MATHEMATICS OF COMPUTATION
Volume 77, Number 264, October 2008, Pages 2309–2338
S 0025-5718(08)02103-0
Article electronically published on April 24, 2008

THE POWER OF ADAPTION FOR APPROXIMATING FUNCTIONS WITH SINGULARITIES

LESZEK PLASKOTA, GRZEGORZ W. WASILKOWSKI, AND YAXI ZHAO

Abstract. Consider approximating functions based on a finite number of their samples. We show that adaptive algorithms are much more powerful than nonadaptive ones when dealing with piecewise smooth functions. More specifically, let F_r^1 be the class of scalar functions $f : [0, T] \to \mathbb{R}$ whose derivatives of order up to r are continuous at any point except for one unknown *singular* point. We provide an adaptive algorithm $\mathcal{A}_n^{\mathrm{ad}}$ that uses at most n samples of f and whose *worst case* L^p error ($1 \le p < \infty$) with respect to 'reasonable' function classes $\mathcal{F}_r^1 \subset F_r^1$ is proportional to n^{-r}. On the other hand, the worst case error of any nonadaptive algorithm that uses n samples is at best proportional to $n^{-1/p}$.

The restriction to only one singularity is necessary for superiority of adaption in the worst case setting. Fortunately, adaption regains its power in the *asymptotic setting* even for a very general class F_r^∞ consisting of piecewise C^r-smooth functions, each having a finite number of singular points. For any $f \in F_r^\infty$ our adaptive algorithm approximates f with error converging to zero at least as fast as n^{-r}. We also prove that the rate of convergence for nonadaptive methods cannot be better than $n^{-1/p}$, i.e., is much slower.

The results mentioned above do not hold if the errors are measured in the L^∞ norm, since no algorithm produces small L^∞ errors for functions with unknown discontinuities. However, we strongly believe that the L^∞ norm is inappropriate when dealing with singular functions and that the *Skorohod metric* should be used instead. We show that our adaptive algorithm retains its positive properties when the approximation error is measured in the Skorohod metric. That is, the worst case error with respect to $\mathcal{F}_r^1$ equals $\Theta(n^{-r})$, and the convergence in the asymptotic setting for F_r^∞ is n^{-r}.

Numerical results confirm the theoretical properties of our algorithms.

1. Introduction

In traditional studies of function approximation one usually assumes that the functions under consideration are smooth in some sense, e.g., they are r-times continuously differentiable, and that they form a separable linear space. On the other hand, functions appearing in practice are very often only piecewise smooth, and their collection is more complicated than a linear space. An important example is provided by image representation. Here edges correspond to discontinuities of the underlying function, and the linear combination of two images need not be any image of practical interest.

Received by the editor July 24, 2007.
2000 *Mathematics Subject Classification.* Primary 65Y20, 65D05, 41A10, 41A25.
Key words and phrases. Function approximation, sampling, singularities, adaptive algorithms.

©2008 American Mathematical Society
Reverts to public domain 28 years from publication

Much attention has been recently devoted to wavelets that are now considered a universal tool to deal with images and signals. A given function f (image, signal) is represented by a few of the largest coefficients of its expansion in a suitable wavelet basis (or curvelet basis in 2-D); see, e.g., [5, 8, 9, 15]. It is also believed that adaptive methods, that rely on adjusting the choice of the basis to the underlying function and its discontinuities (edges), are much more efficient than nonadaptive methods where the basis is fixed for all f's. Note that the wavelet approach is based on the paradigm that one has at his disposal *all* wavelet coefficients of f corresponding to a given basis. This actually means that *complete* information about f is available.

In the present paper, we take an *information-based* approach. It follows the paradigm that only *partial* information about f is available, and that the only source of information is an *oracle*; see, e.g., [16, 20, 24, 25, 28, 29, 30, 32, 33]. This seems to be more suitable for numerical computations where the oracle calls usually correspond to function evaluations. In this approach the term *adaption* means adjusting the selection of successive oracle calls to the results of previous calls. Note that superiority of adaptive methods is here far less obvious since so far adaptive methods have been mainly developed for integration (and not approximation) problems; see, e.g., [2, 6, 7, 12, 13, 14, 19, 22, 23]. Moreover, there are theoretical results showing that adaption does not help for linear problems such as function approximation. This holds in different settings, including the worst case setting with convex classes, and the asymptotic setting with Banach function spaces; see, e.g., [3, 17, 10, 30, 31]. Observe that singular functions of practical interest form neither convex classes nor Banach spaces. Therefore the results mentioned above do not contradict the common belief that adaptive methods should be used when dealing with singularities.

The information-based approach to the approximation of singular functions was presented in a recent paper [1]. The scalar function f to be approximated is assumed to be globally continuous, and C^r-smooth except for exactly one (unknown) point at which f' has nontrivial discontinuity. The authors use the uniform mesh of size h to construct a nonadaptive approximation which converges to f with error proportional to h^r in the uniform norm. This result means that for singular but globally continuous functions, adaption does not help in the asymptotic setting, i.e., nonadaptive methods are optimal.

We consider a class F_r^∞ of functions with more general singularities. We assume that $f : [0,T] \to \mathbb{R}$ and allow discontinuity of the function itself and/or of its derivatives starting at order perhaps higher than one, as assumed in [1]. That is, f may be, say, twice differentiable on $[0,T]$, and $f^{(3)}(s)$ may not exist at some point s. Moreover, f may have more than just one (but finitely many) singular points; their location and number are unknown. We want to approximate functions $f \in F_r^\infty$ by algorithms $\mathcal{A}_n f$ that use function values at some n points $x_1, \ldots, x_n$ as the only available information. We allow *adaptive* algorithms, in which case the choice of x_j depends on previously obtained values of $f(x_1), \ldots, f(x_{j-1})$.

An interesting feature of discontinuous functions is that it is impossible to approximate them with small errors measured in the L^∞ norm. This should not be surprising after realizing how inappropriate the L^∞ norm is for such functions. To see this, let us consider perhaps the simplest discontinuous functions from our class

F_r^∞, indicator functions of subintervals of $[0,T]$ given by

$$f_a(x) \;=\; \mathbb{1}_{[0,a]}(x) \;:=\; \begin{cases} 1 & \text{if } x \in [0,a], \\ 0 & \text{if } x \in (a,T]. \end{cases}$$

Since $\|f_a - f_b\|_{L^\infty} = 1$ whenever $a \neq b$, the sequence $\{f_{a+1/n}\}$ does not converge to f_a in the L^∞ norm even though both functions are *almost identical*. In particular, the function f_a cannot be represented with error smaller than 1 (in the L^∞ norm) in a *floating point arithmetic of finite precision*, unless a is a machine number. Indeed, if $a \neq \mathrm{fl}(a)$, then again $\|f_a - f_{\mathrm{fl}(a)}\|_{L^\infty} = 1$, yet $f_{\mathrm{fl}(a)}$ clearly is the best possible approximation/representation of f_a one could think of. Even worse, the constant function $f \equiv 1/2$ is a better (in the L^∞ norm) approximation to f_a than $f_{\mathrm{fl}(a)}$! (See Section 3 for a more detailed discussion.) Of course, one could use L^p norms with finite p, and we indeed do analyze the algorithms using L^p norms for all $p \in [1,\infty)$. However, L^p norms may not be adequate for applications where introduction of false singularities is undesirable.

In other words, one would like to use a distance (metric) that enjoys all the good properties of the uniform norm, yet is also appropriate for discontinuous functions. Fortunately, the *Skorohod metric* $\mathrm{dist}_{\mathrm{S}}(\cdot,\cdot)$, see (3), enjoys the desiderata above. Roughly speaking, the distance between two functions f and g is small iff the graph of one of them, say f, can be viewed as a slightly perturbed graph of g, i.e., f is close to $g \circ \lambda$ for some homeomorphism λ that is close to the identity function $\mathrm{id}(x) = x$, both in the L^∞ norm. Following our example, note that $\mathrm{dist}_{\mathrm{S}}(f_a, f_{\mathrm{fl}(a)}) = |a - \mathrm{fl}(a)|$ is small, as it should be. Moreover, for *continuous* functions from our class F_r^∞ (e.g., with discontinuity of derivatives but not of the function itself), the Skorohod distance is equivalent to the uniform norm. Hence, one could say that it has the good and practical properties of the uniform norm without the drawbacks that the norm induces for discontinuous functions. We refer again to Section 3 for some relations between the Skorohod metric and L^p norms.

In our paper, we seek optimal nonadaptive and adaptive algorithms in the worst case and asymptotic settings, assuming that the error of approximation is measured in the L^p norm ($1 \le p < \infty$) or in the Skorohod metric.

We are ready to discuss the main results of the paper. Since they strongly depend on the number of singular points, we begin with the class $F_r^1 \subset F_r^\infty$ that consists of functions with at most one singular point. In Section 4 we prove that the best nonadaptive algorithms using n function values have the rather poor convergence rate $n^{-1/p}$ when the L^p norms are used, and n^{-1} when the Skorohod metric is used; see Theorem 2 and Proposition 3. These negative results should be contrasted with the case of globally C^r-smooth functions (having no singularities), where nonadaptive piecewise polynomial approximation yields the error bounded by $CT^r \, \|f^{(r)}\|_{L^\infty} \, n^{-r}$ in the uniform norm, for any f and $n \ge r$.

Can we do better by using adaptive algorithms? Definitely YES. In Section 5 we construct an adaptive algorithm $\mathcal{A}_n^{\mathrm{ad}}$ that achieves the optimal convergence rate of n^{-r} for both the L^p norms and the Skorohod metric. Moreover, the error asymptotically behaves as though there were no singularities, i.e., for any $f \in F_r^1$ we have

$$\lim_{n\to\infty} \|f - \mathcal{A}_n^{\mathrm{ad}} f\|_{L^p} \cdot n^r = T^r \cdot \gamma \cdot \|f^{(r)}\|_{L^p} \tag{1}$$

and

$$\lim_{n\to\infty} \mathrm{dist}_{\mathrm{S}}(f, \mathcal{A}_n^{\mathrm{ad}} f) \cdot n^r = T^r \cdot \gamma \cdot \|f^{(r)}\|_{L^\infty}, \tag{2}$$

where γ depends only on r and the metric used; see Theorems 3 and 4.

We stress that (1) and (2) are only asymptotic results. If we switch to the worst case setting, then the situation is more complicated. For the worst case error to be of order n^{-r} it is *not* enough to assume that the rth derivative is uniformly bounded, which is the case for functions with no singularities. Actually, we prove in Section 5.4 that for any (possibly adaptive) algorithm $\mathcal{A}_n$ its worst case error with respect to the class

$$\{\, f \in F_r^1 \mid \|f^{(r)}\|_{L^\infty} \le L_r \}$$

is infinite, independent of whether the L^p norm or the Skorohod metric is considered. Hence, for positive worst case results we have to put more restrictions on f. Indeed, we show that for our adaptive algorithm $\mathcal{A}_n^{\mathrm{ad}}$ we have

$$\sup\{\, \mathrm{dist}_{\mathrm{S}}(f, \mathcal{A}_n^{\mathrm{ad}} f) \mid f \in F_r^1,\ \|f^{(r)}\|_{L^\infty} \le L_r,\ \|f'\|_{L^\infty} \le L_1 \,\} = \Theta(n^{-r})$$

and

$$\sup\{\, \|f - \mathcal{A}_n^{\mathrm{ad}} f\|_{L^p} | f \in F_r^1, \|f^{(r)}\|_{L^\infty} \le L_r, \|f'\|_{L^\infty} \le L_1, |\Delta_f^{(0)}| \le D_0 \,\} = \Theta(n^{-r}),$$

where L_r, L_1, and D_0 are finite, and $\Delta_f^{(0)}$ is the discontinuity jump of f. For the details, see again Theorems 3 and 4, as well as Section 5.4.

Our algorithm $\mathcal{A}_n^{\mathrm{ad}}$ consists of two steps. First, an adaptive (bisection-like) mechanism is used to detect and localize singularities; next, a piecewise polynomial interpolation of degree $r-1$ is applied. The detection mechanism is based on ideas already developed in [21], where numerical integration of piecewise smooth functions is considered, and relies on getting some quantitative information about the singularity by computing rth order divided differences. Roughly speaking, singularities with *big* discontinuity jumps are easily detected and localized with small errors, and those not detected have *small* enough jumps not to influence the approximation error.

Next, in Section 6, we consider the general case of functions with multiple singularities. We show in Section 6.1 that, unfortunately, the worst case errors of any algorithm with respect to any reasonable function class from F_r^∞ are bounded away from zero, independent of the number n of function evaluations used; see Proposition 4. In other words, the worst case errors of even optimal algorithms do not converge to zero. This is because two or more singularity points can be arbitrarily close to one another, which makes it impossible to distinguish them using a fixed number n of function evaluations.

Fortunately, the lack of convergence in the worst case setting can be reversed in the asymptotic setting, as considered in Section 6.2. Indeed, a modification of $\mathcal{A}_n^{\mathrm{ad}}$ leads to an algorithm $\overline{\mathcal{A}}_n^{\mathrm{ad}}$ that replicates the asymptotic properties (1) and (2); see Theorem 5.

Finally, in Section 7, we report on numerical results that confirm the theoretical properties of $\mathcal{A}_n^{\mathrm{ad}}$ and $\overline{\mathcal{A}}_n^{\mathrm{ad}}$.

2. Basic definitions

We say that a function $g : [a,b] \to \mathbb{R}$ is *r-smooth* iff $g \in C^{r-1}([a,b])$, with $g^{(r)} \in L^\infty(a,b)$ continuous almost everywhere. The set of all r-smooth functions defined on $[a,b]$ will be denoted by $W_r(a,b)$.

