Bulletin of L.N. Gumilyov Eurasian National University. Mathematics, computer science, mechanics series, 2026, Vol. 156, №3, P.6-16
http://bulmathmc.enu.kz, E-mail: vest_math@enu.kz

Article
IRSTI: 27.25.19

HOW USEFUL ARE ADAPTIVE ALGORITHMS FOR FUNCTION APPROXIMATION?

L. Plaskota
University of Warsaw, S. Banach Str., 2, 02-097, Warsaw, Poland
(E-mail: leszekp@mimuw.edu.pl)

Abstract. Adaptive algorithms are widely used to solve numerical problems, with Newton's method for finding zeros and adaptive quadratures for numerical integration being prominent examples. In this paper, we review selected results concerning the application of adaptive algorithms to the problem of approximating scalar functions. Allowable algorithms utilize function values, which can be collected adaptively or nonadaptively. The worst case and asymptotic settings of *information based complexity* theory are considered. In both settings, adaptive algorithms do not provide better convergence rates than nonadaptive algorithms for smooth functions. The advantage of adaption manifests itself in the approximation of piecewise smooth functions and in the optimization of asymptotic constants. In particular, for piecewise smooth functions, the use of adaption is necessary to restore the optimal convergence rate from the smooth case. Adaptive algorithms can also significantly improve the asymptotic constants compared to nonadaptive algorithms, especially for functions whose corresponding derivatives change drastically.

Keywords: function approximation, adaptive algorithms, Lagramge interpolation, asymptotic constants, piecewise smoothness.

DOI: https://doi.org/10.32523/bulmathenu.2026/3.1

2000 Mathematics Subject Classification: 65Y20, 65D15, 41A25

1. Introduction

An algorithm is *adaptive* if its subsequent steps depend on the results of previous steps. For numerical problems, adaption is most often achieved through the process of collecting information about the current problem instance. If the problem is defined on a set of functions then information may consist of function values, with subsequent points being selected depending on the values at previous points. Adaptive methods are especially efficient and widely used for nonlinear problems. Primary examples are Newton's method for zero finding or the method of steepest descent for finding a minimum. Less obvious is the use of adaption for linear problems. Indeed, there are theoretical results saying that adaption does not help in some problem settings. To give an example, suppose a linear problem is defined on a set that is convex and symmetric about zero, and allowable algorithms use linear evaluations of the problem instance. Then adaptive evaluations give no more than twice smaller worst case error than nonadaptive evaluations, i.e., when they are pre-assigned in advance

[2, 3, 16]. On the other hand, adaptive procedures are widely used for, e.g., numerical integration, and adaptive quadratures form part of numerical software packages [9]. This apparent paradox can be easily explained: while adaptive quadratures yield no benefit in the worst case setting over convex and centrosymmetric sets of functions, they can prove more effective for functions that possess some special properties, or in other problem settings, see [4, 10, 12, 14] and also [7] for a survey on the power of adaption for numerical problems.

In this note, we review some old and recent results on the usefulness of adaptive algorithms in the context of numerical approximation of functions of one variable. For $1 \le p \le \infty$, we consider the L^p-approximation of functions

$$f : [0,1] \to \mathbb{R}$$

belonging to a given set $\mathcal{F} \subset L^p(0,1)$. The way the functions are approximated is taken from *information-based complexity* [17]. That is, we assume that our only initial knowledge is that $f \in \mathcal{F}$; however, we have a possibility of evaluating the values of f at finitely many points. This means that allowable approximations (aka. *algorithms*) are of the form $A_n(f) = \varphi(N_n(f))$, where

$$N_n(f) = \big(f(x_1), f(x_2), \ldots, f(x_n)\big), \qquad x_i \in [0,1],$$

is *information* about f, and $\varphi : \mathbb{R}^n \to L^p(0,1)$ is an arbitrary *reconstruction map*. Information is in general adaptive, which means that the subsequent points x_i are intrinsically chosen depending on the values $f(x_1), \ldots, f(x_{i-1})$. If all x_i are pre-selected, i.e., chosen independently of f, then information is nonadaptive.

An approximation A_n is adaptive if it uses adaptive information. The error of A_n at $f \in \mathcal{F}$ is given as

$$\|f - A_n(f)\|_{L^p(0,1)}.$$

The answer to the question of whether adaptive approximations are better than nonadaptive approximations is ambiguous, as it depends on the formulation of the approximation problem itself, the class of functions being approximated, and the choice of criterion by which different approximations are compared. In this review, we distinguish between the classes of smooth and piecewise smooth functions, and different approximations are compared using the worst case approach and the asymptotic approach. We show that the advantage of using adaption is mainly manifested for piecewise smooth functions and in the optimization of asymptotic constants.

Most of the results are here presented in a simplified form. This primarily applies to the assumption that the function values are evaluated exactly. However, most of the cited papers assume that these values are subject to deterministic or random noise. We refer interested readers to [5, 6, 8, 11].

The content of the paper is as follows. In Section 2 we consider the worst case setting, separately for smooth functions in Sebsection 2.1, and for piecewise smooth functions in Subsection 2.2. In Section 3 we study the asymptotic setting, where we are interested in the convergence rate for smooth and piecewise smooth functions in Subsection 3.1, and in the asymptotic constants in Subsection 3.2.

2. Worst case setting

In this section, we consider the worst case approach, where the error of A_n is defined as its worst performance over a function class.

