods. Let Q_n^ω be any adaptive stochastic method for the integration problem on the class F_{uv}. Then the lower bound

$$\Delta_{\max}(Q_n^\omega) \geq \frac{\sqrt{2}}{128}(v-u)n^{-5/2}$$

holds. This lower bound gives the optimal order of convergence because we can construct methods with

$$\Delta_{\max}(Q_n^\omega) = O((v-u)\cdot n^{-5/2}).$$

□

This fact shows that adaptive Monte Carlo methods are much better (in a stochastic sense) than nonadaptive Monte Carlo methods, but nonadaptive Monte Carlo methods are not better than deterministic methods. This is quite different than known results for unit balls of Hölder or Sobolev spaces, but it is similar to the results of Novak (1992) for the class of monotone functions.

Is a similar result also true if we consider one of the larger classes, say, F_2? This is exactly the question that is studied in this paper. The positive answer is contained in our Theorem. Similarly as in the case of deterministic formulas the upper bound cannot be proved using the known results and formulas for the classes F_{uv}.

Theorem. Consider the class F_2. Let Q_n^ω be any nonadaptive stochastic method using n knots. Then the lower bound

$$\Delta_{\max}(Q_n^\omega) \geq \frac{1}{16n^2}$$

holds. Thus the (stochastic) error of nonadaptive stochastic methods can be only slightly smaller than the worst case error of deterministic methods. Let Q_n^ω be any adaptive stochastic method. Then the lower bound

$$\Delta_{\max}(Q_n^\omega) \geq \frac{\sqrt{2}}{16}n^{-5/2}$$

holds. This lower bound gives the optimal order of convergence because we can construct methods with

$$\Delta_{\max}(Q_n^\omega) = O(n^{-5/2}).$$

□

Remark. Our method $(Q_n^\omega)_{n\in\mathbf{N}}$ defined in the proof gives the optimal order $n^{-5/2}$ of convergence on each class of the form $\{f : [0,1] \to \mathbf{R} \mid f \text{ convex}, \ f(0)-2f(1/2)+f(1) \le K\}$ for $K > 0$. Hence, this order of convergence holds for each convex function on $[0,1]$.

□

2. Proof

The lower bounds are a simple consequence from Fact 4 and (4). A more detailed analysis would give these lower bounds also for the class F_1. Therefore the interesting thing is whether the upper bound $n^{-5/2}$ can be proved for the larger class F_2. We describe an adaptive Monte Carlo method with the optimal order of convergence. Let $m \in \mathbf{N}$ and $m \ge 5$.

First step. Define $k_m = 2(m-1)^{-2}$ and

$$a_i = (i-1)^2 k_m = 1 - a_{m+1-i}, \qquad i = 1, \dots, \left\lfloor \frac{m+1}{2} \right\rfloor$$

and compute $f(a_i)$ for each i. This step certainly is nonadaptive and deterministic.

Second step. In each interval $[a_i, a_{i+1}]$ we determine the area F_i between the pointwise largest and smallest convex function (in the given interval) that fits the data at the points a_i. For this purpose, let f_i be the linear interpolant of f at a_i and a_{i+1}, and let s_i be its gradient. On $[a_i, a_{i+1}]$, where $2 \le i \le m-2$, the pointwise largest function is f_i, while the smallest function is $\max\{f_{i-1}, f_{i+1}\}$. Defining $d_i = a_{i+1} - a_i$, the area F_i is now given by the formula

$$\begin{aligned} F_i &= \frac{d_i^2}{2} \frac{(s_{i+1}-s_i)(s_i - s_{i-1})}{s_{i+1}-s_{i-1}} \le \frac{d_i^2}{8}(s_{i+1}-s_{i-1}) \\ &= \frac{d_i^2}{8} \left\{ (a_{i+2}-a_i)\mathrm{dvd}(a_i, a_{i+1}, a_{i+2})[f] \right. \\ &\qquad \left. + (a_{i+1}-a_{i-1})\mathrm{dvd}(a_{i-1}, a_i, a_{i+1})[f] \right\} =: G_i(f), \end{aligned}$$

where $\mathrm{dvd}(x,y,z)$ denotes the divided difference with nodes x, y and z. The modifications on the boundary intervals are obvious and we obtain there,

$$F_1 = \frac{d_1^2}{2}(s_2 - s_1) = \frac{d_1^2}{2}(a_3 - a_1)\mathrm{dvd}(a_1, a_2, a_3)[f] =: G_1(f),$$

as well as

$$F_{m-1} = \frac{d_{m-1}^2}{2}(a_m - a_{m-2})\mathrm{dvd}(a_{m-2}, a_{m-1}, a_m)[f] =: G_{m-1}(f).$$