For given $T > 0$ and $r \geq 1$ we consider the space $F_r^\infty = F_r^\infty(0,T)$ of functions

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

that are piecewise r-smooth. Specifically, $f \in F_r^\infty$ iff there exists an integer $k = k_f \geq 0$, points $0 = s_0 < s_1 < \cdots < s_k < s_{k+1} = T$, and functions $g_i \in W_r(s_i, s_{i+1})$ such that for all $0 \leq i \leq k$ and $x \in (s_i, s_{i+1})$, we have

$$f(x) = g_i(x).$$

Moreover, $f(0) = g_0(0)$, $f(T) = g_k(T)$, and either $f(s_i) = g_{i-1}(s_i)$ or $f(s_i) = g_i(s_i)$, $1 \leq i \leq k$, i.e., f is left- or right-continuous at each s_i.

The points s_i, $1 \leq i \leq k$, are *singularities* of f. The corresponding discontinuity jumps are denoted by

$$\Delta_i^{(j)} := g_i^{(j)}(s_i) - g_{i-1}^{(j)}(s_i), \qquad 0 \leq j \leq r-1.$$

Observe that any $f \in F_r^\infty$ admits the useful representation

$$f(x) \;=\; g(x) \;+\; \sum_{i=1}^{k} \mathbb{1}_{I_i}(x) \sum_{j=0}^{r-1} \Delta_i^{(j)} \, \frac{(x-s_i)^j}{j!},$$

where $g \in W_r(0,T)$ and

$$I_i \;=\; \begin{cases} [s_i, T] & \text{if } f \text{ is right-continuous at } s_i, \\ (s_i, T] & \text{if } f \text{ is left-continuous at } s_i. \end{cases}$$

Here and elsewhere, $\mathbb{1}_A$ denotes the indicator function of the set A.

We distinguish in F_r^∞ the classes F_r^ℓ of functions with no more than ℓ singular points,

$$F_r^\ell := \{\, f \in F_r^\infty \quad | \quad k_f \leq \ell \,\}.$$

In particular, $F_r^0 = W_r(0,T)$. Obviously,

$$F_r^0 \subset F_r^1 \subset \cdots \subset F_r^\ell \subset \cdots \quad \text{and} \quad F_r^\infty \;=\; \bigcup_{\ell=0}^{\infty} F_r^\ell.$$

For fixed ℓ (finite or infinite) our aim is to construct efficient algorithms for approximating functions $f \in F_r^\ell$ in a given *metric (distance)*, choices of which will be discussed later. We assume that the only *a priori* information about f is that $f \in F_r^\ell$. In particular, the locations of singularities are unknown; when $\ell = \infty$, we also assume that we do not have an upper bound on the number of singularities. However, together with this *a priori* information, the algorithms also use some *a posteriori* information that consists of finitely many evaluations of f at certain points x_i. For instance, the algorithm may be given as

$$\mathcal{A}_n f \;=\; \sum_{i=1}^{n} f(x_i) \cdot f_i$$

for some $x_i \in [0,T]$ and functions f_i. These include (piecewise) polynomial or trigonometric interpolation, splines, etc. More generally, by an algorithm we mean a mapping of the form

$$\mathcal{A}_n f \;=\; \varphi(f(x_1),\ldots,f(x_n)).$$

If the points x_i are chosen independently of f, then the algorithm is *nonadaptive*. An important particular case is provided by the equispaced sampling

$$x_i \;=\; \frac{i-1}{n-1}\,T, \qquad 1 \le i \le n.$$

We also consider *adaptive* algorithms in which the choice of the successive x_i's and their number $n = n(f)$ depend on previously obtained information about f. That is, x_1 is fixed and

$$x_i \;=\; x_i\big(f(x_1),\ldots,f(x_{i-1})\big) \quad \text{for} \quad 2 \le i \le n(f).$$

The index n of $\mathcal{A}_n$ means that for any $f \in F_r^\ell$ the algorithm uses no more than n function evaluations, i.e., $n(f) \le n$ for all f.

Note that our problem is determined by the parameters T, r, ℓ, and the metric with respect to which the error of approximation is measured. Hence the results will also depend on them.

Throughout this paper we denote by $\|\cdot\|_{L^p(a,b)}$ the L^p norm on (a,b), for $1 \le p \le \infty$. To simplify notation, we write L^p instead of $L^p(0,T)$ if $(a,b) = (0,T)$. The same applies to the space C of continuous functions and the uniform (Chebyshev) norm.

We will also say that an x is in $[a,b]$ with respect to f (or w.r.t. f in the sequel), and write

$$x \in_f [a,b],$$

iff $x \in (a,b)$, or $x = a$ and f is left-continuous at x, or $x = b$ and f is right-continuous at x. Obviously, if f is continuous at x, then $x \in_f [a,b]$ if and only if $x \in [a,b]$.

3. L^p norms versus the Skorohod metric

It would be desirable to have algorithms that converge fast in the most conservative uniform L^∞ norm. Unfortunately, L^∞ approximation is an ill-posed problem when dealing with discontinuous functions. Indeed, we have the following negative result.

For a given function $f : [0,T] \to \mathbb{R}$, $\varepsilon \in \mathbb{R}$, and $u \in (0,T)$, let

$$f_{\varepsilon,u} \;:=\; f + \varepsilon\, \mathbb{1}_{[0,u)}.$$

Clearly, if $f \in F_r^\ell$, then $f_{\varepsilon,u} \in F_r^{\ell+1}$.

Lemma 1. *Let $\{\mathcal{A}_n\}_{n\ge1}$ be a sequence of (adaptive) algorithms. Then for any function f and $\varepsilon \ne 0$, the set of $u \in [0,T)$ for which*

$$\lim_{n\to\infty} \|f_{\varepsilon,u} - \mathcal{A}_n f_{\varepsilon,u}\|_{L^\infty} \;=\; 0$$

is countable.

Proof. It is clear that $\mathcal{A}_n f_{\varepsilon,u}$ (with $0 \le u < T$) takes at most 2^n different values. Moreover, if $\|f_{\varepsilon,u_0} - g\|_{L^\infty} < \varepsilon/2$ for some u_0 and g, then $\|f_{\varepsilon,u} - g\|_{L^\infty} > \varepsilon/2$ for any $u \ne u_0$. Hence, the error $\|f_{\varepsilon,u} - \mathcal{A}_n f_{\varepsilon,u}\|_{L^\infty}$ can be smaller than $\varepsilon/2$ for at

most 2^n values of u. Consequently, the error converges to zero with $n \to \infty$ for at most countably many $f_{\varepsilon,u}$'s. $\square$

The lack of convergence for the L^∞ approximation can be remedied by using a weaker metric. The first and straightforward choice is L^p with $1 \le p < \infty$. However, there is also another possibility – the *Skorohod metric*.

The Skorohod metric, see, e.g., [4, 11, 18, 26], was introduced in statistics after it had been observed that the L^∞ norm is not adequate for piecewise continuous functions (processes), since it makes the space of such functions nonseparable, i.e., too large. To explain this from a layman's perspective, let us consider for a moment functions $f_{1/2+1/n} := \mathbb{1}_{[0,1/2+1/n]}$. As $n \to \infty$, the functions $f_{1/2+1/n}$ *resemble* $f_{1/2} = \mathbb{1}_{[0,1/2]}$ *almost perfectly*. Yet $\|f_{1/2} - f_{1/2+1/n}\|_{L^\infty} = 1$ for all n. Even worse, the constant function $f^* \equiv 1/2$ is much closer to $f_{1/2}$ in the L^∞ norm than any $f_{1/2+1/n}$ in spite of the fact that $f_{1/2}$ and f^* are so different. The L^p norms (with $p < \infty$) do not have these bad properties; however, they *ignore* singular points, characteristics that might be important to preserve in some applications, e.g., in signal/image processing.

As we shall see, the Skorohod metric is free of all the problems above, and therefore turns out to be a suitable tool for working with discontinuous functions. It is defined as follows. Let f, g be two functions from L^∞. Their Skorohod distance is given by

$$\mathrm{dist}_{\mathrm{S}}(f,g) := \inf_{\lambda \in H} \left(\|f \circ \lambda - g\|_{L^\infty} + \|\lambda - \mathrm{id}\|_{L^\infty} \right), \tag{3}$$

where id is the identity function ($\mathrm{id}(x) = x \ \forall x$) and H is the class of homeomorphisms $\lambda : [0,T] \to [0,T]$ with $\lambda(0) = 0$ and $\lambda(T) = T$.[1]

Stating the definition in a less formal way, the Skorohod distance between f and g is small if there is a *change of variables* $y = \lambda(x)$ such that $f(\lambda(\cdot))$ is close (in the L^∞ topology) to g, and the change of variables λ is close (again in the L^∞ topology) to the identity function.

Another analogy comes from the heart of numerical analysis: $f(\lambda(\cdot))$ is a *slightly perturbed input function*, and any algorithm producing g close to it is *Strongly Numerically Stable*. Even a function as simple as $f_{1/3} = \mathbb{1}_{[0,1/3]}$ cannot be represented exactly in a *floating point arithmetic*, and the best one can do is to approximate it by *a slightly perturbed* function $\tilde{f}_{1/3} = \mathbb{1}_{[0,\mathrm{fl}(1/3)]}$, where $\mathrm{fl}(1/3)$ denotes the machine representation of 1/3. Clearly $\mathrm{dist}_{\mathrm{S}}(f_{1/3}, \tilde{f}_{1/3}) = |1/3 - \mathrm{fl}(1/3)|$ is very small (bounded by one third of the unit round-off); however, $\|f_{1/3} - \tilde{f}_{1/3}\|_{L^\infty} = 1$.

We believe that the Skorohod metric is an important way of measuring errors for a host of applications including image analysis/processing, where preserving discontinuities is essential. There $f(\lambda(\cdot))$ would correspond to a *slightly distorted* image of f.

Even though $\|f - g\|_{L^\infty}$ and $\mathrm{dist}_{\mathrm{S}}(f,g)$ differ so much, it is possible to show some relations. Obviously, we always have

$$\mathrm{dist}_{\mathrm{S}}(f,g) \le \|f - g\|_{L^\infty}.$$

[1]Originally (see, e.g., [18]) the Skorohod metric was defined using the sup (uniform) norm instead of L^∞, however, under the assumption that the functions are right-continuous. It should be clear that under such assumption, there is no difference between using the sup and the L^∞ norms, and our analysis can be carried over.

The following proposition might be known; however, we have not seen it in the literature.

Proposition 1. *Let the functions $f, g \in L^\infty$. If one of them satisfies a Lipschitz condition with constant L, then*

$$\|f - g\|_{L^\infty} \le \max(1, L) \cdot \mathrm{dist}_{\mathrm{S}}(f, g).$$

Hence,

$$\|f - g\|_{L^\infty} = \mathrm{dist}_{\mathrm{S}}(f, g) \quad \text{if } L \le 1.$$

Proof. Suppose f is Lipschitz a Lipschitz function. Then for every $x \in [0, T]$ and every $\lambda \in H$

$$\begin{aligned} |f(x) - g(x)| &\le |f(\lambda(x)) - g(x)| + |f(x) - f(\lambda(x))| \\ &\le |f(\lambda(x)) - g(x)| + L\,|x - \lambda(x)| \\ &\le \max(1, L)\,\left(|f(\lambda(x)) - g(x)| + |x - \lambda(x)|\right). \end{aligned}$$

Now taking the ess sup with respect to x and the inf with respect to λ completes the proof. □

Here is another property showing a relation between the Skorohod metric and the L^1 norm. For $f \in F_1^\infty$, let

$$c(f) := \max\left(1,\, \|f'\|_{L^\infty} + T^{-1} \sum_{j=1}^{k} |\Delta_j^{(0)}|\right),$$

where k is the number of discontinuities of f and $\Delta_j^{(0)}$, $1 \le j \le k$, are the corresponding discontinuity jumps.

Proposition 2. *For any $f, g \in F_1^\infty$ we have*

$$\|f - g\|_{L^1} \le \min\left(c(f), c(g)\right) \cdot T \cdot \mathrm{dist}_{\mathrm{S}}(f, g).$$

Proof. Without loss of generality, let $c(f) \le c(g)$ and the infimum in the definition of $\mathrm{dist}_{\mathrm{S}}(f, g)$ be attained for λ, so that $\mathrm{dist}_{\mathrm{S}}(f, g) = \|\lambda - \mathrm{id}\|_{L^\infty} + \|f \circ \lambda - g\|_{L^\infty}$ (otherwise we would do the standard δ-trick). Then

$$\int_0^T |f(x) - g(x)|\,\mathrm{d}x \le \int_0^T |f(\lambda(x)) - f(x)|\,\mathrm{d}x + \int_0^T |f(\lambda(x)) - g(x)|\,\mathrm{d}x.$$

The latter integral is upper bounded by $T\|f \circ \lambda - g\|_{L^\infty}$. To bound the former integral, we represent f (up to the finite set of discontinuity points s_j) as

$$f = f_1 + \sum_{j=1}^{k} \Delta_j^{(0)}\, \mathbb{1}_{[s_j, T]}$$

where $f_1 \in F_1^\infty \cap C$. Note that $\|f_1'\|_{L^\infty} = \|f'\|_{L^\infty}$ and

$$\mathbb{1}_{[s_j, T]} \circ \lambda - \mathbb{1}_{[s_j, T]} = \mathbb{1}_{[s_j, \lambda^{-1}(s_j))}$$

(by convention, $\mathbb{1}_{[a,b)} = -\mathbb{1}_{[b,a)}$ when $a > b$). Hence

$$\begin{aligned}
&\int_0^T |f(\lambda(x)) - f(x)|\mathrm{d}x \\
&\quad\le \int_0^T |f_1(\lambda(x)) - f_1(x)|\mathrm{d}x + \sum_{j=1}^k |\Delta_j^{(0)}| \int_0^T \left|\mathbb{1}_{[s_j,T]}(\lambda(x)) - \mathbb{1}_{[s_j,T]}(x)\right| \mathrm{d}x \\
&\quad\le T\|\lambda - \mathrm{id}\|_{L^\infty}\|f'\|_{L^\infty} + \sum_{j=1}^k |\Delta_j^{(0)}| \cdot |\lambda^{-1}(s_j) - s_j| \\
&\quad= T\|\lambda - \mathrm{id}\|_{L^\infty}\left(\|f'\|_{L^\infty} + T^{-1}\sum_{j=1}^k |\Delta_j^{(0)}|\right),
\end{aligned}$$

which completes the proof. □

Thus, in the Skorohod metric, any sequence of approximations converges at least as fast as it does in the L^∞ norm, and no faster than in the L^1 norm.