2.1. Approximation of smooth functions

For $r \in \{0, 1, 2, \ldots\}$ and $0 < \varrho \le 1$, let $C^r_\varrho([0,1])$ be the linear space of functions that are r-times differentiable and $f^{(r)}$ are Hölder continuous with exponent ϱ. We denote the Hölder constant of f by

$$[f]_{r,\varrho} = \sup_{0 \le x < y \le 1} \frac{|f^{(r)}(y) - f^{(r)}(x)|}{|y - x|^\varrho} < \infty.$$

Note that $[f]_{r,\varrho}$ is a semi-norm on $C^r_\varrho([0,1])$. Fuctions in $C^r_\varrho([0,1])$ will be called *smooth*.

The standard approximation of smooth functions relies on piecewise Lagrange interpolation $L_{r,m}(f)$. That is, the interval $[0,1]$ is devided into equal-length subintervals $[(i-1)/m, i/m]$

for $1 \le i \le m$, and on each subinterval a polynomial of degree r interpolating f at $r+1$ points in this subinterval is applied. Then the total number of function evaluations is $n \le m(r+1)$. Observe that the above piecewise Lagrange appoximation is nonadaptive. We have the following well-known upper bound on the error of such an approximation for smooth functions.

Proposition 1. *There exists $C > 0$ such that for all $f \in C^r_\varrho([0,1])$ and all $n \ge 1$*

$$\|f - L_{r,m}(f)\|_{L^p(0,1)} \le C\,[f]_{r,\varrho}\,n^{-(r+\varrho)}.$$

We also have the following lower bound on any approximation A_n.

Theorem 1. *There exists $c > 0$ with the following property. Let $M > 0$. For any $n \ge 1$ and any (adaptive or non-adaptive) approximation A_n one can find $f_n \in C^r_\varrho([0,1])$ such that $[f]_{r,\varrho} = M$ and*

$$\|f_n - A_n(f_n)\|_{L^p(0,1)} \ge c\,[f_n]_{r,\varrho}\,n^{-(r+\varrho)}.$$

The proof uses a rather standard technique that is based on 'bump' functions.

Proof. Let $0 \le x_1 < x_2 < \cdots < x_n \le 1$ be the evaluation points of A_n for $f = 0$. (Since information is adaptive, these points may be different for different functions.) Let $\Psi(t) = \alpha[t(1-t)]_+^{r+1}$ and

$$\psi_i(x) = \Psi\left(\frac{x - x_i}{h_i}\right) h_i^{r+\varrho}, \qquad 0 \le i \le n,$$

where $x_0 = 0$, $x_{n+1} = 1$, and $h_i = x_{i+1} - x_i$. (Here $\psi_0 = 0$ if $x_1 = 0$, and $\psi_n = 0$ if $x_n = 1$.) Finally, we let

$$\psi(x) = \sum_{i=0}^{n} \psi_i(x).$$

We can assume that α in the definition of Ψ is such that $[\psi]_{r,\varrho} = M$. Obviously $\pm\psi \in C^r_\varrho([0,1])$. Since $\psi(x_i) = 0 = -\psi(x_i)$ for all i, the functions ψ and $-\psi$ share the same zero information and therefore $A_n(\psi) = A_n(-\psi)$. Hence, for one of them the error equals at least $\|\psi\|_{L^p(0,1)}$. We now estimate this norm from below.

Observe that

$$\|\psi_i\|_{L^p(x_i,x_{i+1})} = h_i^{r+\varrho}\left(\int_{x_i}^{x_{i+1}} \Psi^p\left(\frac{x-x_i}{h_i}\right)\mathrm{d}x\right)^{1/p} = h_i^{r+\rho+1/p}\|\Psi\|_{L^p(0,1)}$$

(which is also valid for $p = \infty$), and therefore

$$\begin{aligned}\|\psi\|_{L^p(0,1)} &= \left(\sum_{i=0}^{n}\|\psi_i\|^p_{L^p(x_i,x_{i+1})}\right)^{1/p} = \|\Psi\|_{L^p(0,1)}\left(\sum_{i=0}^{n} h_i^{p(r+\varrho)+1}\right)^{1/p}\\ &\ge \|\Psi\|_{L^p(0,1)}\left(\frac{1}{n+1}\right)^{r+\varrho}.\end{aligned}$$

It remains to bound $\|\Psi\|_{L^p(0,1)}$ from below by $[\psi]_{r,\varrho}$. If $x_i \le x < y \le x_{i+1}$ then

$$\begin{aligned}|\psi^{(r)}(y) - \psi^{(r)}(x)| &= h_i^\varrho\left|\Psi^{(r)}\left(\frac{y-x_i}{h_i}\right) - \Psi^{(r)}\left(\frac{x-x_i}{h_i}\right)\right|\\ &\le h_i^\varrho\left|\frac{y-x_i}{h_i} - \frac{x-x_i}{h_i}\right|^\varrho [\Psi]_{r,\varrho} = |y-x|^\varrho\,[\Psi]_{r,\varrho}.\end{aligned}$$

On the other hand, if $x \in [x_i, x_{i+1}]$ and $y \in [x_j, x_{j+1}]$ with $i < j$ then

$$\begin{aligned}&|\psi^{(r)}(y) - \psi^{(r)}(x)|\\ &\le |\psi_i^{(r)}(x_i) - \psi_i^{(r)}(x)| + \sum_{s=i+1}^{j-1}|\psi_s^{(r)}(x_{s+1}) - \psi_s^{(r)}(x_s)| + |\psi_j^{(r)}(y) - \psi_j^{(r)}(x_j)|\\ &\le (|y-x_j|^\varrho + |x_i - x|^\varrho)\,[\Psi]_{r,\varrho} \le 2^{1-\varrho}|y-x|^\varrho[\Psi]_{r,\varrho}.\end{aligned}$$