Let now G be the sum of all the functionals G_i. Since G vanishes for all polynomials of degree less than or equal to 1, its second Peano kernel exists. This Peano kernel is a polygonal line with corners

$$(0,0), \quad \left(a_2, \frac{4d_1^2 + d_2^2}{8}\right), \quad \left(a_i, \frac{d_{i-1}^2 + d_i^2}{8}\right), \quad \left(a_{m-1}, \frac{d_{m-2}^2 + 4d_{m-1}^2}{8}\right), \quad (1,0),$$

where $i = 3, 4, \ldots, m-2$, and it therefore lies between the x-axis and the hat function h on $[0, 1]$, given by

$$h(x) = \frac{13}{4(m-1)^2}\left(x - (2x-1)_+\right).$$

By an argument used in Förster and Petras (1990, proof of Theorem 1), we see that if the second Peano kernel of any linear functional G lies completely below the second Peano kernel of a linear functional L, then $G[f] \leq L[f]$ holds for each convex function f on $[0, 1]$. The second Peano kernel of $L = \frac{13}{8}(m-1)^{-2}\mathrm{dvd}(0, \frac{1}{2}, 1)$ is h and, by definition, each function $f \in F_2$ satisfies $\mathrm{dvd}(0, \frac{1}{2}, 1)[f] \leq 8$. Thus, we obtain that

$$F := \sum_{i=1}^{m-1} F_i \leq \frac{13}{(m-1)^2}.$$

In the case

$$\frac{k-1}{m-1}F < F_i \leq \frac{k}{m-1}F$$

we divide the interval $[a_i, a_{i+1}]$ in k smaller intervals of the same length. In this way we obtain $r(f) \leq 2m - 3$ of the smaller intervals I_k. This step is adaptive but still deterministic. We do not compute any function values in this step. Observe that the respective area of each of the smaller intervals is bounded by $\frac{13}{(m-1)^3}$.
Third step. Now we take t_k randomly according to the normalized Lebesgue measure in I_k; the random variables t_k are independent. We also define $t'_k \in I_k$ by the condition that $(t_k + t'_k)/2$ is the midpoint of I_k. We finally take the adaptive Monte Carlo method

$$Q_n^\omega(f) = \sum_{k=1}^{r(f)} \frac{\lambda(I_k)}{2}(f(t_k) + f(t'_k)). \tag{5}$$

Observe that we use $m + 2r(f)$ function values, so we can put $n = m + 4m - 6$. We have to prove that this method has an error of the order $O(n^{-5/2})$ or $O(m^{-5/2})$, respectively.
The random variable

$$\phi_k(\omega) = \frac{\lambda(I_k)}{2}(f(t_k) + f(t'_k))$$

is an unbiased estimator of $\int_{I_k} f(x)\,dx$ with

$$\left|\phi_k(\omega) - \int_{I_k} f(x)\,dx\right| \leq \widetilde{F}_k \leq \frac{13}{(m-1)^3}$$

for all ω. Here $\widetilde{F}_k$ is the respective area in the smaller interval. In particular, we obtain

$$\sigma^2(\phi_k) \leq \frac{169}{(m-1)^6}$$

for the variance of ϕ_k. The ϕ_k are independent and therefore we get

$$\Delta^2_{\max}(Q^\omega_n) \leq \sum_{k=1}^{r(f)} \sigma^2(\phi_k) \leq (2m-3) \cdot \frac{169}{(m-1)^6} = O(m^{-5}).$$

This is just what we claimed.

□

Acknowledgment

One of the authors (E. N.) is supported by a Heisenberg scholarship of the DFG. This paper was written while he was visiting the International Computer Science Institute (ICSI) in Berkeley.

References

[1] H. Braß (1982): Zur Quadraturtheorie konvexer Funktionen. In: G. Hämmerlin (ed.), Numerical Integration, pp. 34–47, ISNM **57**.

[2] K.-J. Förster, K. Petras (1990): On a problem proposed by H. Braß concerning the remainder term in quadrature for convex functions. Numer. Math. **57**, 763–777.

[3] I. A. Glinkin (1984): Best quadrature formula in the class of convex functions. Math. Notes **35**, 368–374. [Russian original: Mat. Zametki **35**, 697–707]

[4] E. Novak (1988): Deterministic and stochastic error bounds in numerical analysis. Lecture Notes in Mathematics **1349**, Springer.

[5] E. Novak (1992): Quadrature formulas for monotone functions. Proc. Amer. Math. Soc. **115**, 59–68.

[6] E. Novak (1993): Quadrature formulas for convex classes of functions. In: H. Braß and G. Hämmerlin (eds.), Numerical Integration IV.

[7] K. Petras (1993a): Quadrature theory of convex functions. In: H. Braß and G. Hämmerlin (eds.), Numerical Integration IV.

[8] K. Petras (1993b): Gaussian quadrature formulae – second Peano kernels, nodes weights and Bessel functions. To appear in Calcolo.

[9] J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski (1988): Information-based Complexity. Academic Press.

[10] D. Zwick (1988): Optimal quadrature for convex functions and generalizations. In: H. Braß and G. Hämmerlin (eds.), Numerical Integration III, ISNM **85**, 310–315.

Erich Novak
Mathematisches Institut
Universität Erlangen-Nürnberg
Bismarckstr. 1 1/2
W-8520 Erlangen
Germany
novak@mi.uni-erlangen.de

Knut Petras
Institut für Angewandte Mathematik
Technische Universität Braunschweig
Pockelsstr. 14
W-3300 Braunschweig
Germany
i1040214@dbstu1.rz.tu-bs.de