There are no such relations between the Skorohod distance and the L^p norms with $1 < p < \infty$. Indeed, if $f \equiv 0$ and $g_n = \mathbb{1}_{[0,1/n]}$, then $\|f - g_n\|_{L^p} = n^{-1/p}$ converges to zero whereas $\mathrm{dist}_{\mathrm{S}}(f, g_n) = 1$. On the other hand, if $f = \mathbb{1}_{[0,T/2]}$ and $g_n = \mathbb{1}_{[0,T/2+1/n]}$, then $\mathrm{dist}_{\mathrm{S}}(f, g_n) = n^{-1}$ converges faster than $\|f - g_n\|_{L^p} = n^{-1/p}$.

4. Nonadaptive algorithms

It is well known that, in the space of r-smooth functions $W_r = F_r^0$, piecewise polynomial interpolation of degree $r-1$ based on equidistant sampling leads to a nonadaptive algorithm with error bounded from above by $C\|f^{(r)}\|_{L^\infty} n^{-r}$ with $C > 0$ independent of f and n. Moreover, n^{-r} is the best rate of convergence among all (including adaptive) algorithms. It turns out that the presence of discontinuities causes all nonadaptive algorithms to have their errors bounded from below by $n^{-1/p}$. The purpose of this section is to show this fact.

We first show the following rather simple bounds.

Theorem 1. *Let $\mathcal{A}_n$ be an arbitrary nonadaptive algorithm. Let $\Delta > 0$.*

(i) There exists a piecewise constant function $f_1 \in F_r^1$ such that $|\Delta_1^{(0)}| \le \Delta$ and

$$\|f_1 - \mathcal{A}_n f_1\|_{L^p} \ge \frac{1}{2}\Delta\left(\frac{T}{n+1}\right)^{1/p}. \tag{4}$$

(ii) There exists a piecewise constant function $f_2 \in F_r^1$ such that $|\Delta_1^{(0)}| \le \Delta$ and

$$\mathrm{dist}_{\mathrm{S}}(f_2, \mathcal{A}_n f_2) \ge \frac{1}{2}\min\left(\frac{T}{n+1}, \Delta\right). \tag{5}$$

Moreover, there exists a piecewise constant function $f_3 \in F_r^2$ such that $|\Delta_1^{(0)}| + |\Delta_2^{(0)}| \le \Delta$ and

$$\mathrm{dist}_{\mathrm{S}}(f_3, \mathcal{A}_n f_3) \ge \frac{1}{4}\Delta. \tag{6}$$

Proof. Suppose that $\mathcal{A}_n$ computes f at points $x_1 < \cdots < x_n$. Let $x_0 = 0$ and $x_{n+1} = T$. Then there exist $0 < a < b < T$ such that $b - a \geq T/(n+1)$ and $[a,b] \subset [x_k, x_{k+1}]$ for some $0 \leq k \leq n$. Take now two functions, $g_1 = \Delta \mathbb{1}_{(a,T]}$ and $g_2 = \Delta \mathbb{1}_{[b,T]}$. Since g_1 and g_2 share the same information (i.e., $g_1(x_i) = g_2(x_i)$ for all $1 \leq i \leq n$) and $\|g_1 - g_2\|_{L^p} \geq \Delta(T/(n+1))^{1/p}$, the algorithm error cannot be smaller than $\Delta(T/(n+1))^{1/p}/2$ for at least one of the functions g_i. This proves (4). Similarly, (5) follows from the fact that for the same functions $\mathrm{dist}_{\mathrm{S}}(g_1, g_2) = \min(b-a, \Delta)$.

To show (6), we choose a, b as before and define $g_1 \equiv 0$ and $g_2 = \Delta \mathbb{1}_{(a,b)}/2$. Then, by Proposition 1, $\mathrm{dist}_{\mathrm{S}}(g_1, g_2) = \|g_2\|_{L^\infty} = \Delta/2$, and (6) follows. □

Theorem 1 is in the spirit of *the worst case setting*, since a single 'bad' function is selected depending on n. The following theorem shows that also asymptotically the L^p error of nonadaptive algorithms cannot decrease faster than $n^{-1/p}$. However, we have the lower bound n^{-1} for the asymptotic convergence in the Skorohod metric, independently of the number of singularities. As we shall see later, these lower bounds are sharp. Hence for nonadaptive algorithms, the switch from the worst case to the asymptotic setting only benefits the convergence in the Skorohod metric.

Let $f_{\varepsilon,u}$ be defined as in Lemma 1.

Theorem 2. *Let $1 \leq p \leq \infty$. Let $\{\mathcal{A}_n\}_{n\geq 1}$ be any sequence of nonadaptive algorithms. Then for any $f \in F_r^\infty$ and $\varepsilon \neq 0$ the sets*

$$A = \left\{ u \in (0,T) \;\middle|\; \lim_{n\to\infty} n^{1/p} \cdot \|f_{\varepsilon,u} - \mathcal{A}_n f_{\varepsilon,u}\|_{L^p} = 0 \right\}$$

and

$$A' = \left\{ u \in (0,T) \;\middle|\; \lim_{n\to\infty} n \cdot \mathrm{dist}_{\mathrm{S}}(f_{\varepsilon,u}, \mathcal{A}_n f_{\varepsilon,u}) = 0 \right\}$$

are of Lebesgue measure zero, i.e., $\mathcal{L}(A) = \mathcal{L}(A') = 0$.

Proof. We first show the theorem for the L^p norm. In view of Lemma 1 it suffices to consider $1 \leq p < \infty$.

Note that for any function g the set of $u \in (0,T)$ such that $\|f_{\varepsilon,u} - g\|_{L^p} \leq a$ is of measure at most $2(2a/|\varepsilon|)^p$. Indeed, for any u, v we have $\|f_{\varepsilon,u} - f_{\varepsilon,v}\|_{L^p} = |\varepsilon||u-v|^{1/p}$. Hence, if $\|f_{\varepsilon,u} - g\|_{L^p} \leq a$ and $|u-v| > (2a/|\varepsilon|)^p$, then

$$\begin{aligned} \|f_{\varepsilon,v} - g\|_{L^p} &\geq \|f_{\varepsilon,u} - f_{\varepsilon,v}\|_{L^p} - \|f_{\varepsilon,u} - g\|_{L^p} \\ &\geq |\varepsilon||u-v|^{1/p} - a \;>\; a. \end{aligned}$$

The proof is similar to the proof of Lemma 4 in [21]. That is, for $c > 0$, we define the sets

$$B_n^c := \left\{ u \in [0,T) \mid \|f_{\varepsilon,u} - \mathcal{A}_n f_{\varepsilon,u}\|_{L^p} \leq c(n+1)^{-1/p} \right\}.$$

Since the information takes no more than $n+1$ different values for functions $f_{\varepsilon,u}$, the algorithm $\mathcal{A}_n$ also takes at most $n+1$ different values for these functions. Hence

$$\mathcal{L}(B_n^c) \leq (n+1) 2 \left(\frac{2c}{|\varepsilon|(n+1)^{1/p}} \right)^p = 2 \left(\frac{2c}{|\varepsilon|} \right)^p.$$

Letting

$$A_n^c := \left\{ u \in [0,T) \mid \|f_{\varepsilon,u} - \mathcal{A}_l f_{\varepsilon,u}\|_{L^p} \leq c(l+1)^{-1/p} \;\forall l \geq n \right\},$$

we have $A_n^c \subset B_n^c$, $A_n^c \subset A_{n+1}^c$ for all n, and $A \subset \bigcup_{n=1}^\infty A_n^c$. Hence

$$\mathcal{L}(A) \le \lim_{n\to\infty} \mathcal{L}(A_n^c) \le 2\left(\frac{2c}{|\varepsilon|}\right)^p.$$

Since c can be arbitrarily small, $\mathcal{L}(A) = 0$.

To show the theorem for the Skorohod metric, observe first that

$$\mathrm{dist}_{\mathrm{S}}(f_{\varepsilon,u}, f_{\varepsilon,v}) \ \ge\ \min\left(|u-v|, \frac{|\varepsilon|}{1+\|f'\|_{L^\infty}}\right). \tag{7}$$

Indeed, let $\lambda \in H$ be arbitrary. If $\lambda(u) = v$, then $\|\lambda - \mathrm{id}\|_{L^\infty} \ge |u-v|$. On the other hand, if $\lambda(u) \ne v$ and $\|\lambda - \mathrm{id}\|_{L^\infty} \le |\varepsilon|/(1+\|f'\|_{L^\infty})$, then there exists $z \in (0,T)$ for which

$$|\mathbb{1}_{[0,u)}(\lambda(z)) - \mathbb{1}_{[0,v)}(z)| \ = \ 1.$$

Then

$$\begin{aligned} |f_{\varepsilon,u}(\lambda(z)) - f_{\varepsilon,v}(z)| \ &= \ |(f(\lambda(z)) - f(z)) + \varepsilon(\mathbb{1}_{[0,u)}(\lambda(z)) - \mathbb{1}_{[0,v)}(z))| \\ &\ge \ -L_f \cdot \|\lambda - \mathrm{id}\|_{L^\infty} + |\varepsilon| \ \ge\ \frac{|\varepsilon|}{1+\|f'\|_{L^\infty}}, \end{aligned}$$

which shows (7).

The proof then proceeds as for the L^p norm by noting that for $a < (1/2)|\varepsilon|/(1+\|f'\|_{L^\infty})$ the set of u for which $\mathrm{dist}_{\mathrm{S}}(f_{\varepsilon,u}, g) \le a$ is of measure at most $2a$. □

To complete this section, we now show that the estimates of Theorems 1 and 2 are sharp. For $n \ge 2$, consider the piecewise constant approximation

$$\mathcal{A}_n^{\mathrm{non}} f \ := \ f(0)\,\mathbb{1}_{[0,c_1)} + \sum_{i=2}^{n-1} f(x_i)\,\mathbb{1}_{[c_{i-1},c_i)} + f(T)\,\mathbb{1}_{[c_{n-1},T]},$$

where

$$x_i := \frac{i-1}{n-1}\,T \qquad \text{and} \qquad c_i := \frac{x_i + x_{i+1}}{2}.$$

Proposition 3. *Suppose that $f \in F_1^\infty$ has exactly k discontinuities s_j with jumps $\Delta_j^{(0)}$, $1 \le j \le k$. Then*

$$\|f - \mathcal{A}_n^{\mathrm{non}} f\|_{L^p} \le \frac{T}{2(n-1)}\left(\frac{T}{p+1}\right)^{1/p}\|f'\|_{L^\infty} + \left(\frac{T}{2(n-1)}\right)^{1/p}\left(\sum_{j=1}^k |\Delta_j^{(0)}|\right). \tag{8}$$

Furthermore, if $k = 1$ or

$$2 \le k \le n-1 \quad \textit{and} \quad \frac{T}{n-1} < \min_{i\ne j}|s_i - s_j|, \tag{9}$$

then

$$\mathrm{dist}_{\mathrm{S}}(f, \mathcal{A}_n^{\mathrm{non}} f) \le \frac{T}{2(n-1)}\|f'\|_{L^\infty} + \min\left(\frac{T}{2(n-1)}(1+\|f'\|_{L^\infty}), \max_{1\le j\le k}|\Delta_j^{(0)}|\right). \tag{10}$$

Proof. We first prove (8). Consider an arbitrary interval $[x_i, c_i]$, where $1 \le i \le n-1$. Let P_i be the set of all indices j such that $s_j \in_f [x_i, c_i]$. Then for any $x \in [x_i, c_i]$ we have

$$f(x) = g_i(x) + \sum_{j\in P_i} \Delta_j^{(0)}\,\mathbb{1}_{I_{i,j}}(x),$$

for some $g_i \in F_1^\infty \cap C$ and $I_{i,j} = [s_j, c_i]$ or $I_{i,j} = (s_j, c_i]$. Hence

$$|f(x) - (\mathcal{A}_n^{\rm non} f)(x)| = |f(x) - f(x_i)| \le (x - x_i)\|f'\|_{L^\infty} + \sum_{j \in P_i} |\Delta_j^{(0)}|.$$

Similarly, for any $x \in [c_i, x_{i+1}]$, $1 \le i \le n-1$, we have

$$|f(x) - (\mathcal{A}_n^{\rm non} f)(x)| \le (x_{i+1} - x)\|f'\|_{L^\infty} + \sum_{j \in Q_i} |\Delta_j^{(0)}|,$$

where Q_i is the set of all indices j such that $s_j \in_f [c_i, x_{i+1}]$. Observe that any s_j is in exactly one of the sets P_i or Q_i. It now follows that the error can be bounded from above by the sum of two functions: a saw-like function

$$A(x) = \|f'\|_{L^\infty} \min_{1 \le j \le n} |x - x_j|$$

and the piecewise constant function

$$B(x) = \sum_{i=1}^{n-1} \left(\mathbb{1}_{[x_i, c_i)}(x) \sum_{j \in P_i} |\Delta_j^{(0)}| + \mathbb{1}_{[c_i, x_{i+1}]}(x) \sum_{j \in Q_i} |\Delta_j^{(0)}| \right).$$

Straightforward calculations give

$$\|A\|_{L^p} = \frac{T}{2(n-1)} \left(\frac{T}{p+1} \right)^{1/p} \|f'\|_{L^\infty}$$

and

$$\begin{aligned} \|B\|_{L^p} &= \left(\frac{T}{2(n-1)} \right)^{1/p} \left(\sum_{i=1}^{n-1} \left(\left(\sum_{j \in P_i} |\Delta_j^{(0)}| \right)^p + \left(\sum_{j \in Q_i} |\Delta_j^{(0)}| \right)^p \right) \right)^{1/p} \\ &\le \left(\frac{T}{2(n-1)} \right)^{1/p} \left(\sum_{j=1}^{k} |\Delta_j^{(0)}| \right). \end{aligned}$$

The bound (8) is now obtained by adding $\|A\|_{L^p}$ and $\|B\|_{L^p}$.