Hence $[\psi]_{r,\varrho} \le 2^{1-\varrho}[\Psi]_{r,\varrho}$ and therefore

$$\|\Psi\|_{L^p(0,1)} = \frac{\|\Psi\|_{L^p(0,1)}}{[\Psi]_{r,\varrho}}\,[\Psi]_{r,\varrho} \ge \left(\frac{\|\Psi\|_{L^p(0,1)}}{2^{1-\varrho}[\Psi]_{r,\varrho}}\right)[\psi]_{r,\varrho}.$$

The theorem follows with $f_n = \psi$ or $f_n = -\psi$, and

$$c = \frac{\|\Psi\|_{L^p(0,1)}}{2^{r+1}[\Psi]_{r,\varrho}}.$$

□

Proposition 1 and Theorem 1 yield the following worst case result. Let

$$\mathcal{F}_{r,\varrho} = \big\{f \in C^r_\varrho([0,1]) :\ [f]_{r,\varrho} \le 1\big\}.$$

Corollary 1. *We have*

$$e_n^{\rm wor}(\mathcal{F}_{r,\varrho}) := \inf_{A_n} \sup_{f\in\mathcal{F}_{r,\varrho}} \|f - A_n(f)\| \asymp n^{-r+\varrho}.$$

Moreover, the above convergence rate of the worst case error is achieved by the piecewise Lagrange interpolation $L_{r,m}$, that is based on uniform partition of the domain $[0,1]$. [1]

Thus adaption does not help for approximation of smooth functions if different approximations are compared with respect to the convergence rate of their worst case errors. [2]

2.2. Approximation of piecewise smooth functions

The situation changes if we switch to piecewise smooth functions. We consider the space $\widehat{C}^r_\varrho([0,1])$ consisting of functions of the form

$$f(x) = \begin{cases} g_f(T - s_f + x), & 0 \le x < s_f, \\ g_f(x - s_f), & s_f \le x \le T, \end{cases}$$

for some $g_f \in C^r_\varrho([0,1])$ and $s_f \in (0,1]$. That is, $f \in \widehat{C}^r_\varrho([0,1])$ iff f is smooth except at one singular point s_f at which f is right-continuous. Since $f^{(j)}(0^+) = f^{(j)}(1^-)$ for $0 \le j \le r$ (unless $s_f = 1$), we view $\widehat{C}^r_\varrho([0,1])$ as the space of 1-periodic functions on $\mathbb{R}$. We denote by

$$\Delta_f^{(j)} = f^{(j)}(s_f^+) - f^{(j)}(s_f^-) = g^{(j)}(0^+) - g^{(j)}(1^-), \quad 0 \le j \le r,$$

the discontinuity jumps of the successive derivatives of f at s_f. We stress that our initial knowledge about f is only that it is a member of $\widehat{C}^r_\varrho([0,1])$; hence, the breakpoint s_f and the jumps $\Delta_f^{(j)}$ are unknown. For $f \in \widehat{C}^r_\varrho([0,1])$ we set $[f]_{r,\varrho} := [g_f]_{r,\varrho}$.

Lemma 1. *Let $1 \le p \le \infty$ and $r \ge 1$. For any approximation A_n we have*

$$\sup_{\{f\in\widehat{C}^r_\varrho([0,1]):\ [f]_{r,\varrho}=0\}} \|f - A_n(f)\|_{L^p(0,1)} = \infty.$$

Proof. Let $S(M) \subset \widehat{C}^r_\varrho([0,1])$ be the family of functions f_s parameterized by $s \in [0,1)$,

$$f_s(x) = M\big(x\mathbf{1}_{[0,s)}(x) + (x-1)\mathbf{1}_{[s,1)}(x)\big), \qquad 0 < x \le 1.$$

Obviously $[f_s]_{r,\varrho} = 0$. Since for any x we have that $f_s(x)$ assumes only two different values depending on whether $s \le x$ or $s > x$, the total number of evaluation points used by A_n for the class $S(M)$ is at most $2^n - 1$. Hence there is $[s_1, s_2] \subset (0,1)$ of length $2^{-(n+1)}$ that does not contain any of such points. This means that f_{s_1} and f_{s_2} share the same information and therefore for one of these two functions the error is at least $\|f_{s_1} - f_{s_2}\|_{L^p(0,1)}/2 = M\,2^{-(1+(n+1)/p)}$. Since M is arbitrarily large, the lemma follows. □

[1] $a_n \asymp b_n$ iff there are $0 < c_1, c_2 < \infty$ such that $c_1 a_n \le b_n \le c_2 a_n$ holds for sufficiently large n.

[2] Actually, we can say much more. Since our approximation problem is linear on $C^r_\varrho([0,1])$ and the class $\mathcal{F}_{r,\varrho}$ is convex and symmetric about zero, adaptive approximations can only be twice better than nonadaptive approximations.

The above negative worst case result was possible, because $\widehat{C}^r_\varrho([0,1])$ contains functions with arbitrarily large jumps $\Delta^{(0)}_f$. However, if we restrict the function class to

$$\begin{aligned}\widehat{\mathcal{F}}^{\rm disc}_{r,\varrho} &= \big\{\, f \in \widehat{C}^r_\varrho([0,1]) :\ [f]_{r,\varrho} \le 1,\ |\Delta^{(0)}_f| \le 1 \,\big\}, \quad \text{or}\\ \widehat{\mathcal{F}}^{\rm cont}_{r,\varrho} &= \big\{\, f \in \widehat{F}_{r,\varrho}([0,1]) :\ [f]_{r,\varrho}| \le 1,\ |\Delta^{(0)}_f| = 0 \,\big\},\end{aligned}$$

then positive results are possible.

