APPROXIMATION CLASSES FOR ADAPTIVE TIME-STEPPING FINITE ELEMENT METHODS

MARCELO ACTIS†, PEDRO MORIN†, AND CORNELIA SCHNEIDER*

ABSTRACT. We study approximation classes for adaptive time-stepping finite element methods for time-dependent Partial Differential Equations (PDE). We measure the approximation error in $L_2([0,T)\times\Omega)$ and consider the approximation with discontinuous finite elements in time and continuous finite elements in space, of any degree. As a byproduct we define Besov spaces for vector-valued functions on an interval and derive some embeddings, as well as Jackson- and Whitney-type estimates.

Contents

1. Introduction and main result

Adaptive time-stepping finite element methods (AFEM) for evolutionary PDE usually lead to a sequence of timesteps and meshes, which yield a partition of the

Date: March 11, 2021.

2010 *Mathematics Subject Classification.* Primary 41A25, 65D05; Secondary 65N30, 65N50.

†Partially supported by Agencia Nacional de Promoción Científica y Tecnológica, through grants PICT-2014-2522, PICT-2016-1983, by CONICET through PIP 2015 11220150100661, and by Universidad Nacional del Litoral through grants CAI+D 2016-50420150100022LI. A research stay of Pedro Morin at Universität Erlangen was partially supported by the Simons Foundation and by the Mathematisches Forschungsinstitut Oberwolfach.

* *(Corresponding author)* Supported by Deutsche Forschungsgemeinschaft (DFG), grant SCHN 1509/1-2.

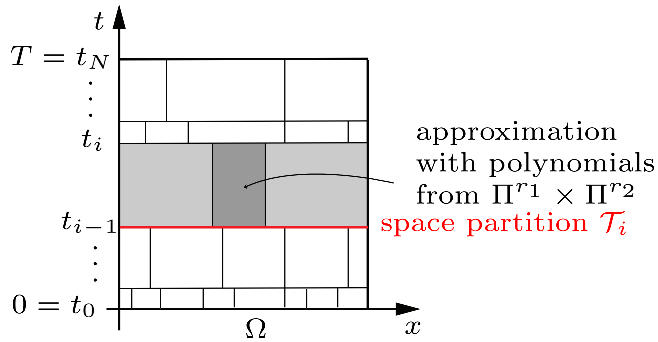


FIGURE 1. Time-space partition $\mathcal{P}$

time interval $0 = t_0 < t_1 < \cdots < t_N = T$ and one triangulation $\mathcal{T}_i$ for each time interval $[t_{i-1}, t_i)$. The complexity of the discrete solution is thus related to the total number of degrees of freedom needed to represent it on the whole interval, which in turn is equivalent to $\sum_{i=1}^N \#\mathcal{T}_i$.

In this article we study spaces of functions which can be approximated using such time-space partitions with an error of order $\left(\sum_{i=1}^N \#\mathcal{T}_i\right)^{-s}$ for different $s > 0$. The results that we obtain are similar in spirit to those of [BDDP02, GM14], where the spaces corresponding to stationary PDE are considered.

Our goal is not to prove the optimality of AFEM but rather to understand which convergence rates are to be expected for the solutions of evolutionary PDE given their regularity. In this paper we aim at establishing the first results in this direction, thus at some points we sacrifice generality in order to have a clearer presentation of the basic ideas and set the foundation for further research in this area.

In order to roughly state our main result, we need to introduce some notation, which will be explained in detail later.

Given a polyhedral space domain $\Omega \subset \mathbb{R}^n$, $n \geq 1$, we let $\mathbb{T}$ denote the set of all triangulations that are obtained through bisection from an initial triangulation $\mathcal{T}_0$ of Ω. For each $\mathcal{T} \in \mathbb{T}$ we denote by $\#\mathcal{T}$ the number of elements of the partition

For $\mathcal{T} \in \mathbb{T}$, we let $\mathbb{V}_\mathcal{T}^r$ denote the finite element space of continuous piecewise polynomial functions of fixed order r, i.e.,

$$\mathbb{V}_\mathcal{T}^r := \{g \in C(\overline{\Omega}) : \; g\big|_T \in \Pi^r \text{ for all } T \in \mathcal{T}\},$$

where Π^r denotes the set of polynomials of total degree (strictly) less than r.