We now show (10). The condition (9) implies that there is no more than one discontinuity in any $[x_j, x_{j+1}]$ w.r.t. f. Assume without loss of generality that $s_j \in_f [x_j, x_{j+1}]$, where $1 \le j \le n-1$. (Otherwise we would insert some trivial jumps.) Then for all $x \in [x_j, x_{j+1}]$ we have

$$|f(x) - (\mathcal{A}_n^{\rm non} f)(x)| \le \frac{T}{2(n-1)} \|f'\|_{L^\infty} + |\Delta_j^{(0)}|$$

which implies

$$\text{dist}_{\rm S}(f, \mathcal{A}_n^{\rm non} f) \le \|f - \mathcal{A}_n^{\rm non} f\|_{L^\infty} \le \frac{T}{2(n-1)} \|f'\|_{L^\infty} + \max_{1 \le j \le k} |\Delta_j^{(0)}|. \tag{11}$$

On the other hand, consider the homeomorphism λ defined as the piecewise linear function interpolating the following data:

$$\lambda(0) = 0, \quad \lambda(T) = T, \quad \text{and} \quad \lambda(c_j) = s_j \qquad 1 \le j \le n-1.$$

Then

$$\|\lambda - \text{id}\|_{L^\infty} = \max_{1 \le j \le n-1} |c_j - s_j| \le \frac{T}{2(n-1)}.$$

Since

$$\lambda([c_j, c_{j+1}]) = [s_j, s_{j+1}] \quad \text{for } 0 \le j \le n-1$$

(where $c_0 = s_0 = 0$ and $c_n = s_n = T$), for any $x \in (c_j, c_{j+1})$ we have

$$|f(\lambda(x)) - (\mathcal{A}_n^{\text{non}} f)(x)| \le \|f'\|_{L^\infty} |\lambda(x) - x_j|$$

and $|\lambda(x) - x_j| \le T/(n-1)$. Hence

$$\|f \circ \lambda - \mathcal{A}_n^{\text{non}} f\|_{L^\infty} \le \frac{T}{n-1} \|f'\|_{L^\infty}$$

and

$$\text{dist}_{\text{S}}(f, \mathcal{A}_n^{\text{non}} f) \le \frac{T}{2(n-1)} (1 + 2\|f'\|_{L^\infty}). \tag{12}$$

Combining (11) and (12) we get (10). □

5. Adaptive algorithms: Single discontinuity

In this section, we provide an adaptive algorithm whose worst case error is proportional to n^{-r}. Clearly, this is a significant improvement over the convergence of nonadaptive methods. We stress that this result holds true because we have restricted the function class to F_r^1. As shown in the next section, adaption loses its superiority in the worst case setting for classes of functions with possibly more than one singular point.

Our adaptive algorithm consists of two stages. First, it tries to detect/localize the discontinuity using adaptive sampling, and next it approximates the function by piecewise polynomial interpolation at points that depend on the location of the singularity.

5.1. **Detection mechanism.** Our mechanism for detecting the singular point s_f in $(0, T)$ with 'large' discontinuity jumps is based on the one already developed in [21]. A theoretical justification of that mechanism is based on Lemma 2 and Lemma 3 that are basically taken from [21]; however, the formulation of Lemma 2 is slightly different than the one in [21]. For completeness, we provide a proof in the Appendix.

For $m \ge 1$, let $h = T/m$ and $t_i := ih$ for all (not necessarily integer) i. Denote by $f[t_i, t_{i+1}, \ldots, t_{i+j}]$ the jth order divided difference of f with respect to $t_i, \ldots, t_{i+j}$. We also define the constant

$$M_r := \max_{0 \le \tau \le 1} \|A_\tau^{-1}\|_\infty,$$

where $A_\tau = (a_{i,j}(\tau))_{i,j=0}^{r-1}$ with

$$a_{i,j}(\tau) = \sum_{l=1}^{i+1} b_{l-i+r-1} \frac{(l-\tau)^j}{j!} \quad \text{and} \quad b_u = \prod_{u \ne l = 0}^{r} (u-l)^{-1}.$$

Lemma 2. *Suppose $f \in F_r^1$ with the singularity $s_f \in_f [t_k, t_{k+1}]$, where $r - 1 \le k \le m - r$. If the divided differences are bounded by*

$$|f[t_i, t_{i+1}, \ldots, t_{i+r}]| \le B \qquad \textit{for} \quad k + 1 - r \le i \le k,$$

then the discontinuity jumps are bounded by

$$|\Delta_f^{(j)}| \le M_r \left(B + \frac{1}{r!} \|f^{(r)}\|_{L^\infty(t_{k+1-r}, t_{k+r})} \right) \cdot h^{r-j} \qquad \textit{for} \quad 0 \le j \le r-1.$$

Lemma 3. *For any integer $i \in \{r, \dots, m\}$,*

$$|f[t_i, t_{i+1}, \dots, t_{i+r}]| \le \max_{2i \le j \le 2i+r} |f[t_{j/2}, t_{(j+1)/2}, \dots, t_{(j+r)/2}]|.$$

Suppose now that the conditions of Lemma 2 are satisfied, i.e.,

$$s_f \in_f [t_{r-1}, t_{m-r+1}]. \tag{13}$$

Then we can compute all possible divided differences $d_i^0 := f[t_i, \dots, t_{i+r}]$, $0 \le i \le m-r$, and compare them to a given *threshold* D. If none of them exceeds D, then by Lemma 2 with $B = D$, the discontinuity jumps are 'small', i.e.,

$$|\Delta_f^{(j)}| \le M_r \left(D + \frac{1}{r!} \|f^{(r)}\|_{L^\infty} \right) h^{r-j}, \qquad 0 \le j \le r-1. \tag{14}$$

Otherwise, if the largest $|d_i^0|$ exceeds D, then we can perform a few, say κ, bisection-like *adaptive* steps (with the use of Lemma 3), each time decreasing the mesh-size twice, to find q such that $|\hat{d}| := |f[t_{q/2^\kappa}, \dots, t_{(q+r)/2^\kappa}]| \ge |d_i^0|$.

Now, if $|\hat{d}| \le \|f^{(r)}\|_{L^\infty}/r!$, then by Lemma 2 with $B = \|f^{(r)}\|_\infty/r!$, we have

$$|\Delta_f^{(j)}| \le \frac{2M_r}{r!} \|f^{(r)}\|_{L^\infty} h^{r-j} \qquad \text{for all} \quad 0 \le j \le r-1, \tag{15}$$

and if $|\hat{d}| > \|f^{(r)}\|_{L^\infty}/r!$, then

$$s_f \in_f [t_{q/2^\kappa}, t_{(q+r)/2^\kappa}]. \tag{16}$$

In addition, if (16) holds, then f is r-smooth on $[0, t_{q/2^\kappa}]$ and on $[t_{(q+r)/2^\kappa}, T]$.

Obviously, for fixed $f \in F_r^1$ with nontrivial discontinuity of one of the derivatives $f^{(j)}$, where $0 \le j \le r-1$, condition (13) is satisfied for the mesh-size h sufficiently small (or m sufficiently large). Even more, if $D < \infty$, then for h small enough the largest $|d_i^0|$ exceeds D so that we only have (15) or (16). However, we aim at stronger, nonasymptotic results and therefore have to extend applicability of the detection mechanism to the intervals $(t_{(r-1)/2^\kappa}, t_{r-1}]$ and $[t_{m-r+1}, t_{m-(r-1)/2^\kappa})$. As we shall see, this is sufficient to construct an algorithm with small worst case error.

The extension works for $m \ge 2(r-1)$ and is done by computing additional divided differences with decreasing mesh-size in the vicinities of 0 and T. Specifically, for $1 \le j \le \kappa$, $0 \le i \le 2(r-1)$, and $2^j m - 3(r-1) - 1 \le i \le 2^j m - r$ we compute

$$d_i^j := f[t_{i/2^j}, t_{(i+1)/2^j}, \dots, t_{(i+r)/2^j}].$$

If $s_f \in_f [t_{(r-1)/2^j}, t_{(r-1)/2^{j-1}}]$ and $|d_i^j| \le D$ for all i (or $s_f \in_f [t_{m-(r-1)/2^{j-1}}, t_{m-(r-1)/2^j}]$ and $|d_i^j| \le D$ for all i), then Lemma 2 can be applied with $B = D$ and the mesh-size h replaced by $h/2^j$.

Observe that the detection mechanism with additional divided differences uses

$$(m+1) + \kappa(r + 2\lceil 3r/2 \rceil)$$

function evaluations in the worst case.

Before providing a corresponding pseudocode, we comment on the choice of the threshold D. Observe that if

$$D \le \|f^{(r)}\|_{L^\infty}/r! \tag{17}$$

and $s_f \in_f [(r-1)h/2^\kappa, T - (r-1)h/2^\kappa]$, then either inequalities (15) hold true, or $s_f \in_f [qh/2^\kappa, (q+r)h/2^\kappa]$ (or both). It is clear that the best choice of D would be $D = \|f^{(r)}\|_{L^\infty}/r!$ since then the adaptive steps are performed only when the

singularity has been detected. Unfortunately (a good estimate for) $\|f^{(r)}\|_{L^\infty}$ is usually *not* available, and this is what we also assume in the present paper. In such cases, to avoid a possible overestimation of $|\Delta_f^{(j)}|$ in (14), one can take $D = 0$. Then (17) holds for any f; however, the detection part is triggered regardless of whether f has a singularity or not.

In the following procedure the parameters are: the smoothness r, a positive integer sequence $\{\kappa_m\}_{m\geq 1}$, and a threshold D.

```
procedure DETECT;
begin
  input m ≥ 2(r − 1); h := T/m;
  κ := κ_m; δ := (r − 1)h2^{−κ};
  for i := 0 to m − r do d_i^0 := f[t_i, ..., t_{i+r}];
  for j := 1 to κ do
  begin
    for i := 0 to 2(r − 1) do d_i^j := f[t_{i/2^j}, ..., t_{(i+r)/2^j}];
    for i := 2^j m − 3(r − 1) − 1 to 2^j m − r do d_i^j := f[t_{i/2^j}, ..., t_{(i+r)/2^j}]
  end;
  (i*, j*) := arg max_{(i,j)} |d_i^j|;
  if |d_{i*}^{j*}| ≤ D then return ∅ else
  begin
    for j := j* + 1 to κ do
    begin
      for i := 0 to r do e_i := f[t_{i*+i/2^j}, ..., t_{i*+(i+r)/2^j}];
      i* := i* + (arg max_i |e_i|) /2^j
    end;
    v_l := max(δ, i*h); v_r := min((i* + r2^{−κ})h, T − δ);
    return [v_l, v_r]
  end
end.
```

5.2. **Adaptive piecewise polynomial interpolation.** In this section, we define and analyze the error of piecewise polynomial interpolation that is based on the adaptive partition of $[0, T]$ produced by DETECT of Section 5.1.

For fixed points τ_i satisfying

$$0 \leq \tau_1 < \tau_2 < \cdots < \tau_r \leq 1,$$

let $\mathcal{P}_r(f; a, b)$ be the polynomial of degree at most $r-1$ interpolating f at $a + \tau_j(b-a)$, with $1 \leq j \leq r$. Further, for any collection U of points $u_0 < u_1 < u_2 < \cdots < u_k$, let $\overline{\mathcal{P}}_r(f; U)$ be the piecewise polynomial interpolation of f determined by U, i.e.,

$$\overline{\mathcal{P}}_r(f; U) = \sum_{i=1}^{k} \mathcal{P}_r(f; u_{i-1}, u_i)\, \mathbb{1}_{[u_{i-1}, u_i)}.$$

Fix a threshold $D \geq 0$ and an integer sequence $\{\kappa_m\}_{m\geq 1}$ satisfying $\kappa_m \geq \log_2(r-1)$, so that

$$\delta_m := (r-1)h2^{-\kappa_m} \leq \frac{T}{m} \quad \text{for all } m.$$

Our algorithm $\mathcal{A}_n^{\rm ad}$ is determined by r, $\{\kappa_m\}_m$, and D, and is defined as follows. We choose $m \ge 2(r-1)$ and run DETECT. If the procedure returns $\emptyset$, then we let

$$U_m := \{t_i \mid 1 \le i \le m-1\} \cup \{\delta_m, T-\delta_m\}$$

and

$$\begin{aligned}\mathcal{A}_n^{\rm ad} f \;&:=\; \overline{\mathcal{P}}_r(f;U_m)\\ &\quad + f(0)\,\mathbb{1}_{[0,\delta_m/2)} \,+\, f(\delta_m)\,\mathbb{1}_{[\delta_m/2,\delta_m)}\\ &\quad + f(T-\delta_m)\,\mathbb{1}_{[T-\delta_m,T-\delta_m/2)} \,+\, f(T)\,\mathbb{1}_{[T-\delta_m/2,T]}.\end{aligned}$$

Otherwise, the procedure produces an interval $[v_l, v_r]$ with $\delta_m \le v_l \le v_r \le T-\delta_m$. Then we let

$$U_m^l := \{t_i \mid 1 \le i \le m-1,\, t_i < v_l\} \cup \{\delta_m, v_l\},$$

$$U_m^r := \{t_i \mid 1 \le i \le m-1,\, t_i > v_r\} \cup \{v_r, T-\delta_m\},$$

and construct our algorithm as

$$\begin{aligned}\mathcal{A}_n^{\rm ad} f \;&:=\; \overline{\mathcal{P}}_r(f;U_m^l) \,+\, \overline{\mathcal{P}}_r(f;U_m^r)\\ &\quad + f(0)\,\mathbb{1}_{[0,\delta_m/2)} \,+\, f(\delta_m)\,\mathbb{1}_{[\delta_m/2,\delta_m)}\\ &\quad + f(v_l)\,\mathbb{1}_{[v_l,(v_l+v_r)/2)} \,+\, f(v_r)\,\mathbb{1}_{[(v_l+v_r)/2,v_r)}\\ &\quad + f(T-\delta_m)\,\mathbb{1}_{[T-\delta_m,T-\delta_m/2)} \,+\, f(T)\,\mathbb{1}_{[T-\delta_m/2,T]}.\end{aligned}$$

That is, $\mathcal{A}_n^{\rm ad} f$ is the piecewise polynomial of degree $r-1$ except on $[0,\delta_m]$, $[T-\delta_m, T]$, and $[v_l, v_r]$ (if this interval is created), where it is given by a piecewise constant function.