The main difficulty in approximating piecewise smooth functions lies in locating the breakpoint s_f. This is where adaption comes in handy. The following adaptive procedure, which was first presented in [5], locates s_f sufficiently well. In the following, $m \ge 2r+1$ and $\omega > 0$, and $t_i = ih$, $h = 1/n$.

Detection procedure

Step 1. Evaluate the divided differences $d_i = f[t_i, \dots, t_{i+r+1}]$ for $1 \le i \le n$, and find

$$i^* = \arg\max_{1\le i\le m} |d_i|.$$

Return $u_1 := t_{i^*}$ and $v_1 := t_{i^*+r+1}$.

Step 2. Denote by p_- and p_+ the polynomials of degree r interpolating the data $(t_j, f(t_j))$ for $i^* - r \le j \le i^*$, and for $i^* + r + 1 \le j \le i^* + 2r + 1$, correspondingly.

$u := u_1$, $v := v_1$ $v - u > \omega$
$z_j := u + j(v-u)/(r+2), \quad j = 1, 2, \dots, r+1$
$j^* := \arg\max_{1\le j\le r+1} |p_+(z_j) - p_-(z_j)|$
if $|f(z_{j^*}) - p_-(z_{j^*})| \le |f(z_{j^*}) - p_+(z_{j^*})|$ **then** $u := z_{j^*}$ **else** $v := z_{j^*}$

Return $u_2 := u$ and $v_2 := v$.

Step 3. Do another iteration.

$u := u_2$, $v := v_2$ there is a local max of $|p_+ - p_-|$ in (u, v)
$z :=$ largest local max of $|p_+ - p_-|$ in (u, v)
if $|f(z) - p_-(z)| \le |f(z) - p_+(z)|$ **then** $u := z$ **else** $v := z$

Return $u_3 := u$ and $v_3 := v$.

Step 4. Set the final approximation of s_f to

$$\xi := \arg\min_{u_3\le x\le v_3} |p_+(x) - p_-(x)|.$$

Given an approximation ξ of s_f, the algorithm uses the polynomials p_- and p_+ to approximate f respectively in $[u_1, \xi)$ and $[\xi, v_1)$, and outside of $[u_1, v_1)$ it uses the piecewise Lagrange intepolation of degree r. We denote this algorithm by $A^*_{r,m}$.

It is worth noting that the singular point s_f is properly approximated, i.e., $s_f \in [u_3, v_3)$, if the discontinuity jumps $\Delta^{(j)}_f$ are large enough. On the other hand, the detection may fail. However, this only happens when the jumps are small relative to the resolution h, so the approximation error is not significantly affected.

The key property of $A^*_{r,m}$ is that for the pointwise error of any $f \in \widehat{C}^r_\varrho([0,1])$ we have

$$|f(x) - (A^*_{r,m}(f))(x)| \ \le\ D\,\big([f]_{r,\varrho} + 1\big) \cdot \begin{cases} h^{r+\varrho} + |\Delta^{(0)}_f|, & \text{for } x \in (u_2, v_2],\\ h^{r+\varrho}, & \text{otherwise,}\end{cases}$$

where D only depends on r and ϱ. Furthermore, the total number n of function evaluations is asymptotically equal to n. These observations lead to the following error estimations, for details see [5].

Theorem 2. *Let the parameter ω in the* Detection procedure *be chosen as:*

$$\omega = \begin{cases} h^{(r+\varrho)p+1}, & \textit{for } \widehat{\mathcal{F}}^{\rm disc}_{r,\varrho},\\ 1, & \textit{for } \widehat{\mathcal{F}}^{\rm cont}_{r,\varrho}.\end{cases}$$

Then

$$e^{\rm wor}\big(A^*_{r,m},\widehat{\mathcal{F}}^{\rm disc}_{r,\varrho}\big) \;:=\; \sup_{f\in\widehat{\mathcal{F}}^{\rm disc}_{r,\varrho}} \|f-A^*_m(f)\|_{L^p(0,1)} \;\asymp\; n^{-(r+\varrho)}, \qquad 1\le p<\infty,$$

$$e^{\rm wor}\big(A^*_{r,m},\widehat{\mathcal{F}}^{\rm cont}_{r,\varrho}\big) \;:=\; \sup_{f\in\widehat{\mathcal{F}}^{\rm cont}_{r,\varrho}} \|f-A^*_m(f)\|_{L^\infty(0,1)} \;\asymp\; n^{-(r+\varrho)}.$$

Theorem 2 does not cover the case $f\in\widehat{\mathcal{F}}^{\rm disc}_{r,\varrho}$ and $p=\infty$. This is no accident, because then the worst case error does not converge to zero, which follows from Lemma 2 of Section 3. In the other cases, the worst-case convergence $n^{-(r+\varrho)}$ from the case of smooth functions is recovered.

Observe that although the algorithm $A^*_{r,m}$ uses adaption to locate the breakpoint s_f, the number of adaptively chosen points is asymptotically negligible as $m\to\infty$. Indeed, for $\widehat{\mathcal{F}}^{\rm disc}_{r,\varrho}$ it is proportional to $\ln m$, and for $\widehat{\mathcal{F}}^{\rm cont}_{r,\varrho}$ it is a constant, because in this case Step 2 is skipped. If $r=1$ then the procedure is fully non-adaptive for continuous case. Thus, for piecewise smooth functions, adaption is used to a different extent, depending on the properties of the approximated functions. However, we want to stress that adaption is necessary to get the convergence $n^{-(r+\varrho)}$. This follows from Lemma 3 of Section 3, which says that using only nonadaptive approximations one can obtain the rate $n^{-1/p}$ at most.

