On measuring the efficiency of kernel operators in $L_p\left(\mathbb{R}^d\right)$

S. Dekel and D. Leviatan

Abstract: It is well known that it is possible to enhance the approximation properties of an kernel operator by increasing its support size. There is an obvious tradeoff between higher approximation order of a kernel and the time complexity of algorithms that employ it. A question is then asked: how do we compare the efficiency of kernels with comparable support size? We follow Blu and Unser and choose as a measure of the efficiency of the kernels the first leading constant in a certain error expansion. We use time domain methods to treat the case of globally supported kernels in $L_p\left(\mathbb{R}^d\right)$, $1 \le p \le \infty$.

AMS Subject clas sifications: 41A35, 41A44, 41A15.

Key words: approximation order, quasi-interpolation, shift-invariant spaces, kernel operators.

Authors' affiliation:

School of Mathematical sciences
Tel-Aviv University
Tel-Aviv 69978, Israel

E-mail:

shai.dekel@turboimage.com

leviatan@math.tau.ac.il

1 Introduction

In this work we perform a "fine" analysis of the approximation properties of integral operators $Q_h : L_p\left(\mathbb{R}^d\right) \to L_p\left(\mathbb{R}^d\right)$, $1 \le p \le \infty$, $h > 0$ defined by a **kernel** K

$$Q_h(f) := \int_{\mathbb{R}^d} f(t)\, K_h(t,\cdot)\, dt, \qquad K_h(t,x) := h^{-d} K\left(h^{-1}t, h^{-1}x\right). \tag{1.1}$$

The approximation error is defined by

$$\left\| f - Q_h(f) \right\|_{L_p\left(\mathbb{R}^d\right)}.$$

Efficient kernels have sufficient decay properties and reproduce polynomials of degree $m-1$, for some $m \ge 1$

$$p(x) = \int_{\mathbb{R}^d} p(t) K(t,x)\, dt, \quad \forall p \in \Pi_{m-1}\left(\mathbb{R}^d\right).$$

Such kernels provide **approximation order** m. Namely, for f in the Sobolev space $W_p^m\left(\mathbb{R}^d\right)$ (see Section 2 for the multi-index notation we use below)

$$\left\| f - Q_h(f) \right\|_{L_p\left(\mathbb{R}^d\right)} \le C h^m \left| f \right|_{m,p}, \qquad \left| f \right|_{m,p} := \sum_{|g|=m} \left\| D^g f \right\|_{L_p\left(\mathbb{R}^d\right)}.$$

An important class of kernel operators is that of **quasi-interpolation** kernels. These kernels are defined by means of two sets of functions, the **generating set** $\Phi = \left\{ f_a \right\}_{a \in \Lambda}$ and the **dual generating set** $\tilde{\Phi} = \left\{ \tilde{f}_a \right\}_{a \in \Lambda}$ in the form

$$K(t,x) := \sum_{k \in \mathbb{Z}^d} \sum_{a \in \Lambda} \tilde{f}_a(t-k) f_a(x-k). \tag{1.2}$$

For example, let $f \in L_2\left(\mathbb{R}^d\right)$ be stable. Then its **natural dual** $\tilde{f}$, defined by its Fourier transform

$$\hat{\tilde{f}} = \frac{\hat{f}}{\sum_{k \in \mathbb{Z}^d} \left| \hat{f}(\cdot + 2p k) \right|^2}, \tag{1.3}$$

provides the orthogonal projection P_h into the shift invariant space $\overline{span}\left\{ f\left(h^{-1} \cdot - k\right) \ : \ k \in \mathbb{Z}^d \right\}$ by

$$P_h f = Q_h f = h^{-d} \sum_{k \in \mathbb{Z}^d} \left\langle f, \tilde{f}\left(h^{-1} \cdot - k\right) \right\rangle f\left(h^{-1} \cdot - k\right).$$

It is known [R] that any univariate generator f that provides approximation order m can be represented as a convolution of the B-spline of order m and a tempered distribution. Therefore, m is also the minimal possible support for a univariate generator that provides approximation order m. At the other end of the scale there is the infinitely supported sinc function given by

$$\operatorname{sinc}(x) := \prod_{n=1}^{d} \frac{\sin(p\, x_n)}{p\, x_n}.$$

The sinc function provides infinite approximation order. That is, the kernel (1.2) with the choice $f = \tilde{f} = \operatorname{sinc}$ has the remarkable property that for any $m \ge 1$ and $f \in W_2^m(\mathbb{R}^d)$

$$\| f - Q_h(f) \|_{L_2(\mathbb{R}^d)} \le h^m |f|_{m,2}.$$

For results on approximation from shift-invariant spaces we refer the reader to [LJC] for the general case $1 \le p \le \infty$ and to [JP] for a survey on harmonic analysis techniques in the case $p = 2$.

In this work we follow Unser et al. ([U], [BTU], [BU]) and attempt to measure the efficiency of kernels by the leading coefficient in a certain error expansion. The purpose of this work is to extend their approach to the multivariate case and to the range $1 \le p \le \infty$.