Note that in the worst case the algorithm performs

$$n = n(m) := (r+1)(m+1) + \kappa_m\,(r + 2\lceil 3r/2\rceil) + 1$$

function evaluations (those related to the detection mechanism plus those related to the piecewise polynomial interpolation).

To analyze the error of $\mathcal{A}_n^{\rm ad}$, we need some auxiliary results. Let $1 \le p \le \infty$. It is well known that if $f \in W_r(a,b)$, i.e., f is r-smooth on $[a,b]$, then

$$\|f - \mathcal{P}_r(f;a,b)\|_{L^p(a,b)} \;\le\; \alpha \cdot \|f^{(r)}\|_{L^\infty(a,b)}\,(b-a)^{r+1/p}, \tag{18}$$

where

$$\alpha = \alpha(r,p) := \frac{1}{r!}\left\| \prod_{j=1}^{r} |\cdot - \tau_j| \right\|_{L^p(0,1)}.$$

(Here and later we adopt the convention that $1/p = 0$ when $p = \infty$.)

Inequality (18) can be easily generalized to the case of functions with one singularity as follows. For $0 \le j \le r-1$, let

$$\alpha_j \;:=\; \alpha_j(r,p) \;:=\; \max_{0\le s\le 1} \|\phi_{j,s} - \mathcal{P}_r(\phi_{j,s};0,1)\|_{L^p(0,1)},$$

where

$$\phi_{j,s}(x) \;=\; \frac{(x-s)^j}{j!}\,\mathbb{1}_{[s,T]}(x).$$

Then we have the following bound, the proof of which is in the Appendix.

Lemma 4. *Let $1 \le p \le \infty$. Let $f \in F_r^1$ with the singular point $s_f \in_f [a,b]$ and discontinuity jumps $\Delta_f^{(j)}$, where $0 \le j \le r-1$. Then*

$$\begin{aligned} &\|f - \mathcal{P}_r(f;a,b)\|_{L^p(a,b)} \\ &\quad \le \quad \alpha \cdot \|f^{(r)}\|_{L^\infty(a,b)}(b-a)^{r+1/p} + \sum_{j=0}^{r-1} \alpha_j \cdot |\Delta_f^{(j)}|(b-a)^{j+1/p}. \end{aligned}$$

We are now ready to state and prove the main result about the L^p error of our adaptive algorithm. Let

$$D_f := \max\left(D, \frac{\|f^{(r)}\|_{L^\infty}}{r!}\right).$$

Theorem 3. *Let $1 \le p < \infty$. If the sequence $\{\kappa_m\}_{m\ge1}$ satisfies $\kappa_m = O(m)$, then $\mathcal{A}_n^{\mathrm{ad}}$ uses $n = O(m)$ function evaluations. Furthermore, for any $f \in F_r^1$ and $m \ge 2(r-1)$ we have*

$$\|f - \mathcal{A}_n^{\mathrm{ad}} f\|_{L^p} \le C \cdot \left(D_f m^{-r} + \delta_m^{1/p}(|\Delta_f^{(0)}| + \delta_m \|f'\|_{L^\infty})\right), \tag{19}$$

for some C independent of f. If in addition

$$\lim_{m\to\infty} \kappa_m - (rp-1)\log_2 m = \infty, \tag{20}$$

then

$$\lim_{m\to\infty} \|f - \mathcal{A}_n^{\mathrm{ad}} f\|_{L^p} \cdot m^r = T^r \alpha \|f^{(r)}\|_{L^p}. \tag{21}$$

Proof. First consider the case when s_f is w.r.t. f in one of the 'small' subintervals $[0,\delta_m]$, $[T-\delta_m]$, or $[v_l, v_r]$, each of length at most δ_m. Then the L^p error of $\mathcal{A}_n^{\mathrm{ad}}$ can be upper bounded by

$$\begin{aligned} \|f - \mathcal{A}_n^{\mathrm{ad}} f\|_{L^p} \quad \le \quad & T^{1/p}\alpha\|f^{(r)}\|_{L^\infty} h^r + 2 \cdot \frac{\delta_m^{1+1/p}}{2(1+p)^{1/p}}\|f'\|_{L^\infty} \\ & + \left(\frac{\delta_m^{1+1/p}}{2(1+p)^{1/p}}\|f'\|_{L^\infty} + \left(\frac{\delta_m}{2}\right)^{1/p} |\Delta_f^{(0)}|\right), \end{aligned}$$

where the three successive components of the sum are: the L^p error outside of the 'small' subintervals derived from (18), the error on the two (or one) 'small' subintervals not containing s_f (which follows from (8) of Proposition 3 with $n = 2$ and T replaced by δ_m), and the error on the 'small' subinterval containing s_f (which again follows from (8)).

Now consider the second, opposite case when s_f is w.r.t. f in one of the 'large' subintervals of length at most h. This is possible only when all the divided differences computed in lines 04 to 09 of DETECT are not larger than D_f and, consequently, the jumps $|\Delta_f^{(j)}| \le 2M_r D_f h^{r-j}$. Hence in this case, the error can be bounded by

$$\begin{aligned} \|f - \mathcal{A}_n^{\mathrm{ad}} f\|_{L^p} \quad \le \quad & T^{1/p}\alpha\|f^{(r)}\|_{L^\infty} h^r + 3 \cdot \frac{\delta_m^{1+1/p}}{2(1+p)^{1/p}}\|f'\|_{L^\infty} \\ & + \left(\alpha\|f^{(r)}\|_{L^\infty} + \frac{2M_r}{r!}\left(\sum_{j=0}^{r-1}\alpha_j\right) D_f\right) h^{r+1/p}, \end{aligned}$$

where the last component of the sum is the error on the 'large' interval containing s_f, derived from Lemma 4 and inequalities (14) and (15).

Inequality (19) is now obtained by combining both cases.

To prove (21), observe first that $s_f \in_f [v_l, v_r]$ for sufficiently large m. For any interval $[t_{i-1}, t_i]$ that has empty intersection with $(0, \delta_m) \cup (v_l, v_r) \cup (T - \delta_m, T)$ (i.e., except for 3 or 4 intervals), the approximation $(\mathcal{A}_n^{\mathrm{ad}} f)(x) = \mathcal{P}_r(f; t_{i-1}, t_i)(x)$. Hence for any such interval the error can be bounded by

$$\alpha |f^{(r)}(\xi_i^-)| h^{r+1/p} \le \|f - \mathcal{A}_n f\|_{L^p(t_{i-1},t_i)} \le \alpha |f^{(r)}(\xi_i^+)| h^{r+1/p},$$

where $\xi_i^-, \xi_i^+ \in [t_{i-1}, t_i]$. Since by (20) the error on the 'small' subintervals is asymptotically negligible, it follows that

$$h^{-rp} \|f - \mathcal{A}_n^{\mathrm{ad}} f\|_{L^p}^p = h^{-rp} \sum_{i=1}^m \|f - \mathcal{A}_n^{\mathrm{ad}} f\|_{L^p(t_{i-1},t_i)}^p$$

can be asymptotically bounded from below and from above by Riemann sums

$$\sum_{j=1}^m h \alpha^p |f^{(r)}(\xi_j^{\pm})|^p$$

of the function $\alpha^p |f^{(r)}|^p$. By definition of the space, any function $f \in F_r^\infty$ is Riemann integrable. (If necessary, we formally set zero for $f^{(r)}$ at points where this derivative is not defined.) Hence we obtain (21) by letting $m \to \infty$. □

A corresponding result about the error of $\mathcal{A}_n^{\mathrm{ad}}$ with respect to the Skorohod metric is as follows.

Theorem 4. *For any $f \in F_r^1$ and $m \ge 2(r-1)$ we have*

$$\mathrm{dist}_{\mathrm{S}}(f, \mathcal{A}_n^{\mathrm{ad}} f) \le \hat{C} \cdot \left(D_f m^{-r} + \delta_m \|f'\|_{L^\infty} + \min(\delta_m, |\Delta_f^{(0)}|) \right), \tag{22}$$

for some $\hat{C}$ independent of f. Moreover, if

$$\lim_{m \to \infty} \kappa_m - (r-1) \log_2 m = \infty,$$

then

$$\lim_{m \to \infty} \mathrm{dist}_{\mathrm{S}}(f, \mathcal{A}_n^{\mathrm{ad}} f) \cdot m^r = T^r \alpha \|f^{(r)}\|_{L^\infty}. \tag{23}$$

Proof. First consider the case when s_f is in one of the three 'small' intervals of length δ_m. We can assume without loss of generality that $s_f \in_f [v_l, v_r]$. Then, taking $\lambda = \mathrm{id}$ we have

$$\mathrm{dist}_{\mathrm{S}}(f, \mathcal{A}_n^{\mathrm{ad}} f) \le \max \left(|\Delta_f^{(0)}| + \frac{\delta_m}{2} \|f'\|_{L^\infty}, \alpha \|f^{(r)}\|_{L^\infty} h^r \right).$$

On the other hand, define the homeomorphism λ as the piecewise linear interpolation of the data: $\lambda(0) = 0$, $\lambda(T) = T$, and $\lambda(c) = s_f$ with $c = (v_l + v_r)/2$. Then $\|\lambda - \mathrm{id}\|_{L^\infty} = \delta_m/2$. For $x \notin (v_l, v_r)$ the error can be estimated as

$$\begin{aligned} |f(\lambda(x)) - (\mathcal{A}_n^{\mathrm{ad}} f)(x)| &\le |f(\lambda(x)) - f(x)| + |f(x) - (\mathcal{A}_n^{\mathrm{ad}} f)(x)| \\ &\le \begin{cases} \delta_m \|f'\|_{L^\infty} & \text{if } x \in [0, \delta_m] \cup [T - \delta_m, T], \\ \delta_m \|f'\|_{L^\infty}/2 + \alpha \|f^{(r)}\|_{L^\infty} h^r & \text{otherwise.} \end{cases} \end{aligned}$$

For $x \in (v_l, v_r)$ we show, similar to the proof of (10), that the error is at most $\delta_m \|f'\|_{L^\infty}$. Thus, in this case we also have

$$\begin{aligned} \mathrm{dist}_{\mathrm{S}}(f, \mathcal{A}_n^{\mathrm{ad}} f) &\le \|\lambda - \mathrm{id}\|_{L^\infty} + \|f \circ \lambda - \mathcal{A}_n^{\mathrm{ad}} f\|_{L^\infty} \\ &\le \frac{\delta_m}{2}(1 + \|f'\|_{L^\infty}) + \max\left(\frac{\delta_m}{2}\|f'\|_{L^\infty}, \alpha \|f^{(r)}\|_{L^\infty} h^r\right). \end{aligned}$$

Now consider the case when s_f is not in any of the 'small' intervals. Then we have (14) or (15) and

$$\begin{aligned} \mathrm{dist}_{\mathrm{S}}(f, \mathcal{A}_n^{\mathrm{ad}} f) &\le \|f - \mathcal{A}_n^{\mathrm{ad}} f\|_{L^\infty} \\ &\le \max\left(\frac{\delta_m}{2}\|f'\|_{L^\infty}\left(\alpha\|f^{(r)}\|_{L^\infty} + \frac{2M_r}{r!}\left(\sum_{j=0}^{r-1}\alpha_j\right) D_f\right) h^r\right). \end{aligned}$$

Inequality (22) is obtained by combining both cases.

To show (23) we use arguments similar to those in the proof of (21). □

5.3. **Worst case setting.** The bound (19) in Theorem 3 establishes the following result in the *worst case setting*. Let $\mathcal{F}_r^1$ be the class of functions $f \in F_r^1$ with uniformly bounded jumps $|\Delta_f^{(0)}|$ and first and rth derivatives of f, i.e.,

$$\mathcal{F}_r^1 = \mathcal{F}_r^1(L_r, L_1, D_0) := \{\, f \in F_r^1 \mid \|f^{(r)}\|_{L^\infty} \le L_r, \|f'\|_{L^\infty} \le L_1, |\Delta_f^{(0)}| \le D_0 \,\}$$

for some finite L_1, L_r, and D_0. Then the worst case L^p error of $\mathcal{A}_n^{\mathrm{ad}}$ with respect to $\mathcal{F}_r^1$ satisfies

$$\sup_{f \in \mathcal{F}_r^1} \|f - \mathcal{A}_n^{\mathrm{ad}} f\|_{L^p} \le C \cdot n^{-r},$$

where C is independent of n and f (but it depends on L_r, L_1, D_0). Similarly, for the Skorohod metric we have

$$\sup_{f \in \hat{\mathcal{F}}_r^1} \mathrm{dist}_{\mathrm{S}}(f, \mathcal{A}_n^{\mathrm{ad}}) \le \hat{C} \cdot n^{-r}$$

with respect to the class

$$\hat{\mathcal{F}}_r^1 = \mathcal{F}_r^1(L_r, L_1, \infty) := \{\, f \in F_r^1 \mid \|f^{(r)}\|_{L^\infty} \le L_r, \|f'\|_{L^\infty} \le L_1 \,\}.$$

Note that $\mathcal{F}_r^1 \subset \hat{\mathcal{F}}_r^1$.