We have considered only functions that have one nontrivial singular point. We conclude this section by noting that for functions with two or more singularities, the worst case results are negative. This is because the two singularities can be arbitrarily close to each other and thus cannot be efficiently detected with a fixed number n of function evaluations, see [8].

3. Asymptotic setting

In this section, we consider the asymptotic setting. That is, for a sequence of approximations $\{A_n\}_{n\ge1}$ we study how the error $\|f-A_n(f)\|_{L^p(0,1)}$ behaves as $n\to\infty$, for each problem instance $f\in C^r_\varrho([0,1])$.

3.1. Convergence rate

Proposition 1 establishes the upper bound $n^{-(r+\varrho)}$ on the convergence rate for each individual function from the smooth space $C^r_\varrho([0,1])$. Moreover, this rate can be obtained by the piecewise Lagrange interpolation $L_{r,m}$, which is nonadaptive. The following result, that follows from [18], says that the convergence $n^{-(r+\varrho)}$ is in a sense best we can get with the asymptotic approach. This means that adaption does not help in the asymptotic setting for smooth functions.

Theorem 3. *Let $\{\tau_n\}_{n\ge1}$ be an arbitrary positive sequence converging to zero. For any (adaptive or non-adaptive) approximations $\{A_n\}_{n\ge1}$ the set of all $f\in C^r_\varrho([0,1])$ such that*

$$\limsup_{n\to\infty}\frac{\|f-A_n(f)\|_{L^p}}{\tau_n\, n^{-(r+\varrho)}}<\infty$$

does not contain any non-trivial ball with respect to the norm $\|\cdot\|_{C([0,1])}+[\,\cdot\,]_{r,\varrho}$.

We now switch to piecewise smooth functions. First we show that, unlike in the smooth case, it is impossible to construct approximations that converge in $L^\infty(0,1)$ for all discontinuous functions. The following lemma (as well as Lemma 3 below) is a slight modification of the corresponding result from [15]. In what follows, for $s\in(0,1)$, we let

$$f_s(x)\;=\;M\big(x\mathbf{1}_{[0,s)}(x)+(x-1)\mathbf{1}_{[s,1)}(x)\big), \qquad 0<x\le1,$$

as in the proof of Lemma 1.

Lemma 2. *Let $r\ge1$. For any sequence $\{A_n\}_{n\ge1}$ of approximations the set*

$$\big\{\,s\in(0,1):\ \lim_{n\to\infty}\|f_s-A_n(f_s)\|_{L^\infty(0,1)}=0\big\}$$

is countable.

Proof. It is clear that $A_n(f_s)$ takes no more than 2^n different values. Furthermore, if $\|f_s - g\|_{L^\infty(0,1)} \le M/2$ then $\|f_t - g\|_{L^\infty(0,1)} > M/2$ for $t \ne s$. Hence $\|f_s - A_n(f_s)\|_{L^\infty(0,1)} < M/2$ for no more than 2^n values of t, which implies that the error converges to 0 for at most countably many s. □

In all other cases, we apply the algorithm $A^*_{r,m}$ from Section 2.2. However, to obtain the optimal rate $n^{-(r+\varrho)}$ for discontinuous functions, the use of adaption is again necessary.

Lemma 3. *Let $r \ge 1$. For any sequence $\{A_n\}_{n\ge1}$ of nonadaptive approximations the set*

$$B = \Big\{ s \in (0,1) : \lim_{n\to\infty} n^{1/p} \|f_s - A_n(f_s)\|_{L^p(0,1)} = 0 \Big\}$$

is of measure zero, i.e., $|B| = 0$.

Proof. For any function g the set of $s \in (0,1)$ such that $\|f_s - g\|_{L^p(0,1)} \le a$ is of measure at most $2(2a/M)^p$. Indeed, for any s, t we have $\|f_s - f_t\|_{L^p(0,1)} = M|s-t|^{1/p}$. Hence, if $\|f_t - g\|_{L^p(0,1)} \le a$ and $|s-t| > (2a/M)^p$ then

$$\|f_t - g\|_{L^p(0,1)} \ge \|f_s - f_t\|_{L^p(0,1)} - \|f_s - g\|_{L^p(0,1)} \ge M|s-t|^{1/p} - a > a.$$

Now, for $c > 0$, we define the sets

$$B^c_n = \big\{ s \in (0,1) : \|f_s - A_n(f_s)\|_{L^p(0,1)} \le c(n+1)^{-1/p} \big\}.$$

Since the information used by A_n takes at most $n+1$ different values for functions f_s, so does A_n. Hence

$$|B^c_n| \le 2(n+1)\left(\frac{2c}{M(n+1)^{1/p}}\right)^p = 2\left(\frac{2c}{M}\right)^p.$$

Letting

$$D^c_n = \big\{ s \in (0,1) : \forall \ell \ge n \ \|f_{g,s} - A_\ell(f_{g,s})\|_{L^p(0,1)} \le c(\ell+1)^{-1/p} \big\}$$

we have $D^c_n \subset B^c_n$, $D^c_n \subset D^c_{n+1}$ for all n, and $B \subset \cup_{n=1}^\infty D^c_n$. Hence

$$|B| \le \lim_{n\to\infty} |D^c_n| \le 2(2c/M)^p.$$

Since c can be arbitrarily small, $|B| = 0$. □

Furthermore, unlike in the wost case setting, for continuous functions the convergence rate $n^{-(r+\varrho)}$ can be obtained using fully nonadaptive approximation. It is enough to use $A^*_{r,m}$ skipping Steps 2 and 3 in the DETECTION PROCEDURE, as shown in Proposition 5 of [5] (see also [1]).