We impose the following conditions on the kernel K. First we assume the kernel is shift-invariant in the sense that

$$K(t-k, x-k) = K(t,x), \qquad \forall k \in \mathbb{Z}^d. \tag{1.4}$$

This condition clearly holds for quasi-interpolation kernels of type (1.2). The second assumption is that the kernel has sufficient decay, namely, that for some $n \ge 1$, $C > 0$ and $e > 0$

$$|K(t,x)| \le C\left(1 + |t-x|\right)^{-(n+d+e)}. \tag{1.5}$$

For example, (1.5) holds whenever $\Phi, \tilde{\Phi}$ in (1.2) are finite sets of generators with sufficient decay. We note that in general the natural dual (1.3) of a compactly supported $f \in L_2(\mathbb{R}^d)$ is of infinite support. However, in applications it is common practice to use kernels that have a compactly supported band, especially in the multivariate case where $d \ge 2$.

Property (1.5) also ensures that the operators Q_h, $h > 0$, are bounded operators in $L_p(\mathbb{R}^d)$, $1 \le p \le \infty$. This is proved by first using (1.5) to show that they are bounded operators in $L_1(\mathbb{R}^d)$ and $L_\infty(\mathbb{R}^d)$, and then applying an interpolation argument for the case $1 < p < \infty$ (see e.g. Theorem 13.7.3 in [DL] and the details that follow).

Finally, we assume that the kernel reproduces polynomials of degree $m-1$ for some $m \ge 1$, i.e.,

$$x^g = \int_{\mathbb{R}^d} t^g K(t,x)\, dt, \qquad \forall g,\ |g| \le m-1. \tag{1.6}$$

It is known (see e.g. [LJC]) that a kernel that satisfies the above conditions for some $n = m \ge 1$ in (1.5) yields approximation order m. Our main result augments previous results as follows.

Theorem 1.1 Let $m \ge 1$. Assume that a kernel K satisfies (1.4), (1.5) for $n = m+1$ and (1.6). Let $\{Q_h\}_{h \in \mathbb{R}_+}$ be the operators (1.1) and let $1 \le p \le \infty$. Then for any function $f \in W_p^{m+d}(\mathbb{R}^d)$

$$\| f - Q_h(f) \|_{L_p(\mathbb{R}^d)} \le C_{p,K}^{-} h^m |f|_{m,p} + C \sum_{n=1}^{d} h^{m+n} |f|_{m+n,p}, \tag{1.7}$$

where

$$C^{-}_{p,K} := \max_{g\in\mathbb{Z}^d_+,\ |g|=m} \frac{\left\| e_{g,K} \right\|_{L_p\left([0,1]^d\right)}}{g\,!}, \tag{1.8}$$

and

$$e_{g,K}(x) := \int_{\mathbb{R}^d} (t-x)^g\, K(t,x)\, dt, \qquad g \in \mathbb{Z}^d_+. \tag{1.9}$$

Using (1.6) it can be shown that in the univariate case the constant $C^{-}_{p,K}$ in (1.8) is given by

$$C^{-}_{p,K} = \frac{\left\| x^m - \int_{\mathbb{R}} t^m K(t,x)\, dt \right\|_{L_p([0,1])}}{m!}.$$

It is not surprising to see that we obtain a Chebyshev-type result, namely, the leading constant $C^{-}_{p,K}$ is determined by how well the kernel, which reproduces polynomials of degree $m-1$, approximates the monomial x^m. Obviously, if K reproduces polynomials of a degree higher than $m-1$, then as expected, $C^{-}_{p,K} = 0$. The constant $C^{-}_{p,K}$ can serve as a "fine" measure of the approximation properties of kernels that provide approximation order m. Since in applications the support size of the kernel determines the complexity of the algorithms, we should strive for kernels with the highest possible approximation order for a given support size and the smallest possible leading constant $C^{-}_{p,K}$. Indeed, Blu, Thevenaz and Unser construct in [BTU] univariate generators (O-moms) that are asymptotically optimal in the following sense. Of all the generators that provide approximation order m and have the minimal support size m, these generators have the smallest constant $C^{-}_{2,K}$. For instance, for $m = 4, 6$, these "optimal" generators are

$$OM_4 = N_4 + \frac{1}{42} N_4^{(2)}, \qquad OM_6 = N_6 + \frac{1}{33} N_6^{(2)} + \frac{1}{7920} N_6^{(4)},$$

where N_m is the univariate B-spline of order m. They also demonstrate the advantage of the kernel operators defined by these generators in image processing applications. However, in these constructions the kernels are not compactly supported.