Remark 1. It is clear that the adaptive algorithm $\mathcal{A}_n^{\mathrm{ad}}$ can be generalized by applying interpolation (or extrapolation) by polynomials of order $s \in \{0, 1, \ldots, r\}$ in the intervals $[0, \delta_m]$, $[v_l, v_r]$, $[T - \delta_m, T]$ (where zero approximation is used for $s = 0$). It can be shown that the L^p error of such a modified algorithm $\widetilde{\mathcal{A}}_n^{\mathrm{ad}}$ can be bounded as

$$\|f - \widetilde{\mathcal{A}}_n^{\mathrm{ad}} f\|_{L^p} \le C \cdot \left(\|f^{(r)}\|_{L^\infty} n^{-r} + \delta_m^{1/p}\Big(\sum_{j=0}^{s-1} \delta_m^j |\Delta_f^{(j)}| + \delta_m^s \|f^{(s)}\|_{L^\infty}\Big)\right),$$

where dependence on the successive discontinuity jumps $|\Delta_f^{(j)}|$ rapidly decreases (e.g., if $\kappa_m = m$, then $\delta_m^{1/p}$ vanishes at an exponential rate). Consequently, we have the corresponding worst case results for the classes

$$\mathcal{F}_r^1(s) := \{\, f \in F_r^1 \mid \|f^{(r)}\|_{L^\infty} \le L_r, |\Delta_f^{(j)}| \le D_j, 0 \le j \le s-1, \|f^{(s)}\|_{L^\infty} \le L_s \,\}.$$

We chose to work with $s = 1$ for a number of reasons. First, dependence of the error on quantities other than $f^{(r)}$ is practically negligible. Second, if κ_m increases faster than $\log_2 m$ and slower than m, then the same algorithm $\mathcal{A}_n^{\rm ad}$ (i.e., with the same set of parameters) satisfies the estimates of Theorems 3 and 4. Third, in the Skorohod metric, the algorithm $\widetilde{\mathcal{A}}_n^{\rm ad}$ does not work well for $s = 0$, and taking $s \ge 2$ is useless because in this way dependence on f' cannot be removed from the error formula, due to the L^∞ norm of the homeomorphism λ.

5.4. **Lower bounds.** It is well known that for functions with no singularities ($f \in W_r = F_r^0$) the worst case L^p error ($1 \le p \le \infty$) is of order n^{-r} with respect to the class of functions with uniformly bounded $\|f^{(r)}\|_{L^\infty}$. Moreover, this convergence cannot be improved, and uniform boundeness of the rth derivative is crucial.

One can ask if a similar result holds true in F_r^1. That is, if it is possible to construct an algorithm with the worst case error of order n^{-r} in the class

$$\widetilde{\mathcal{F}}_r^1 \;=\; \mathcal{F}_r^1(L_r, \infty, \infty) \;:=\; \{\, f \in F_r^1 \mid \|f^{(r)}\|_{L^\infty} \le L_r \,\}$$

that is larger than $\mathcal{F}_r^1$ and $\hat{\mathcal{F}}_r^1$. Unfortunately, the answer to this question is negative since any algorithm has an infinite worst case error even for the subclass $\mathcal{F}_r^1(0, \infty, \infty)$. To show this, let g_1 and g_2 be two r-smooth functions that will be specified later. Consider the following functions dependent on g_i's and parameterized by $s \in (0, T)$:

$$f_s \;:=\; g_1\, \mathbb{1}_{[0,s)} + g_2\, \mathbb{1}_{[s,T]}.$$

Of course, they belong to F_r^1. Let $\mathcal{A}_n$ be an arbitrary adaptive algorithm that evaluates f at n points $x_j = x_j(f(x_1), \ldots, f(x_{j-1}))$. To simplify the notation, we will write $x_j = x_j(f)$. Since the value of f_s at x depends only on whether $x < s$ or $x \ge s$, we have that

$$\left| \bigcup_{s \in (0,T)} \{x_1, x_2(f_s), \ldots, x_n(f_s)\} \right| \;\le\; 2^n - 1.$$

Therefore, there is an interval $(a, b) \subset (0, T)$ of length $T2^{-n}$ that contains no point $x_j(f_s)$, i.e.,

$$b - a \;=\; T\,2^{-n} \quad \text{and} \quad (a, b) \cap \bigcup_{s \in (0,T)} \{x_1, x_2(f_s), \ldots, x_n(f_t)\} \;=\; \emptyset.$$

This means that the algorithm $\mathcal{A}_n$ cannot distinguish between the functions $f_{a+\delta}$ and $f_{b-\delta}$ (for sufficiently small δ), and its error is bounded from below by $\|f_{a+\delta} - f_{b-\delta}\|_{L^p}/2$ and $\mathrm{dist}_{\rm S}(f_{a+\delta}, f_{b-\delta})/2$, respectively.

Now consider $g_1 \equiv 0$ and $g_2 \equiv D_0$. Clearly, $f_{a+\delta}$ and $f_{b-\delta}$ belong to $\mathcal{F}_r^1(0, 0, D_0)$, which is a very small subset of $\mathcal{F}_r^1(0, \infty, \infty)$. Since

$$\|f_{a+\delta} - f_{b-\delta}\|_{L^p} = D_0(T2^{-n} - 2\delta)^{1/p},$$

we have

$$\sup_{f \in \mathcal{F}_r^1(0,0,D_0)} \|f - \mathcal{A}_n f\|_{L^p} \;\ge\; \frac{D_0}{2}\, \left(T2^{-n}\right)^{1/p} \quad \text{and} \quad \sup_{f \in \mathcal{F}_r^1(0,0,\infty)} \|f - \mathcal{A}_n f\|_{L^p} \;=\; \infty$$

for any algorithm $\mathcal{A}_n$. The result above does not hold for the Skorohod distance since the class $\mathcal{F}_r^1(0, 0, \infty)$ admits an adaptive algorithm with its worst case error equal to $T2^{-n}$. Therefore to get a meaningful lower bound, we use different functions g_i. Namely, we take $g_1(x) = A_1 + Lx$ and $g_2(x) = A_2 + Lx$ for some

constants A_i and L. Note that if $\lambda(a+\delta) \neq b-\delta$, then $\|f_{a+\delta} \circ \lambda - f_{b-\delta}\|_{L^\infty} \geq |A_1 - A_2|$. However, when $\lambda(a+\delta) = b-\delta$, then $\|\lambda - \mathrm{id}\|_{L^\infty} \geq b-a-2\delta$ and $\|f_{a+\delta} \circ \lambda - f_{b-\delta}\|_{L^\infty} \geq |L|(b-a-2\delta)$. This shows that $\mathrm{dist}_{\mathrm{S}}(f_{a+\delta}, f_{b-\delta}) \geq \min(|A_1 - A_2|, (|L|+1)(b-a-2\delta))$. Note that for both functions, $|A_1 - A_2|$ is the discontinuity, L is the norm of the first derivative, and the norm of the rth derivative is zero. Hence we conclude that

$$\sup_{f \in F_r^1(0,L_1,D_0)} \mathrm{dist}_{\mathrm{S}}(f, \mathcal{A}_n f) \geq \frac{1}{2} \min\left(D_0\,,\, (1+L_1)T2^{-n}\right)$$

for any algorithm $\mathcal{A}_n$. In particular, $\sup_{f \in F_r^1(0,\infty,\infty)} \mathrm{dist}_{\mathrm{S}}(f, \mathcal{A}_n f) = \infty$, as claimed.

These lower estimates show that uniform boundeness of only $\|f^{(r)}\|_{L^\infty}$ is not sufficient for the worst case error (measured in the L^p norm or the Skorohod metric) to be finite.

Remark 2. The convergence rate n^{-r} cannot be improved, not only in the worst case, but also in the *asymptotic setting*. This follows from the general results of Trojan [27] that can be rephrased as follows. Let $\{\mathcal{A}_n\}_n$ be an arbitrary sequence of algorithms and $\{\gamma_n\}_n$ be any positive sequence converging to infinity (however slowly). Then the set of $f \in C^r$ for which $\limsup_{n\to\infty} \|f - \mathcal{A}_n f\|_{L^p} n^r/\gamma_n < \infty$ is nowhere dense. Although the results of [27] were proven for errors measured in norms of Banach spaces, due to Proposition 1, the same result holds for the approximation problem with the Skorohod metric.

6. Adaptive algorithms: Multiple discontinuities

In this section, we consider functions with possibly multiple singularities. Unlike the case of one singularity, the results now depend significantly on whether the problem is considered in the worst case setting or in the asymptotic setting.

6.1. **Worst case setting.** A natural generalization of the class $\mathcal{F}_r^1$ to the case of multiple discontinuities is

$$\mathcal{F}_{r,q}^\ell = \mathcal{F}_{r,q}^\ell(L_r, L_1, D_0) := \{\, f \in F_r^\ell \mid \|f^{(r)}\|_{L^\infty} \leq L_r, \|f'\|_{L^\infty} \leq L_1, \|\overline{\Delta}_f^{(0)}\|_q \leq D_0 \,\},$$

where $\overline{\Delta}_f^{(0)} = \left(\Delta_1^{(0)}, \ldots, \Delta_{k_f}^{(0)}\right) \in \mathbb{R}^{k_f}$ is the vector of all discontinuity jumps of f,

$$\|\overline{\Delta}_f^{(0)}\|_q = \left(\sum_{j=1}^{k_f} |\Delta_j^{(0)}|^q\right)^{1/q} \qquad \text{for } 1 \leq q < \infty,$$

and $\|\overline{\Delta}_f^{(0)}\|_\infty = \max_{1\leq j\leq k_f} |\Delta_j^{(0)}|$. Let $\mathcal{A}_n^{\mathrm{non}}$ be the nonadaptive algorithm from Proposition 3. Due to (8), for $\ell < \infty$ the worst case error of $\mathcal{A}_n^{\mathrm{non}}$ over $\mathcal{F}_{r,q}^\ell$ is bounded as

$$\sup_{f \in \mathcal{F}_{r,q}^\ell} \|f - \mathcal{A}_n^{\mathrm{non}} f\|_{L^p} \leq C_\ell \cdot n^{-1/p}.$$

Surprisingly, this error bound cannot be improved by any other (adaptive) algorithm. We also have that for the worst case approximation in the Skorohod metric with respect to the class

$$\hat{\mathcal{F}}_r^\ell = \hat{\mathcal{F}}_r^\ell(L_r, L_1, \infty) := \{\, f \in F_r^\ell \mid \|f^{(r)}\|_{L^\infty} \leq L_r, \|f'\|_{L^\infty} \leq L_1 \,\},$$

the error equals infinity.

Proposition 4. *Let $2 \le \ell \le \infty$ and $1 \le q \le \infty$. For any (adaptive) algorithm $\mathcal{A}_n$ using n function evaluations we have*

$$\sup_{f \in \mathcal{F}^{\ell}_{r,q}} \|f - \mathcal{A}_n f\|_{L^p} \ge D_0 2^{-1/q} \lfloor \ell/2 \rfloor^{1-1/q} \left(\frac{T}{n+1} \right)^{1/p}$$

(with the convention that $\infty^a = \infty$ for $a > 0$, and $\infty^0 = 1$) and

$$\sup_{f \in \hat{\mathcal{F}}^{\ell}_{r}} \operatorname{dist}_{\mathrm{S}}(f, \mathcal{A}_n f) = \infty.$$

Proof. Let $x_1 \le x_2 \le \cdots \le x_n$ be the points at which the function values are computed for $f \equiv 0$. Let $0 \le a < b \le T$ be such that $b - a \ge T/(n+1)$ and $[a, b] \subset [x_k, x_{k+1}]$ for some k (where $x_0 = 0$, $x_{n+1} = T$).

For $2 \le \ell < \infty$, we denote $k := \lfloor \ell/2 \rfloor$ and $\Delta := D_0 (2k)^{-1/q}$, and define

$$f^* := \Delta \sum_{j=1}^{k} \mathbb{1}_{(a+\delta j, b-\delta j)}, \qquad 0 < \delta < \frac{b-a}{2k}.$$

Observe that $f_1 := f^*$ and $f_{-1} := -f^*$ share the same information, and they are both in $\mathcal{F}^{\ell}_{r,q}$. Note also that $f_1(x) = \Delta k$ for any $x \in (a + \delta k, b - \delta k)$. Therefore, for any δ, the worst case L^p error of $\mathcal{A}_n$ is at least

$$\frac{1}{2} \|f_1 - f_{-1}\|_{L^p} \ge \Delta k (b - a - 2\delta k)^{-1/p} \ge D_0 2^{-1/q} \lfloor \ell/2 \rfloor^{1-1/q} \left(\frac{T}{n+1} - 2\delta \left\lfloor \frac{\ell}{2} \right\rfloor \right)^{1/p}.$$

The bounds for $\ell = \infty$ are obtained by taking $\ell \to \infty$ above.