Another advantage of using the asymptotic approach is that the error rate $n^{-(r+\varrho)}$ can also be obtained for functions f that possess more than just one singular point. Then we can apply the DETECTION PROCEDURE appropriately many times, and for sufficiently large number n of function evaluations (where n depends on f) all the essential singular points of f can be located. (This would not be possible with the worst case approach, where n is fixed). For details, see [8].

3.2. Asymptotic constants

So far we have compared different approximations in terms to the convergence rate of the worst case error in a given function class, or in terms of the convergence for each individual f in the class, as $n \to \infty$. Using this criterion, we argued that adaption does not help in approximating smooth functions, both in the worst case and in the asymptotic settings. So why are adaptive methods used for smooth functions? The point is that even if two approximations have the same convergence rate, their asymptotic constants can differ significantly, making one of them much more effective than the other. While there are no satisfactory general results regarding the asymptotic constants for the worst case setting, much more can be said about them in the asymptotic setting. This section in entirely based on [13].

For an integer $r \ge 1$, we consider the space $C^r([0,1])$ of functions $f : [0,1] \to \mathbb{R}$ that are r-times continuously differentiable. Such functions can be approximated using piecewise Lagrange

interpolation of degree $r-1$ with possibly non-uniform partition of $[0,1]$. Specifically, we first fix points

$$0 \le t_1 < t_2 < \cdots < t_r \le 1.$$

For a given $f \in C^r([0,1])$, the interval $[0,1]$ is subdivided into m subintervals that are determined by a choice of points

$$0 = x_0 < x_1 < \cdots < x_m = 1.$$

In each $[x_{j-1}, x_j]$, the function is approximated by its Lagrange polynomial of degree $r-1$ interpolating f at

$$x_{j,i} = x_{j-1} + h_j t_i, \qquad 1 \le i \le r,$$

where $h_j = x_j - x_{j-1}$. If the partition is equispaced then this results in the nonadaptive approximation $L_{r-1,m}(f)$ considered earlier. We established the optimality of this approximation with respect to the convergence rate. Now we are interested in the asymptotic constants.

Proposition 2. *Let $1 \le p \le \infty$. For any $f \in C^r([0,1])$ we have*

$$\|f - L_{r-1,m}(f)\|_{L^p(0,1)} \approx \frac{\alpha_{r,p}}{r!} \left\|f^{(r)}\right\|_{L^p(0,1)} m^{-r},$$

where $a_{r,p} = \|P_r\|_{L^p(0,1)}$ with $P_r(t) = (t-t_1)(t-t_2)\cdots(t-t_r)$. [3]

It turns out that an adaptive choice of the partition may lead to much better asymptotic constants. Indeed, for a given f, consider the partition $0 = x_0^* < x_1^* < \cdots < x_m^* = 1$ that makes all local errors equal, i.e., for the corresponding approximation $L^*_{r-1,m}(f)$

$$\|f - L^*_{r-1,m}(f)\|_{L^p(x^*_{j-1},x^*_j)} = \frac{1}{m} \|f - L^*_{r-1,m}(f)\|_{L^p(0,1)}, \quad 1 \le j \le m.$$

Such a partition exists since the local errors continuously depend on the points x_i.

Proposition 3. *The equal-local error piecewise Lagrange interpolations $L^*_{r-1,m}(f)$ enjoy the best asymptotic constants for all $f \in C^r([0,1])$. Furthermore,*

$$\|f - L^*_{r-1,m} f\|_{L^p(0,1)} \approx \frac{\alpha_{r,p}}{r!} \left\|f^{(r)}\right\|_{L^{1/(r+1/p)}(0,1)} m^{-r}.$$

(Note that $\|\cdot\|_{L^{1/(r+1/p)}(0,1)}$ is not a norm if $r + 1/p > 1$, where $1/\infty = 0$ by convention.)

Thus the optimal partition is asymptotically better than the uniform partition by the factor of

$$R_{r,p}(f) = \frac{\|f^{(r)}\|_{L^p(0,1)}}{\|f^{(r)}\|_{L^{1/(r+1/p)}(0,1)}}.$$

We have $1 \le R_{r,p}(f) < \infty$. It equals 1 for polynomials of degree r, and the more $f^{(r)}$ varies the bigger $R_{r,p}(f)$. The gain from using adaption can be significant. For instance, let $f(x) = 1/(x + 10^{-d})$. If $d = 2$ then $R_{4,\infty} \cong 10^6$, and for $d = 8$ we have $R_{r,\infty} \cong 10^{29}$.

Now the question is whether the optimal partition can be realized algorithmically using only function evaluations. The following procedure almost does the job, at least for functions f for which the r th derivate does not change its sign.

We fix $t_0 \in [0,1]$ that is different from all t_i's. For an interval $I = [c,d] \subset [0,1]$ of length $h = d - c$, define the functional $\mathcal{L}_I : C^r([0,1]) \to \mathbb{R}$ as

$$\mathcal{L}_I(f) = f(u_0) - \sum_{i=1}^{r} w_i f(u_i), \quad \text{where} \quad w_i = \prod_{i \ne k = 1}^{r} \frac{t_0 - t_k}{t_i - t_k}$$

and $u_i = c + h t_i$, $0 \le i \le r$.

The procedure uses a priority queue S whose elements are subintervals. For each $I \in S$ of length h its priority is given by $p_f(I) = h^{1/p} |\mathcal{L}_I(f)|$. In the following pseudocode, insert(S, I)

[3] $a_m \approx b_m$ iff $\lim_{m\to\infty} a_m/b_m = 1$.

and $I := \text{extract_max}(S)$ denote correspondingly inserting an interval to S, and extracting from S an interval with the highest priority.