Below is a table with examples for known kernels K and the values of their corresponding leading constants $C^{-}_{p,K}$ for several values of p.

m	Kernel	$\left\|\mathrm{supp}(\tilde{f})\right\| \times \left\|\mathrm{supp}(f)\right\|$	$m!C^{-}_{1,K}$	$m!C^{-}_{2,K}$	$m!C^{-}_{10,K}$	$m!C^{-}_{100,K}$	$m!C^{-}_{\infty,K}$
1	Daub1, N_1	1×1	0.25	0.2888	0.4289	0.9138	1.0
2	Daub2	3×3	0.1933	0.2236	0.3502	0.4717	0.5
	Coif1	5×5	0.1760	0.2125	0.3920	0.5868	0.6376
	$\left(\widetilde{N_2}, N_2\right)$	$\infty\times2$		0.07454			
3	Daub3	4×4	0.2582	0.2988	0.4362	0.5672	0.5974
	$\left(\widetilde{N_3}, N_3\right)$	$\infty\times3$		0.03450			
	(1,3), [2,6]	2×6	0.2813	0.2989	0.3453	0.3847	0.3951
4	Daub4	6×6	0.4854	0.5557	0.7141	0.8342	0.8658
	Coif2	11×11	0.4332	0.4953	0.6671	0.8196	0.8560

	$\left(\widetilde{N_4}, N_4\right)$	$\infty\times 4$		0.02182			
	$\left(\widetilde{OM_4}, OM_4\right)$	$\infty\times 4$		0.004763			
	$(2,4)$	2×8	1.3756	1.6360	2.4745	3.2882	3.5031
	$(4,4)$, [9,7]	9×7	0.3063	0.3474	0.4569	0.5532	0.5760
	$(4,4)$, [7,9]	7×9	0.7662	0.8929	1.2751	1.6277	1.7169
5	Daub5	8×8	1.1997	1.3161	1.6086	1.8429	1.9085
	$\left(\widetilde{N_5}, N_5\right)$	$\infty\times 5$		0.01734			
	$(1,5)$	1×9	1.1718	1.2719	1.5075	1.6771	1.7192
	(3,5)	3×11	11.4494	11.7480	12.8894	14.5878	15.3649
6	Daub6	10×10	3.4417	3.7788	4.5029	4.9661	5.1092
	Coif3	17×17	2.8730	3.2314	4.1054	4.7317	4.9022
	$\left(\widetilde{N_6}, N_6\right)$	$\infty\times 6$		0.01655			
	$\left(\widetilde{OM_6}, OM_6\right)$	$\infty\times 6$		0.0003			
7	Daub7	12×12	11.5241	12.7441	15.6353	17.6203	18.1699
	(3,7)	3×15	100.6437	106.7492	122.4140	136.1411	139.5817

Table 1-1 Values of $C^{-}_{p,K}$ for known univariate kernels K.

Remarks

1. All of the entries in Table 1-1 correspond to univariate kernels of type (1.2)
2. The kernels Daub$m$ are the Daubechies orthogonal generators with $f=\tilde{f}$ (see [D]).
3. The kernels Coif n with $2n=m$ are the Coiflets orthogonal generators with $f=\tilde{f}$ (see [D]).
4. The kernels $(\tilde{m}, m)$ are the CDF biorthogonal generators taken from [D] and are identified (as in [D]) by the approximation orders of $\left(\tilde{f}, f\right)$.
5. The kernel $\left(\widetilde{N_m}, N_m\right)$ is the B-spline of order m and its (globally supported) natural dual.
6. The kernels $\left(\widetilde{OM_m}, OM_m\right)$ are the "optimal" generators for $p=2$ and their (globally supported) natural dual constructed in [BTU].
7. The notation $[l1, l2]$ is used in the signal processing community to represent the length of the corresponding filters. These are also the support sizes of $\left(\tilde{f}, f\right)$.
8. Some of the entries for the case $p=2$ are given in [U].

Our measure for the efficiency of kernels seems to correlate well with empirical results in signal processing. For example, the popular $[9,7]$ filters that serve as the default filters in the image compression standard JPEG2000, correspond to a kernel with a relatively small leading constant compared to other compactly supported kernels that provide approximation order 4. Observe that the $[7,9]$ filter obtained by switching the roles of f and $\tilde{f}$ in (1.2), has the same support size and provides the same approximation order but with a bigger leading constant. Indeed, it does not perform as well as the $[9,7]$ in image compression.

We also observe that the relative efficiency of a kernel changes for different values of p. For example for $p=1$ the kernel Daub3 has a smaller constant than $(1,3)$ while for $p=\infty$, the opposite is true.

2 Preliminaries

We recall some basic definitions of multivariate polynomials, differentials and Taylor series. For a multi-index $a\in\mathbb{Z}_+^d$ we write $|a|:=\sum_{k=1}^d a_k$. The factorial of a multi-index is given by $a!=\prod_{k=1}^d a_k!$. For $a\in\mathbb{Z}_+^d$, the function $x^a:=\prod_{k=1}^d x_k^{a_k}$ is a **multivariate monomial** of (total) **degree** $|a|$. The monomials of degree $\le m$ are the building blocks of the **multivariate polynomials**

$$\Pi_m\left(\mathbb{R}^d\right):=\left\{p(x)=\sum_{|a|\le m} c_a x^a\right\}.$$

The n **th order differential** of a sufficiently smooth function f at a point $x\in\mathbb{R}^d$ can be represented for our purpose in a simple form as an operator $\mathbb{D}^n:\mathbb{R}^d\to\mathbb{R}$ defined by

$$\mathbb{D}^n(f,x)\cdot v:=n!\sum_{|a|=n}\frac{D^a f(x)}{a!}v^a\ ,\qquad v\in\mathbb{R}^d,$$