For the result in the Skorohod metric it suffices to note that for arbitrary D_0 the worst case error is at least

$$\frac{1}{2} \operatorname{dist}_{\mathrm{S}}(f_1, 0) = \frac{1}{2} D_0 2^{-1/q} \lfloor \ell/2 \rfloor^{1-1/q}.$$

□

Remark 3. Consider the classes

$$\mathcal{F}^{\ell}_{r,q}(s) := \{ f \in F^{\ell}_{r} \mid \|f^{(r)}\|_{L^\infty} \le L_r, \|\overline{\Delta}^{(j)}_f\|_q \le D_j, 0 \le j \le s-1, \|f^{(s)}\|_{L^\infty} \le L_s \}$$

that are generalizations of $\mathcal{F}^1_r(s)$ from Remark 1. Observe that Proposition 4 also remains valid in $\mathcal{F}^{\ell}_{r,q}(s)$ with $1 \le s \le r$, since the functions constructed in the proof belong to those classes. The situation is slightly different for $s = 0$, where, for any $\mathcal{A}_n$, one can show that

$$\sup_{f \in \mathcal{F}^{\ell}_{r,q}(0)} \|f - \mathcal{A}_n f\|_{L^p} \ge L_0 T^{1/p} \min \left(\left(\frac{\lfloor \ell/2 \rfloor}{n+1} \right)^{1/p}, 1 \right)$$

and

$$\sup_{f \in \hat{\mathcal{F}}^{\ell}_{r,q}(0)} \operatorname{dist}_{\mathrm{S}}(f, \mathcal{A}_n f) \ge \frac{1}{2} L_0.$$

The results of this section can be summarized as follows. In the worst case setting, adaption helps for the classes F^{ℓ}_r only when $\ell = 1$. This is in agreement with the corresponding results for the integration problem; see [21].

Fortunately, adaption regains its superiority in the asymptotic setting discussed below.

6.2. **Asymptotic setting.** In this section, we show that the convergence rate of n^{-r} can be preserved even in the class F_r^∞, if we switch from the worst case setting to the *asymptotic setting*. That is, instead of fixing n and considering the 'worst' functions, we have a sequence of algorithms $\mathcal{A}_n$ and, for a fixed function f, we inspect how fast the error goes to zero when $n \to \infty$. We assume that $f \in F_r^\infty$, i.e., the only *a priori* knowledge used is that f is r-smooth except at a finite (but unknown) number of singular points.

Since the number of singularities is unknown, the DETECT procedure of Section 5.1 needs to be modified to accommodate this situation. For a given m, our modified procedure will work as though there were up to ℓ_m singularities, where ℓ_m converges with m to infinity. Moreover, as we look at the asymptotic behavior, the procedure will not need to sample with increasing resolution in vicinities of 0 and T.

Indeed, for sufficiently large m (depending on f) we have that ℓ_m exceeds the number k_f of singular points of f. Moreover, $h = T/m$ is small enough to separate those singular points, and to separate them from the boundary of $[0, T]$. Then the detection mechanism works for each of the singular points as though it was unique.

The modified detection mechanism is given by the following pseudocode, where the parameters are: the smoothness r, a threshold D, and positive integer sequences $\{\kappa_m\}_{m\geq 1}$ and $\{\ell_m\}_{m\geq 1}$. It returns the set V of subintervals suspected to contain essential singularities. S is an auxiliary set of indices.

```
procedure DETECT;
begin
  input m ≥ (r − 1); h := T/m; S := V := ∅;
  κ := κ_m; ℓ := ℓ_m;
  for i := 0 to m − r do d_i := f[t_i, ..., t_{i+r}];
  while ( |S| < ℓ ) and ( max_i |d_i| > D ) do
  begin
    i* := arg max_i |d_i|; S := S ∪ {i*};
    for i := max(0, i* − r + 1) to min(i* + r − 1, m − r) do d_i := 0;
  end;
  for every i* ∈ S do
  begin
    for j := 1 to κ do
    begin
      for i := 0 to r do e_i := f[t_{i*+i/2^j}, ..., t_{i*+(i+r)/2^j}];
      i* := i* + (arg max_i |e_i|) /2^j
    end;
    v_l := i*h; v_r := (i* + r2^{−κ})h;
    V := V ∪ {[v_l, v_r]}
  end;
  return V
end.
```

The resulting algorithm $\overline{\mathcal{A}}_n^{\mathrm{ad}}$ relies on piecewise polynomial approximation of degree $r-1$, except for the subintervals in V, where it is given by piecewise constant approximation. Of course, there are up to ℓ_m detected subintervals.

Formally, $\overline{\mathcal{A}}_n^{\rm ad}$ is defined as follows. Let

$$P_m := \bigcup_{[v_l,v_r]\in V} (v_j, v_r)$$

and

$$U_m := \{\, v_l, v_r \mid [v_l, v_r] \in V \,\} \cup \{\, t_i \mid 0 \le i \le m, t_i \notin P_m \,\}.$$

Then

$$\begin{aligned}\overline{\mathcal{A}}_n^{\rm ad} f \quad &:= \quad \overline{\mathcal{P}}_r(f; U_m)\mathbb{1}_{[0,T)\setminus P_m} + f(T)\mathbb{1}_{\{T\}} \\ &\quad + \sum_{[v_l,v_r]\in V} f(v_l)\mathbb{1}_{[v_l,(v_l+v_r)/2)} + f(v_r)\mathbb{1}_{[(v_l,v_r)/2,v_r)}.\end{aligned}$$

Observe that in the worst case, the number of function evaluations is now at most

$$n = n(m) = (r+1)m + r\ell_m(\kappa_m + 1) + 1.$$

We have the following theorem, the proof of which follows from our discussion and the proofs of Theorems 3 and 4.

Theorem 5. *Let* $\lim_{m\to\infty} \ell_m = \infty$ *and* $\ell_m\kappa_m = O(m)$. *Then* $\overline{\mathcal{A}}_n^{\rm ad}$ *uses* $n = O(m)$ *function evaluations. Additionally,*

(a) if $\lim_{m\to\infty} \kappa_m - (rp-1)\log_2(m) = \infty$, *then*

$$\lim_{m\to\infty} \|f - \overline{\mathcal{A}}_n^{\rm ad} f\|_{L^p} \cdot m^r \;=\; T^r\, \alpha(r,p)\, \|f^{(r)}\|_{L^p}, \qquad 1 \le p < \infty;$$

(b) if $\lim_{m\to\infty} \kappa_m - (r-1)\log_2(m) = \infty$, *then*

$$\lim_{m\to\infty} {\rm dist}_S(f, \overline{\mathcal{A}}_n^{\rm ad} f) \cdot m^r \;=\; T^r\, \alpha(r,\infty)\, \|f^{(r)}\|_{L^\infty}.$$

7. Test results

We implemented and tested both $\mathcal{A}_n^{\rm ad}$ and $\overline{\mathcal{A}}_n^{\rm ad}$ algorithms for regularity $r = 2, 4$ with errors measured both in the Skorohod metric and the L_2 norm. We report the test results for the following four functions:

$$\begin{aligned} f_1(x) \;&=\; \begin{cases} x(\sin(x) + 0.1), & x \in [0,\pi), \\ (x - 2\pi)(0.1 - \sin(x)), & x \in [\pi, 2\pi], \end{cases} \\ f_2(x) \;&=\; \sin(8x)/16 + e^{-16|x-1|}, \;\; x \in [0,\pi], \\ f_{3,4}(x) \;&=\; e^{-(x-1)^2} + \sum_{j=1}^{5} \Delta_j^{(0)}\, \mathbb{1}_{[0,s_j)}(x), \;\; x \in [0,\pi]. \end{aligned}$$

The functions f_3 and f_4 have five singular points with the same discontinuity jumps given by $\Delta_i^{(0)}$, $[\Delta_1^{(0)}, \Delta_2^{(0)}, \Delta_3^{(0)}, \Delta_4^{(0)}, \Delta_5^{(0)}] = [0.8, -0.14, 0.06, -0.1, 0.2]$ and with singular points given, respectively, by $s_j = j\pi/6$ for f_3, and $[s_1, s_2, s_3, s_4, s_5] = [\pi/6, \pi/6 + 0.03, \pi/2, \pi/2 + 0.07, \pi/2 + 0.073]$ for f_4. Note that the singular points of f_3 are evenly spaced, whereas singular points of f_4 are clustered, $s_1 \sim s_2$ and $s_3 \sim s_4 \sim s_5$.

For interpolations $\mathcal{P}_r(\cdot, a, b)$ we used points equidistributed over (a, b) (including the endpoints). All the tests are performed under the choice of

$$\kappa_m = \lceil 2r * (\log_2(m) - \log_2(T)) \rceil.$$

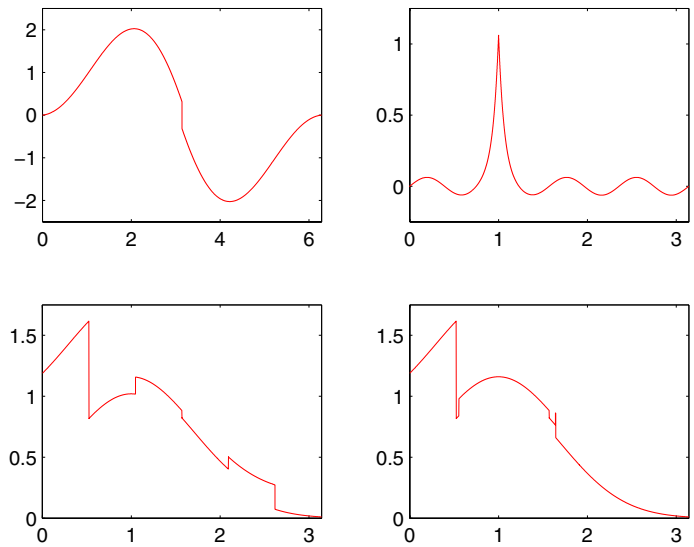

FIGURE 1. Graphs of test functions

When estimating the Skorohod distance, we used the corresponding piecewise linear interpolation as λ. Unless stated otherwise, we set the parameter $D = 0$ to show that the algorithms work well even for this most conservative choice of D.

The tables provide the total number of function evaluations used (denoted by n) as well as the estimated errors (denoted respectively by sko_dist and l2_error).

We begin with the first four tables, where the results for $r = 2$ and $r = 4$ are provided, respectively, in the left and right parts. The function f_1 has only discontinuity at $s_f = \pi$, with the discontinuity jump 0.2π. Table 1 clearly shows that the errors of $\mathcal{A}_n^{\mathrm{ad}}$ are proportional to m^{-r} and that n is proportional to $m(r-1)$.

The function f_2 has only one singular point, $s_f = 1$; however, this time $|\Delta_1^{(0)}| = 0$ and $|\Delta_1^{(1)}| = 32$. Also here, the errors are proportional to m^{-r}, and n is close to $m(r-1)$; see Table 2.

TABLE 1. Errors of $\mathcal{A}_n^{\mathrm{ad}}$ for f_1

m	n	sko_dist	l2_error	n	sko_dist	l2_error
50	129	6.13E-03	7.57E-03	515	6.14E-07	1.57E-05
100	203	1.53E-03	1.89E-03	777	3.83E-08	6.95E-07
150	271	6.80E-04	8.42E-04	997	7.56E-09	1.36E-08
200	327	3.83E-04	4.73E-04	1189	2.39E-09	3.07E-08
300	445	1.70E-04	2.10E-04	1559	4.72E-10	5.18E-10

Since the functions f_3 and f_4 have five singular points, we use them to test $\overline{\mathcal{A}}_n^{\mathrm{ad}}$ with

$$\ell(m) = \frac{m}{2r\log_2(m)}.$$

TABLE 2. Errors of $\mathcal{A}_n^{\rm ad}$ for f_2

m	n	sko_dist	l2_error	n	sko_dist	l2_error
50	159	5.16E-02	1.35E-02	643	2.14E-04	4.15E-05
100	233	1.90E-02	4.16E-03	905	1.91E-05	3.34E-06
150	295	1.07E-02	2.27E-03	1123	4.73E-06	8.34E-07
200	357	5.80E-03	1.16E-03	1317	1.43E-06	2.36E-07
300	469	2.91E-03	5.70E-04	1685	3.16E-07	5.24E-08

In addition to the errors and the number of function evaluations, we provide the number of subintervals chosen by the algorithm in the detection process. This number is marked by n_sin in the second and the sixth column. Since $D = 0$, it is not surprising that the number n_sin of detected subintervals equals $\lfloor \ell(m) \rfloor$.

Tables 3 and 4 show that the errors are quite large unless m is large enough. For f_3 we have the errors not proportional to m^{-r} until $m = 150$ for $r = 2$ and $m = 400$ for $r = 4$. The corresponding values of m are even much higher (in the thousands) for f_4. This is because the singular points of f_4 are clustered and require a very small spacing h in order to be distinguished from each other during the detection stage. Nevertheless, both tables are consistent with our theoretical results on the worst case and asymptotic settings for functions with multiple singularities: the former setting does not work, but the latter does.

TABLE 3. Errors of $\overline{\mathcal{A}}_n^{\rm ad}$ for f_3

m	n_sin	n	sko_dist	l2_error	n_sin	n	sko_dist	l2_error
100	3	227	6.54E-02	1.14E-02	1	471	1.99E-01	1.91E-02
150	5	391	1.10E-04	7.42E-05	2	827	1.25E-01	2.35E-02
200	6	501	6.16E-05	4.07E-05	3	1207	9.93E-02	7.34E-03
250	7	629	3.94E-05	2.57E-05	3	1393	9.65E-02	1.02E-02
300	9	805	2.74E-05	1.82E-05	4	1787	5.35E-02	5.49E-03
400	11	1039	1.54E-05	1.01E-05	5	2369	2.35E-11	1.30E-11

TABLE 4. Errors of $\overline{\mathcal{A}}_n^{\rm ad}$ for f_4

m	n_sin	n	sko_dist	l2_error	n_sin	n	sko_dist	l2_error
100	3	227	8.56E-02	1.49E-02	1	471	1.97E-01	2.11E-02
500	13	1307	7.55E-02	3.46E-03	6	2975	8.32E-02	1.46E-03
1000	25	2751	7.14E-02	3.11E-03	12	6335	9.10E-02	1.31E-03
2000	45	5511	6.17E-07	4.06E-07	22	12819	3.80E-14	2.06E-14
4000	83	11139	1.54E-07	1.01E-07	41	26021	2.89E-15	1.31E-15

In the rest of this section, we discuss how the choice of D and/or $\ell(m)$ affects the efficiency of the algorithm. This is illustrated by the errors of $\overline{\mathcal{A}}_n^{\rm ad}$ for the function f_3; see Tables 5 and 6. We begin by discussing the choice of D.