Optimal Partition

$S := \emptyset$; insert$(S, [0,1])$;
for $k = 2, 3, \ldots, m$
  $[l, r] := \text{extract_max}(S)$;
  $c := (l + r)/2$;
  insert$(S, [l, c])$; insert$(S, [c, r])$
end

Denote by $L^{**}_{m,r}(f)$ the approximation corresponding to the partition above.

Theorem 4. *If the derivative $f^{(r)}$ does not change its sign in $[0,1]$, then*

$$\|f - L^{**}_{r-1,m}(f)\|_{L^p(0,1)} \lessapprox \kappa_{r,p} \frac{\alpha_{r,p}}{r!} \|f^{(r)}\|_{L^{1/(r+1/p)}(0,1)} m^{-r},$$

where $\kappa_{r,p}$ is a 'small' factor, e.g., $\kappa_{4,1} \cong 2.9$, $\kappa_{4,2} \cong 4.7$, $\kappa_{4,\infty} = 16$. [4]

Comparing Theorem 4 with Proposition 2 we coclude that adaption can significantly help in optimazing asymptotic constants in approximating smooth functions.

References

1 F.Arandiga, A. Cohen, R. Donat, N.Dyn, Interpolation and approximation of piecewise smooth functions, SIAM Journal of Numerical Analysis, 43 (2006), No. 1, 41–57.

2 N. S.Bakhvalov, On the optimality of linear methods for operator approximation in convex classes of functions, Zh. Vychisl. Mat. Mat. Fiz., 11 (1971), No. 4, 1014–1018; U.S.S.R. Comput. Math. Math. Phys., 11 (1971), No. 4, 244–249.

3 S.Gal, C.A. Micchelli, , Optimal sequential and non-sequential procedures for evaluating a functional, Applicable Analysis, 10 (1980), 105–120.

4 A.Kałuża, L. Plaskota, Asymptotic analysis of adaptive Simpson quadratures for piecewise smooth functions, Monte Carlo and Quasi-Monte Carlo, 2024, Ch. Lemieux and B. Feng eds., Springer Proceedings in Mathematics and Statistics, 52, Springer, 2026, 297–315.

5 P. M.Morkisz, L.Plaskota, Approximation of piecewise Hölder functions from inexact information, Journal of Complexity, 32(2016), No. 1, 122–136.

6 P.M. Morkisz, L.Plaskota, Complexity of approximating Hölder classes from information with varying Gaussian noise, Journal of Complexity, 60 (2020), 101497.

7 E.Novak, On the power of adaption, Journal of Complexity, 12(1996), 199-237.

8 M.Perlik, L.Plaskota, Approximation of piecewise smooth functions from information contaminated with random noise, Submitted, 2026.

9 R.Piessens, E.de Doncker-Kapenga, C. W. Ueberhuber, D. K.Kahaner, QUADPACK: A subroutine package for automatic integration, Springer Verlag, 1983.

10 L.Plaskota, Automatic integration using asymptotically optimal adaptive Simpson quadrature, Numerische Mathematik, 131(2015), 173–198.

11 L.Plaskota, Noisy Information and Computational Complexity, Cambridge Univ. Press, Cambridge, 1996.

12 L.Plaskota, P.Przybyłowicz, Stępień, Monte Carlo integration of C^r functions with adaptive variance reduction: an asymptotic analysis, BIT Numerical Analysis, 63(2023), 1–24.

13 L.Plaskota, P.Samoraj, Automatic approximation using asyptotically optimal adaptive interpolation, Numerical Algorithms, 89(2021), 277–302.

14 L.Plaskota, G.W. Wasilkowski, Adaption allows efficient integration of functions with singularities, Numerische Mathematik, 10 (2005), 123–144.

15 L.Plaskota, G. W. Wasilkowski, Y.Zhao, The power of adaption for approximating functions with singularities, Mathematics of Computation, 77(2008), 2309-2338.

16 J. F.Traub, H.Woźniakowski, A General Theory of Optimal Algorithms, Academic Press, New York, 1980.

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

18 G. M.Trojan,Asymptotic setting for linear problems, manuscript (1983) (See also [17].)

[4] $a_m \lessapprox b_m$ iff $\limsup_{m\to\infty} a_m/b_m \le 1$.

Функцияларды жуықтаудағы адаптивті алгоритмдер қаншалықты пайдалы?

Л. Пласкота
Варшава университеті, С. Банах көш., 2, 02-097, Варшава, Польша

Аннотация. Адаптивті алгоритмдер сандық есептерді шешуде кеңінен қолданылады, функциялардың нөлдерін табуға арналған Ньютон әдісі мен сандық интегралдауға арналған адаптивті квадратуралық формулалар — олардың белгілі мысалдары. Бұл мақалада скаляр функцияларды жуықтау есебіне адаптивті алгоритмдерді қолдануға қатысты жекелеген нәтижелер қарастырылады. Бұнда рұқсат етілетін алгоритмдер функцияның адаптивті немесе адаптивті емес тәсілмен таңдалатын мәндерін пайдаланады. Ақпараттық күрделілік теориясы аясында ең нашар және асимптотикалық жағдайдағы есеп қойылымдары қарастырылады. Екі қойылымда да тегіс функциялар үшін адаптивті алгоритмдер адаптивті емес алгоритмдермен салыстырғанда жинақталудың жоғары жылдамдығын қамтамасыз ете алмады. Адаптацияның артықшылықтары бөлік-бөлік тегіс функцияларды жуықтауда және асимптотикалық тұрақтыларды оңтайландыруда көрінеді. Атап айтқанда, бөліктеп тегіс функциялар үшін тегіс функциялар жағдайында қол жеткізілетін оңтайлы жинақталу жылдамдығын жуықтау үшін адаптация қажет. Сонымен қатар, адаптивті алгоритмдер адаптивті емес алгоритмдермен салыстырғанда асимптотикалық тұрақтыларды, әсіресе тиісті туындылары күрт өзгеретін функциялар үшін, едәуір жақсартуға мүмкіндік береді.