where

$$D^a f:=\frac{\partial^n f}{\partial x_1^{a_1}\cdots\partial x_d^{a_d}}.$$

We find it convenient to also use the general form of the nth order differential as an operator in the space $L^n\left(\mathbb{R}^d,\mathbb{R}\right)$ (see e.g. [AMR] pp. 76-93). This is the space of all real valued multilinear operators defined inductively by $L^n\left(\mathbb{R}^d,\mathbb{R}\right):=L\left(\mathbb{R}^d,L^{n-1}\left(\mathbb{R}^d,\mathbb{R}\right)\right)$, where $L(X,Y)$ is the space of linear operators from X to Y. The space $L^n\left(\mathbb{R}^d,\mathbb{R}\right)$ is a finite dimensional Banach space of dimension $n\times d$ equipped with the norm

$$\|A\|_{L^n\left(\mathbb{R}^d,\mathbb{R}\right)}=\sup_{e_1,\dots,e_n\neq 0}\frac{|A(e_1,\dots,e_n)|}{\|e_1\|\cdots\|e_n\|}.$$

Denoting $\Lambda:=\left\{a\in\mathbb{Z}_+^d \mid |a|=n\right\}$ we have for $1\le p\le\infty$ the following norm equivalence

$$\left\|\mathbb{D}^n f(x)\right\|_{L^n\left(\mathbb{R}^d,\mathbb{R}\right)}\sim\left\|D^a f(x)\right\|_{l_p(\Lambda)},\tag{2.1}$$

where the equivalence constants depend on d, n and p but are independent of the point x. If for $A \in L^n\left(\mathbb{R}^d, \mathbb{R}\right)$ and $v \in \mathbb{R}^d$ we denote

$$A \cdot v^n := A\underbrace{\left(v, \dots, v\right)}_{\text{n times}},$$

then the **Taylor polynomial** of degree $n-1$ of a sufficiently smooth function f about the point $x \in \mathbb{R}^d$ is given by

$$T_{n-1}\left(x,t\right) = T_{n-1}\left(f, x\right)\left(t\right) := \sum_{k=0}^{n-1} \frac{\mathbb{D}^k f\left(x\right)}{k!} \cdot \left(t-x\right)^k .$$

The **Taylor remainder** of order n of a sufficiently smooth function f at a point $x \in \mathbb{R}^d$ is given by

$$R_n\left(x,t\right) = R_n\left(f, x\right)\left(t\right) := \int_0^1 \frac{\left(1-u\right)^{n-1}}{\left(n-1\right)!} \mathbb{D}^n\left(f, x+u\left(t-x\right)\right) du \cdot \left(t-x\right)^n, \tag{2.2}$$

and we have that

$$f\left(t\right) = T_{n-1}\left(x,t\right) + R_n\left(x,t\right).$$

3 Proof of the main result

This section is devoted to the proof of Theorem 1.1. The following lemma is required in cases where we wish to estimate a discrete sum of samples of a function in the p norm using the (integral) function norm.

Lemma 3.1 For any $f \in C^d\left(\mathbb{R}^d\right) \cap W_p^d\left(\mathbb{R}^d\right)$, $1 \le p < \infty$ and $h > 0$ we have the following 'numerical integration' inequality

$$\left\| h^d \sum_{k \in \mathbb{Z}^d} \left| f\left(x + hk\right)\right|^p \right\|_{L_\infty\left(\mathbb{R}^d\right)}^{1/p} \le \left\| f \right\|_p + C \sum_{n=1}^{d} h^n \left| f \right|_{n,p} . \tag{3.1}$$

Proof Without loss of generality we can assume that $x = 0$, else we take the function $f\left(x - \cdot\right)$. Define the following step function

$$f_h\left(t\right) := \sum_{k \in \mathbb{Z}^d} f\left(hk\right) \mathbf{1}_{S\left(hk,h\right)}\left(t\right),$$

where for $y = \left(y_1, \dots, y_d\right)$, $S\left(y,h\right)$ is the box $\left[y_1, y_1 + h\right) \times \cdots \times \left[y_d, y_d + h\right)$. Then,

$$\begin{aligned} \sum_{k \in \mathbb{Z}^d} h^d \left| f\left(hk\right)\right|^p &= \sum_{k \in \mathbb{Z}^d} \int_{S\left(hk,h\right)} \left| f\left(hk\right)\right|^p dt \\ &\le \sum_{k \in \mathbb{Z}^d} \int_{S\left(hk,h\right)} \left(\left| f\left(hk + x\right)\right| + \left| f\left(hk + x\right) - f\left(hk\right)\right| \right)^p dx \end{aligned}$$

$$= \left\| |f| + |f - f_h| \right\|_p^p$$
$$\le \left(\|f\|_p + \|f - f_h\|_p \right)^p .$$

Therefore, it is sufficient to prove that $\|f - f_h\|_p \le C \sum_{n=1}^{d} h^d |f|_{n,p}$. This last inequality states that numerical integration using interpolation and a step size h provides 1st order accuracy, which is well known for the univariate case. Now, write

$$f(t) = \sum_{k \in \mathbb{Z}^d} f(t) \mathbf{1}_{S(h,k,h)}(t) ,$$

and we have

$$f(t) - f_h(t) = \sum_{k \in \mathbb{Z}^d} \left(f(t) - f(hk) \right) \mathbf{1}_{S(h,k,h)}(t) .$$