As mentioned before, a good choice would be $\|f^{(r)}\|_{L^\infty}/r!$ or its decent estimation, which in some cases might be impossible to get. In such cases one could use the most conservative choice $D = 0$ (as we have done above). However, then, too,

TABLE 5. Different D and $\ell(m)$ for f_3 and $r = 2$

m	n_sin	n	sko_dist	n_sin	n	sko_dist
	$D = 0$	$\ell = \ell_1$		$D = 1$	$\ell = \ell_1$	
100	3	227	6.54E-02	3	227	6.54E-02
150	5	391	1.10E-04	5	391	1.10E-04
200	6	501	6.16E-05	5	451	6.17E-04
250	7	629	3.94E-05	5	521	3.95E-05
300	9	805	2.74E-05	5	581	2.74E-05
400	11	1039	1.54E-05	5	691	1.54E-05
	$D = 0$	$\ell = \ell_2$		$D = 1$	$\ell = \ell_2$	
100	15	731	2.46E-04	5	311	2.47E-04
150	20	1111	1.10E-04	5	391	1.10E-04
200	26	1501	6.16E-05	5	451	6.17E-05
250	31	1925	3.94E-05	5	521	3.95E-05
300	36	2317	2.74E-05	5	581	2.74E-05
400	46	3069	1.54E-05	5	691	1.54E-05

many subintervals might be selected in the detection process, resulting in a higher number of function evaluations. This indeed can be seen by comparing the numbers n_sko for $D = 0$ and $D = 1$ (note that $\|f_3^{(r)}\|_{L^\infty}/r! \leq 1$). We report results only for errors measured in the Skorohod distance since they are identical to those with the L_2 norm.

When the value of D is too small (say $D = 0$), the choice of the function $\ell(m)$ can worsen this problem. Indeed, this is illustrated in the following tables, where we used two different choices:[2]

$$\ell_1(m) = \frac{m}{2r\log_2(m)} \quad \text{and} \quad \ell_2(m) = \frac{m}{\log_2(m)}.$$

For instance, in Table 6 we have the following situation. When $m = 400$, the algorithm uses 2369 points when $\ell(m) = \ell_1(m)$ and over 5 times more when $\ell(m) = \ell_2(m)$, even though the errors are the same. However, the choice of $\ell(m) = \ell_1(m)$ is not good for smaller values of m, since $\ell_1(m) < 5$; recall that f_3 has five singular points.

Note that when D is properly chosen ($D = 1$ for f_3), the choice of $\ell(m)$ is not so important as long as $\ell(m)$ is not smaller than the number of singularities, since then the number n_sin does not change and equals 5. However, if $\ell(m) < 5$, then we have the same situation as for $D = 0$: the errors are too large, since n_sin is smaller than 5 and some singular points are not detected/localized.

However, even for poor choices of D and $\ell(m)$, we always have n proportional to m and the error proportional to m^{-r}, the latter holding when m is sufficiently large.

[2]Of course, given the function f, the best choice for $\ell(m)$ would be the actual number of singularities. However, this number is often unavailable.

TABLE 6. Different D and $\ell(m)$ for f_3 and $r = 4$

m	n_sin	n_sko	sko_dist	n_sin	n_sko	sko_dist
	$D = 0$	$\ell = \ell_1$		$D = 1$	$\ell = \ell_1$	
100	1	471	1.99E-01	1	471	1.99E-01
150	2	827	1.25E-01	2	827	1.25E-01
200	3	1207	9.93E-02	3	1207	9.93E-02
250	3	1393	9.65E-02	3	1393	9.65E-02
300	4	1787	5.35E-02	4	1787	5.35E-02
400	5	2369	2.35E-11	5	2369	2.35E-11
	$D = 0$	$\ell = \ell_2$		$D = 1$	$\ell = \ell_2$	
100	15	2849	6.00E-09	5	1149	6.00E-09
150	20	4243	1.19E-09	5	1393	1.19E-09
200	26	5851	3.75E-10	5	1609	3.75E-10
250	31	7383	1.54E-10	5	1819	1.54E-10
300	36	8889	7.41E-11	5	2007	7.42E-11
400	46	11963	2.35E-11	5	2369	2.35E-11

APPENDIX

Proof of Lemma 2. Suppose that f is right-continuous at s_f. (The other case is symmetric.) Then

$$f(x) \;=\; g(x) + \sum_{j=0}^{r-1} \Delta_f^{(j)} \frac{(x - s_f)^j}{j!} \, \mathbb{1}_{[s_f,T]}(x)$$

with $g \in W_r(0,T)$. We also have that, for any function ϕ,

$$\phi[t_i, t_{i+1}, \ldots, t_{i+r}] \;=\; h^{-r} \sum_{l=i}^{i+r} b_{l-i} \phi(t_l).$$

By linearity of divided differences we then have

$$\begin{aligned} f[t_i, \ldots, t_{i+r}] \;&=\; g[t_i, \ldots, t_{i+r}] \\ &\quad + \sum_{j=0}^{r-1} \Delta_f^{(j)} \left(h^{-r} \sum_{l=i}^{i+r} b_{l-i} \frac{(t_l - s_f)^j}{j!} \mathbb{1}_{[s_f,T]}(t_l) \right). \end{aligned}$$

Put $\tau = (s_f - t_k)/h$, $t = l - k$, and $s = i - k + r - 1$. Then

$$\begin{aligned} h^{-r} \sum_{l=i}^{i+r} b_{l-i} \frac{(t_l - s_f)^j}{j!} \mathbb{1}_{[s_f,T]}(t_l) \;&=\; h^{j-r} \sum_{l=k+1}^{i+r} b_{l-i} \frac{(l - k - \tau)^j}{j!} \\ &=\; h^{j-r} \sum_{t=1}^{s+1} b_{t-s+r-1} \frac{(t - \tau)^j}{j!}. \end{aligned}$$

Hence

$$f[t_i, \ldots, t_{i+r}] \;=\; g[t_i, \ldots, t_{i+r}] \;+\; \sum_{j=0}^{r-1} \Delta_f^{(j)} h^{j-r} a_{i+r-k-1,j}(\tau).$$

The rest of the proof goes as the proof of Lemma 3.3 in [21]. □

Proof of Lemma 4. Assume without loss of generality that f is right-continuous at s_f. Then

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

where $g \in W_r(0,T)$. Since polynomial interpolation is linear, we have

$$\begin{aligned} &\|f - \mathcal{P}_r(f;a,b)\|_{L^p(a,b)} \\ &\quad \leq \quad \|g - \mathcal{P}_r(g;a,b)\|_{L^p(a,b)} + \sum_{j=0}^{r-1} |\Delta_f^{(j)}| \cdot \|\phi_{j,s_f} - \mathcal{P}_r(\phi_{j,s_f};a,b)\|_{L^p(a,b)}. \end{aligned}$$

Changing variables we obtain

$$\|\phi_{j,s_f} - \mathcal{P}_r(\phi_{j,s_f};a,b)\|_{L^p(a,b)} \leq \alpha_j \cdot (b-a)^{j+1/p},$$

which, together with (18) and the equality $\|g^{(r)}\|_{L^p(a,b)} = \|f^{(r)}\|_{L^p(a,b)}$, gives the desired result. □

Acknowledgments

We are grateful to Art Werschulz for many valuable suggestions on how to improve the presentation of the results. We also thank Jakob Creutzig and Klaus Ritter for helpful discussions. The first author was supported by the Polish Ministry of Science and Higher Education under Grant 1 P03A 039 28. The second author was partially supported by the National Sciences Foundation under Grant DMS-0608727, and part of this work was done when he visited the University of Warsaw.

References

[1] F. Arandiga, A. Cohen, R. Donat, and N. Dyn, Interpolation and approximation of piecewise smooth functions, *SIAM J. of Numerical Analysis* **43** (2005), pp. 41–57. MR2177955 (2006h:41002)

[2] K. E. Atkinson, *An Introduction to Numerical Analysis,* 2nd ed., Wiley, New York, 1989. MR1007135 (90m:65001)

[3] N. S. Bakhvalov, On the optimality of linear methods for the operator approximation in convex classes of functions (in Russian), *Zh. Vychisl. Mat. Mat. Fiz.* **11** (1971), pp. 1014–1018. MR0290523 (44:7703)

[4] P. Billingsley, *Convergence of Probability Measures,* Wiley, New York, 1968. MR0233396 (38:1718)

[5] E.J. Candès and D.L. Donoho, New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities, *Comm. Pure Appl. Math.* **57** (2004), pp. 219–266. MR2012649 (2004k:42052)

[6] P. J. Davis and P. Rabinowitz, *Methods of Numerical Integration,* Academic Press, New York, 1975. MR0448814 (56:7119)

[7] C. de Boor, CADRE: An algorithm for numerical quadrature, *Mathematical Software*, pp. 201–209, Academic Press, New York, 1971.

[8] I. Daubechies, *Ten lectures on wavelets*, SIAM, Philadelphia, 1992. MR1162107 (93e:42045)

[9] D.L. Donoho, M. Vetterli, R.A. DeVore, and I. Daubechies, Data compression and harmonic analysis, *IEEE Trans. Inform. Theory* **44** (1998), pp. 2435–2476. MR1658775 (99i:94028)

[10] S. Gal and C. A. Micchelli, Optimal sequential and non-sequential procedures for evaluating a functional, *Appl. Anal.* **10** (1980), pp. 105–120. MR575536 (82c:65033)

[11] A. N. Kolmogorov, On Skorohod convergence, *Theor. Prob. Appl.* **1** (1956) *(in Russian).* MR0085638 (19:69i)

[12] A. R. Krommer and C. W. Ueberhuber, *Computational Integration*, SIAM, Philadelphia, 1998. MR1625683 (99g:65027)

[13] J. M. Lyness, Notes on the adaptive Simpson quadrature routine, *J. Assoc. Comput. Mach.* **16** (1969), pp. 483–495. MR0240981 (39:2326)

[14] W. M. McKeeman, Algorithm 145, adaptive numerical integration by Simpson's rule, *Comm. ACM* **5** (1962), pp. 604.

[15] H. N. Mhaskar and J. Prestin, Polynomial frames: a fast tour, in *Approximation Theory XI*, Gatlinburg 2004 (Eds. C.K. Chui, M. Neamtu, L.L. Schumaker), pp. 287–318, Nashboro Press, Brentwood, TN, 2005 MR2126687 (2005k:42087)

[16] E. Novak, *Deterministic and Stochastic Error Bounds in Numerical Analysis,* Springer-Verlag, Berlin, 1988. MR971255 (90a:65004)

[17] E. Novak, On the power of adaption, *J. Complexity* **12** (1996), pp. 199–238. MR1408328 (97g:65008)

[18] K. R. Parthasarathy, *Probability Measures on Metric Spaces,* Academic Press, London, 1967. MR0226684 (37:2271)

[19] R. Piessens, E. deDoncker-Kapenga, C. Überhuber, and D. Kahaner, *QUADPACK: A Subroutine Package for Automatic Integration*, Springer Verlag, New York, 1983. MR712135 (85b:65022)

[20] L. Plaskota, *Noisy Information and Computational Complexity*, Cambridge University Press, Cambridge, 1996. MR1446005 (99b:65189)

[21] L. Plaskota and G. W. Wasilkowski, Adaption allows efficient integration of functions with unknown singularities, *Numerische Mathematik* **102** (2005), pp. 123–144. MR2206675 (2006k:65066)

[22] A. Ralston and P. Rabinowitz, *A First Course in Numerical Analysis,* 2nd ed., McGraw-Hill, New York, 1978. MR0494814 (58:13599)

[23] J. R. Rice, A metalgorithm for adaptive quadrature, *J. Assoc. Comput. Mach.* **22** (1975), pp. 61–82. MR0483327 (58:3340)

[24] K. Ritter, *Average-Case Analysis of Numerical Problems,* Springer-Verlag, Berlin, 2000. MR1763973 (2001i:65001)

[25] K. Sikorski, *Optimal Solution of Nonlinear Equations,* Oxford Press, Oxford, 2000. MR1827804 (2002h:65002)

[26] A. V. Skorohod, Limit theorem for stochastic processes, *Theory Prob. Appl.* **1** (1956), pp. 261–290. MR0084897 (18:943c)

[27] G. M. Trojan, Asymptotic Setting for Linear Problems, unpublished Ph.D. dissertation, 1980. (See also [28, Sect. 2, Chap. 10].)

[28] J. F. Traub, G. W. Wasilkowski and H. Woźniakowski, *Information-Based Complexity*, Academic Press, New York, 1988. MR958691 (90f:68085)

[29] J. F. Traub and A. G. Werschulz, *Complexity and Information,* Cambridge University Press, Cambridge, 1998. MR1692462 (2000m:65170)

[30] J. F. Traub and H. Woźniakowski, *A General Theory of Optimal Algorithms*, Academic Press, New York, 1980. MR584446 (84m:68041)

[31] G. W. Wasilkowski, Information of varying cardinality, *J. Complexity* **1** (1986), pp. 107–117. MR922813 (88m:65099)

[32] G. Wahba, *Spline Models for Observational Data*, SIAM, Philadelphia, 1990. MR1045442 (91g:62028)

[33] A. G. Werschulz, *The Computational Complexity of Differential and Integral Equations,* Oxford University Press, Oxford, 1991. MR1144521 (93a:68061)

Faculty of Mathematics, Informatics, and Mechanics, University of Warsaw, Banacha 2, 02-097 Warsaw, Poland

E-mail address: `leszekp@mimuw.edu.pl`

Department of Computer Science, University of Kentucky, 773 Anderson Hall, Lexington, Kentucky 40506-0046

E-mail address: `greg@cs.uky.edu`

Department of Computer Science, University of Kentucky, 773 Anderson Hall, Lexington, Kentucky 40506-0046

E-mail address: `yaxi@uky.edu`