Түйін сөздер: функцияларды жуықтау, адаптивті алгоритмдер, Лагранж интерполяциясы, асимптотикалық тұрақтылар, бөлік-бөлік тегістік.

Насколько полезны адаптивные алгоритмы для аппроксимации функций?

Л. Пласкота
Варшавский университет, ул. С. Банаха, 2, 02-097, Варшава, Польша

Аннотация. Адаптивные алгоритмы широко используются для решения численных задач. Классическими примерами являются метод Ньютона для нахождения нулей функций и адаптивные квадратурные формулы для численного интегрирования. В настоящей статье рассматриваются отдельные результаты, касающиеся применения адаптивных алгоритмов к задаче приближения скалярных функций. Допустимые алгоритмы используют значения функции, которые могут выбираться адаптивно или неадаптивно. Рассматриваются постановки в наихудшем и в асимптотическом случаях в рамках теории информационной сложности. В обеих постановках адаптивные алгоритмы не обеспечивают более высоких скоростей сходимости по сравнению с неадаптивными алгоритмами для гладких функций. Преимущества адаптации проявляются при приближении кусочно-гладких функций и оптимизации асимптотических констант. В частности, для кусочно-гладких функций адаптация необходима для восстановления оптимальной скорости сходимости, достигаемой в случае гладких функций. Адаптивные алгоритмы также позволяют существенно улучшить асимптотические константы по сравнению с неадаптивными алгоритмами, особенно для функций, соответствующие производные которых резко изменяются.

Ключевые слова: Приближение функций, адаптивные алгоритмы, интерполяция Лагранжа, асимптотические константы, кусочная гладкость.

References

1 F.Arandiga, A. Cohen, R. Donat, N.Dyn, Interpolation and approximation of piecewise smooth functions, SIAM Journal of Numerical Analysis, 43 (2006), No. 1, 41–57.

2 N. S.Bakhvalov, On the optimality of linear methods for operator approximation in convex classes of functions, Zh. Vychisl. Mat. Mat. Fiz., 11 (1971), No. 4, 1014–1018; U.S.S.R. Comput. Math. Math. Phys., 11 (1971), No. 4, 244–249.

3 S.Gal, C.A. Micchelli, , Optimal sequential and non-sequential procedures for evaluating a functional, Applicable Analysis, 10 (1980), 105–120.
4 A.Kałuża, L. Plaskota, Asymptotic analysis of adaptive Simpson quadratures for piecewise smooth functions, Monte Carlo and Quasi-Monte Carlo, 2024, Ch. Lemieux and B. Feng eds., Springer Proceedings in Mathematics and Statistics, 52, Springer, 2026, 297–315.
5 P. M.Morkisz, L.Plaskota, Approximation of piecewise Hölder functions from inexact information, Journal of Complexity, 32(2016), No. 1, 122–136.
6 P.M. Morkisz, L.Plaskota, Complexity of approximating Hölder classes from information with varying Gaussian noise, Journal of Complexity, 60 (2020), 101497.
7 E.Novak, On the power of adaption, Journal of Complexity, 12(1996), 199-237.
8 M.Perlik, L.Plaskota, Approximation of piecewise smooth functions from information contaminated with random noise, Submitted, 2026.
9 R.Piessens, E.de Doncker-Kapenga, C. W. Ueberhuber, D. K.Kahaner, QUADPACK: A subroutine package for automatic integration, Springer Verlag, 1983.
10 L.Plaskota, Automatic integration using asymptotically optimal adaptive Simpson quadrature, Numerische Mathematik, 131(2015), 173–198.
11 L.Plaskota, Noisy Information and Computational Complexity, Cambridge Univ. Press, Cambridge, 1996.
12 L.Plaskota, P.Przybyłowicz, Stępień, Monte Carlo integration of C^r functions with adaptive variance reduction: an asymptotic analysis, BIT Numerical Analysis, 63(2023), 1–24.
13 L.Plaskota, P.Samoraj, Automatic approximation using asyptotically optimal adaptive interpolation, Numerical Algorithms, 89(2021), 277–302.
14 L.Plaskota, G.W. Wasilkowski, Adaption allows efficient integration of functions with singularities, Numerische Mathematik, 10 (2005), 123–144.
15 L.Plaskota, G. W. Wasilkowski, Y.Zhao, The power of adaption for approximating functions with singularities, Mathematics of Computation, 77(2008), 2309-2338.
16 J. F.Traub, H.Woźniakowski, A General Theory of Optimal Algorithms, Academic Press, New York, 1980.
17 J. F.Traub, G.W. Wasilkowski, H.Woźniakowski , Information-Based Complexity, Academic Press, New York, 1988.
18 G. M.Trojan,Asymptotic setting for linear problems, manuscript (1983) (See also [17].)

Information about the author:

Leszek Plaskota– Professor, Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, S. Banach Str., 2, 02-097, Warsaw, Poland.

Автор туралы мәлімет:

Лешек Пласкота – профессор, математика, информатика және механика факультеті, Варшава университеті, С. Банах көшесі, 2, 02-097, Варшава, Польша.

Received: 21.09.2026.
Approved: 17.10.2026. Available online: 01.11.2026.

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY NC) license (https://creativecommons.org/licenses/by-nc/4.0/).