Hence

$$\|f - f_h\|_p^p = \sum_{k \in \mathbb{Z}^d} \int_{S(hk,h)} \left| f(t) - f(hk) \right|^p dt . \tag{3.2}$$

We first demonstrate the proof for the case $d = 2$. In this case, we need to estimate

$$\int_{hk_1}^{h(k_1+1)} \int_{hk_2}^{h(k_2+1)} \left| f(t_1, t_2) - f(hk_1, hk_2) \right|^p dt_1 dt_2 ,$$

or if we denote $g(u_1, u_2) = f(u_1 + hk_1, u_2 + hk_2)$, then a change of variables yields that we need to estimate

$$\int_0^h \int_0^h \left| g(u_1,u_2) - g(0,0) \right|^p du_1 du_2 \le 2^{p-1} \left(\int_0^h \int_0^h \left| g(u_1,u_2) - g(u_1,0) \right|^p du_1 du_2 + \int_0^h \int_0^h \left| g(u_1,0) - g(0,0) \right|^p du_1 du_2 \right)$$
$$=: 2^{p-1} \left(I_1 + I_2 \right).$$

We estimate the first term by

$$I_1 \le \int_0^h \int_0^h \left(\int_0^{u_2} \left| \frac{\partial g}{\partial u_2}(u_1, v) \right| dv \right)^p du_1 du_2$$
$$\le h \int_0^h \left(\int_0^h \left| \frac{\partial g}{\partial u_2}(u_1, v) \right| dv \right)^p du_1 .$$

Hölder's inequality yields

$$I_1 \le h h^{p-1} \int_0^h \int_0^h \left| \frac{\partial g}{\partial u_2}(u_1, v) \right|^p dv du_1$$
$$= h^p \left\| \frac{\partial f}{\partial t_2} \right\|_{L_p(S(hk,h))}^p .$$

As for the second term

$$
\begin{aligned}
I_2 &\le \int_0^h\int_0^h \left(\int_0^{u_1}\left|\frac{\partial g}{\partial u_1}(v,0)\right|dv\right)^p du_1 du_2 \\
&\le 2^{p-1}h\int_0^h\left(\int_0^h\left|\frac{\partial g}{\partial u_1}(v,u_2)\right|dv\right)^p du_2 + 2^{p-1}h\int_0^h\left(\int_0^h\int_0^{u_2}\left|\frac{\partial^2 g}{\partial u_2\partial u_1}(v_1,v_2)\right|dv_2 dv_1\right)^p du_2 \\
&\le 2^{p-1}h^p\int_0^h\int_0^h\left|\frac{\partial g}{\partial v_1}(v_1,v_2)\right|^p dv_1 dv_2 + 2^{p-1}h^{2p}\int_0^h\int_0^h\left|\frac{\partial^2 g}{\partial u_2\partial u_1}(v_1,v_2)\right|^p dv_1 dv_2 \\
&\le 2^{p-1}h^p\left\|\frac{\partial f}{\partial t_1}\right\|^p_{L_p(S(hk,h))} + 2^{p-1}h^{2p}\left\|\frac{\partial^2 f}{\partial t_1\partial t_2}\right\|^p_{L_p(S(hk,h))}.
\end{aligned}
$$

Substituting into (3.2) we obtain

$$
\| f - f_h\|_p \le C\left(h\left(\left\|\frac{\partial f}{\partial t_1}\right\|_p + \left\|\frac{\partial f}{\partial t_2}\right\|_p\right) + h^2\left\|\frac{\partial^2 f}{\partial t_1\partial t_2}\right\|\right).
$$

This completes the proof for $d=2$. The case $d>2$ follows in similar manner, but requires an induction process. We denote $g(u) := f(u+hk)$ for any (fixed) $k\in\mathbb{Z}^d$ and estimate

$$
\begin{aligned}
&\int_0^h\cdots\int_0^h\left|g(u_1,\ldots,u_d)-g(0,\ldots,0)\right|^p du_1\cdots du_d \\
&\qquad\le C(p,d)\sum_{i=1}^d\int_0^h\cdots\int_0^h\left|g(0,\ldots,0,u_i,\ldots,u_d)-g(0,\ldots,0,u_{i+1},\ldots,u_d)\right|^p du_1\cdots du_d .
\end{aligned}
$$

We will prove by induction that

$$
\begin{aligned}
&\int_0^h\cdots\int_0^h\left|g(0,\ldots,0,u_i,\ldots,u_d)-g(0,\ldots,0,u_{i+1},\ldots,u_d)\right|^p du_1\cdots du_d \\
&\qquad\le C(p,d)\sum_{j=1}^i h^{jp}\sum_{1\le k_1<\cdots<k_j\le i}\int_0^h\cdots\int_0^h\left|\frac{\partial^j g}{\partial u_{k_1}\cdots\partial u_{k_j}}g(u_1,\ldots,u_d)\right|^p du_1\cdots du_d .
\end{aligned}
\tag{3.3}
$$

For $i=1$

$$
\int_0^h\cdots\int_0^h\left|g(u_1,\ldots,u_d)-g(0,u_2,\ldots,u_d)\right|^p du_1\cdots du_d \le h\int_0^h\cdots\int_0^h\left(\int_o^h\left|\frac{\partial g}{\partial u_1}(v,u_2,\ldots,u_d)\right|dv\right)^p du_2\cdots du_d
$$

by Hölder's inequality

$$
\le h^p\int_0^h\cdots\int_0^h\left|\frac{\partial g}{\partial u_1}(u_1,u_2,\ldots,u_d)\right|^p du_1\cdots du_d ,
$$

and this completes the proof for $i=1$. Assume (3.3) holds for all $1\le m<i$. Then as above

$$\int_0^h\cdots\int_0^h\left|g\left(0,\ldots,0,u_i,\ldots,u_d\right)-g\left(0,\ldots,0,u_{i+1},\ldots,u_d\right)\right|^p du_1\cdots du_d$$

$$\le h^p\int_0^h\cdots\int_0^h\left|\frac{\partial g}{\partial u_i}\left(0,\ldots,0,u_i,\ldots,u_d\right)\right|^p du_1\cdots du_d =: I_0^{(i)}(i)$$

$$\le 2^{p-1}h^p\int_0^h\cdots\int_0^h\left|\frac{\partial g}{\partial u_i}\left(0,\ldots,0,u_i,\ldots,u_d\right)-\frac{\partial g}{\partial u_i}\left(0,\ldots,u_{i-1},u_i,\ldots,u_d\right)\right|^p du_1\cdots du_d$$
$$+2^{p-1}h^p\int_0^h\cdots\int_0^h\left|\frac{\partial g}{\partial u_i}\left(0,\ldots,u_{i-1},u_i,\ldots,u_d\right)\right|^p du_1\cdots du_d =: I_1^{(i)}(i-1)+I_2^{(i)}(i-1).$$

Note that $I_2^{(i)}(i-1)=2^{p-1}I_0^{(i)}(i-1)$. Hence

$$I_0^{(i)}(i)\le I_1^{(i)}(i-1)+2^{p-1}I_1^{(i)}(i-1)+\cdots+2^{(p-1)(i-2)}I_1^{(i)}(1)+2^{(p-1)(i-2)}I_2^{(i)}(1)\,. \tag{3.4}$$

Now by the induction assumption for $1\le m<i$

$$I_1^{(i)}(m)\le C(p,d)h^p\sum_{j=1}^m h^{jp}\sum_{1\le k_1<\cdots<k_j\le m}\int_0^h\cdots\int_0^h\left|\frac{\partial g}{\partial u_{k_1}\cdots\partial u_{k_j}}\frac{\partial g}{\partial u_i}\left(u_1,\ldots,u_d\right)\right|^p du_1\cdots du_d\,. \tag{3.5}$$

Also

$$I_2^{(i)}(1)=2^{p-1}h^p\int_0^h\cdots\int_0^h\left|\frac{\partial g}{\partial u_i}\left(u_1,\ldots,u_d\right)\right|^p du_1\cdots du_d\,. \tag{3.6}$$

Combining (3.4) with (3.5) for $m=1,\ldots,i-1$ and (3.6), we complete the induction.

♦

Lemma 3.2 Assume K satisfies (1.5) for some $n\ge 1$, $C,e>0$. Then for all $f\in C^n\left(\mathbb{R}^d\right)\cap W_p^n\left(\mathbb{R}^d\right)$, $1\le p\le\infty$

$$\left\|\int_{\mathbb{R}^d}K_h(t,x)R_n(f,x)(t)\,dt\right\|_p\le Ch^n|f|_{n,p}\,. \tag{3.7}$$

Proof Observe that is sufficient to show that for all $f\in C^n\left(\mathbb{R}^d\right)\cap W_p^n\left(\mathbb{R}^d\right)$

$$\left\|\int_{\mathbb{R}^d}K(t,x)R_n(x,t)\,dt\right\|_p\le C|f|_{n,p}\,,$$

since (3.7) follows by the change of variables

$$\left\|\int_{\mathbb{R}^d}K_h(t,x)R_n(x,t)\,dt\right\|_p=\left\|\int_{\mathbb{R}^d}K(t,x)R_n\left(f(h\cdot),x\right)(t)\,dt\right\|_p\,.$$

First, we bound the Taylor remainder using (2.2)

$$\left|R_n(x,t)\right| = \left|\left(\int_0^1 \frac{(1-u)^{n-1}}{(n-1)!}\mathbb{D}^n f\left(x+(t-x)u\right)du\right)\cdot(t-x)^n\right|$$
$$\le |t-x|^n \int_0^1 \left\|\mathbb{D}^n f\left(x+(t-x)u\right)\right\|_{L^n\left(\mathbb{R}^d,\mathbb{R}\right)} du.$$

Let $p=\infty$ and fix $x\in\mathbb{R}^d$. Then (1.5) and (2.1) yield

$$\left|\int_{\mathbb{R}^d} K(t,x)R_n(x,t)\,dt\right| \le \int_0^1\int_{\mathbb{R}^d}|K(t,x)||t-x|^n \left\|\mathbb{D}^n f\left(x+u\ (t-x)\right)\right\|_{L^n\left(\mathbb{R}^d,\mathbb{R}\right)} dtdu$$
$$\le \operatorname*{ess\,sup}_{y\in\mathbb{R}^d}\left\|\mathbb{D}^n f(y)\right\|_{L^n\left(\mathbb{R}^d,\mathbb{R}\right)}\int_{\mathbb{R}^d}|K(t,x)||t-x|^n\,dt$$
$$\le C|f|_{n,\infty}\int_{\mathbb{R}^d}\left(1+|t-x|\right)^{-(d+e)}dt$$
$$\le C|f|_{n,\infty}.$$

For $1\le p<\infty$ we apply $(1.5), (2.1)$ and twice the Minkowski inequality to obtain

$$\left\|\int_{\mathbb{R}^d} K(t,x)R_n(x,t)\,dt\right\|_p \le \left(\int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d}|K(t,x)||t-x|^n\int_0^1\left\|\mathbb{D}^n f\left(x+u(\ t-x)\right)\right\|_{L^n\left(\mathbb{R}^d,\mathbb{R}\right)} dudt\right)^p dx\right)^{1/p}$$
$$\le \int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d}|K(t,x)|^p|t-x|^{np}\left(\int_0^1\left\|\mathbb{D}^n f\left(x+u(\ t-x)\right)\right\|_{L^n\left(\mathbb{R}^d,\mathbb{R}\right)} du\right)^p dx\right)^{1/p} dt$$
$$\le \int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d}\left(1+|y|\right)^{-(d+e)p}\left(\int_0^1\left\|\mathbb{D}^n f\left(x+uy\right)\right\|_{L^n\left(\mathbb{R}^d,\mathbb{R}\right)} du\right)^p dx\right)^{1/p} dy$$
$$\le \int_{\mathbb{R}^d}\left(1+|y|\right)^{-(d+e)}\left(\int_0^1\left(\int_{\mathbb{R}^d}\left\|\mathbb{D}^n f\left(x+uy\right)\right\|^p_{L^n\left(\mathbb{R}^d,\mathbb{R}\right)} dx\right)^{1/p} du\right)dy$$
$$\le C|f|_{n,p}\int_{\mathbb{R}^d}\left(1+|y|\right)^{-(d+e)}dy$$
$$\le C|f|_{n,p}.$$

♦

Proof of Theorem 1.1 First assume that $f\in C^{m+d}\left(\mathbb{R}^d\right)\cap W_p^{m+d}\left(\mathbb{R}^d\right)$**.** We generally follow the method that was used in [U] for the case $d=1, p=2$. For a fixed $x\in\mathbb{R}^d$ we have by (1.6)

$$f(x)-Q_h(f,x) = \int_{\mathbb{R}^d} R_m(x,t)K_h(t,x)dt$$
$$= \int_{\mathbb{R}^d}\left(\frac{\mathbb{D}^m f(x)\cdot(t-x)^m}{m!}+R_{m+1}(x,t)\right)K_h(t,x)dt$$
$$= h^m\sum_{|g|=m}\frac{D^g\ f(x)}{g\,!}\int_{\mathbb{R}^d}\left(t-h^{-1}x\right)^g K\left(t,h^{-1}x\right)dt+\int_{\mathbb{R}^d}R_{m+1}(x,t)\,K_h(t,x)dt.$$

Using the notation (1.9) we obtain a bound with two terms

$$\left\| f - Q_h(f) \right\|_p \le h^m \left\| \sum_{|g|=m} \frac{D^g f}{g!} e_{g,K}\left(h^{-1}\cdot\right) \right\|_p + \left\| \int_{\mathbb{R}^d} R_{m+1}(\cdot,t) K_h(t,\cdot) dt \right\|_p . \qquad (3.8)$$

First, we bound the second term in (3.8). Since we assumed that the kernel K satisfies (1.5) for $n = m+1$, by Lemma 3.2

$$\left\| \int_{\mathbb{R}^d} R_{m+1}(\cdot,t) K_h(t,\cdot) dt \right\|_p \le C h^{m+1} |f|_{m+1,p} .$$

We now assume that $1 \le p < \infty$, since for $p = \infty$ the proof follows almost immediately from the above arguments. We proceed with the estimate of the first term in (3.8). It is easy to verify that property (1.4) implies that for each $g \in \mathbb{Z}_+^d$ the function $e_g(x)$ is 1-periodic. Therefore, for each $g \in \mathbb{Z}_+^d$, $|g| = m$

$$\begin{aligned}
\left\| D^g f(x) e_{g,K}\left(h^{-1}x\right) \right\|_p^p &= \sum_{k\in\mathbb{Z}^d} \int_{S(hk,h)} \left| D^g f(x) \right|^p \left| e_{g,K}\left(h^{-1}x\right) \right|^p dx \\
&= \int_{S(0,h)} \left| e_{g,K}\left(h^{-1}x\right) \right|^p \sum_{k\in\mathbb{Z}^d} \left| D^g f(x+hk) \right|^p dx \\
&= \int_{S(0,1)} \left| e_{g,K}(y) \right|^p h^d \sum_{k\in\mathbb{Z}^d} \left| D^g f(hy+hk) \right|^p dy \\
&\le \left\| e_{g,K} \right\|_{L_p\left([0,1]^d\right)}^p \left\| h^d \sum_{k\in\mathbb{Z}^d} \left| D^g f(\cdot + kh) \right|^p \right\|_{L_\infty\left(\mathbb{R}^d\right)} .
\end{aligned}$$

Since $f \in C^{m+d}\left(\mathbb{R}^d\right) \cap W_p^{m+d}\left(\mathbb{R}^d\right)$, the partial derivative $D^g f$ is in $C^d\left(\mathbb{R}^d\right) \cap W_p^d\left(\mathbb{R}^d\right)$ for each $g \in \mathbb{Z}_+^d$, $|g| = m$ and thus we may apply (3.1) to get

$$\left\| h^d \sum_k \left| D^g f(\cdot + kh) \right|^p \right\|_{L_\infty(\mathbb{R})}^{1/p} \le \left\| D^g f \right\|_p + C \sum_{n=1}^d h^n |f|_{m+n,p} .$$

Thus we can combine our estimates so far in the following manner

$$\begin{aligned}
\left\| f(x) - Q_h(f,x) \right\|_p &\le h^m \sum_{|g|=m} \frac{\left\| e_{g,K} \right\|_{L_p\left([0,1]^d\right)}}{g!} \left(\left\| D^g f \right\|_p + C \sum_{n=1}^d h^n |f|_{m+n,p} \right) + C h^{m+1} |f|_{m+1,p} \\
&\le \max_{|g|=m} \frac{\left\| e_{g,K} \right\|_{L_p\left([0,1]^d\right)}}{g!} h^m |f|_{m,p} + C \sum_{n=1}^d h^{m+n} |f|_{m+n,p} .
\end{aligned}$$

This proves (1.7) for $f \in C^{m+d}\left(\mathbb{R}^d\right) \cap W_p^{m+d}\left(\mathbb{R}^d\right)$. To complete the proof for arbitrary functions in $W_p^{m+d}\left(\mathbb{R}^d\right)$ we use a standard ‘regularization’ argument. Let $f \in W_p^{m+d}\left(\mathbb{R}^d\right)$, then by Lemma 2.1.3 in [Z] there exists a sequence $f_j \in C^\infty\left(\mathbb{R}^d\right) \cap W_p^{m+d}\left(\mathbb{R}^d\right)$, such that $\left\| f - f_j \right\|_{W_p^{m+d}\left(\mathbb{R}^d\right)} \underset{j\to\infty}{\longrightarrow} 0$. Since Q_h is a bounded operator in $L_p\left(\mathbb{R}^d\right)$ we have

$$\begin{aligned}
\left\| f - Q_h f \right\|_p &\le \left\| f - f_j \right\|_p + \left\| f_j - Q_h f_j \right\|_p + \left\| Q_h f - Q_h f \right\|_p \\
&\le \left(1 + \left\| Q_h \right\|\right) \left\| f - f_j \right\|_p + C_{K,p}^{-} h^m \left| f_j \right|_{m,p} + C \sum_{n=1}^{d} h^{m+n} \left| f_j \right|_{m+n,p} \\
&\underset{j\to\infty}{\longrightarrow} C_{p,K}^{-} h^m \left| f \right|_{m,p} + C \sum_{n=1}^{d} h^{m+n} \left| f \right|_{m+n,p} .
\end{aligned}$$

♦

Acknowledgements: We are grateful to the referee for pointing out a mistake in the proof of Lemma 3.1 and we thank Sasha Sherman for his help in fixing it.

References

[AMR] R. Abraham, J. E. Marsden and T. Ratiu, Manifolds, tensor analysis and applications, Springer-Verlag, 1988.

[BTU] T. Blu, P. Thévenaz and M. Unser, MOMS: Maximal-Order Interpolation of Minimal Support, IEEE Trans. on Image Proc. 10 (2001), 1069-1080.

[BU] T. Blu and M. Unser, Approximation error for quasi-interpolators and (multi-) wavelet expansions, App. Comp. Harmonic Anal. 6 (1999), 219-251.

[D] I. Daubechies, Ten lectures on wavelets, CBMS-NSF Reg. Conf. Series in Applied Math., Vol. 61, SIAM, 1992.

[DL] R. DeVore and G. Lorentz, Constructive Approximation, Springer-Verlag, 1991.

[JP] K. Jetter and G. Plonka, "A survey on L_2 – aproximation order from shift invariant spaces", in Multivariate approximation and applications (eds. N. Dyn, D. Leviatan, D. Levin and A. Pinkus), Cambridge Univ. Press, 2001.

[LJC] J. Lei, R. Q. Jia and W. Cheney, Approximation from shift-invariant spaces by integral operators, SIAM J. Math. Anal. 28 (1997), 481-498.

[R] A. Ron, Factorization theorems for univariate splines on regular grids, Israel J. Math. 70 (1990), 48-68.

[U] M. Unser, Approximation power of biorthogonal wavelet expansions, Trans. Signal Proc. 44 (1996), 519-527.

[Z] W. Ziemer, Weakly differentiable functions, Springer-Verlag, 1989.