Let $r_1, r_2 \in \mathbb{N}$ denote the polynomial orders in time and space, respectively. Let $\{0 = t_0 < t_1 < \ldots < t_N = T\}$ be a partition of the time interval and $\mathcal{T}_1, \ldots, \mathcal{T}_N \in \mathbb{T}$ be partitions of the space domain Ω, where $\mathcal{T}_i$ corresponds to the subinterval $[t_{i-1}, t_i)$, $i = 1, \ldots N$. The time-space partition as illustrated in Figure 1 is then given by

$$\mathcal{P} = (\{0 = t_0 < t_1 < \ldots < t_N = T\}, \{\mathcal{T}_1, \ldots, \mathcal{T}_N\}) \quad \text{with} \quad \#\mathcal{P} = \sum_{i=1}^N \#\mathcal{T}_i$$

and $\mathbb{P}$ is the set of all those time-space partitions. This is the precise kind of time-space partitions produced by time-stepping adaptive methods.

The finite element space $\overline{\mathbb{V}}_{\mathcal{P}}^{r_1,r_2}$ subject to such a partition $\mathcal{P}$ is defined as

$$\overline{\mathbb{V}}_{\mathcal{P}}^{r_1,r_2} := \{g : [0,T)\times\Omega \to \mathbb{R} : g_{[t_{i-1},t_i)\times\Omega} \in \Pi^{r_1}\otimes \mathbb{V}_{\mathcal{T}_i}^{r_2},\ \text{for all } i=1,2,\dots,N\},$$

i.e., $g \in \overline{\mathbb{V}}_{\mathcal{P}}^{r_1,r_2}$ if and only if $g(t,\cdot)\in\mathbb{V}_{\mathcal{T}_i}^{r_2}$ for all $t\in[t_{i-1},t_i)$ and $g(\cdot,x)\big|_{[t_{i-1},t_i)}\in\Pi^{r_1}$ for all $x\in\Omega$, and all $i=1,2,\dots,N$. Discrete solutions of adaptive time-stepping methods, e.g. those which use Discontinuous Galerkin (DG) in time, belong to spaces of this type.

We define the best m-term approximation error by

$$\overline{\sigma}_m(f) = \inf_{\#\mathcal{P}\le m}\ \inf_{g\in\overline{\mathbb{V}}_{\mathcal{P}}^{r_1,r_2}} \|f-g\|_{L_2([0,T)\times\Omega)}.$$

In this article we measure the error in $L_2([0,T)\times\Omega)$ and leave the general case of $L_p([0,T),L_q(\Omega))$ and other generalizations as future work.

For $s>0$ we define the approximation class $\overline{\mathbb{A}}_s$ as the set those functions whose best m-term approximation error is of order m^{-s}, i.e.,

$$\overline{\mathbb{A}}_s := \{f\in L_2([0,T)\times\Omega):\ \exists c>0 \text{ such that } \overline{\sigma}_m(f)\le c\,m^{-s},\ \forall m\in\mathbb{N}\}.$$

Equivalently, we can define $\overline{\mathbb{A}}_s$ through a semi-norm as follows:

$$\overline{\mathbb{A}}_s := \{f\in L_2([0,T)\times\Omega):\ |f|_{\overline{\mathbb{A}}_s}<\infty\} \quad\text{with}\quad |f|_{\overline{\mathbb{A}}_s} := \sup_{m\in\mathbb{N}} m^s\,\overline{\sigma}_m(f).$$

Alternatively, this definition is equivalent to saying that $f\in\overline{\mathbb{A}}_s$ if there is a constant c such that for all $\varepsilon>0$, there exists a time-space partition $\mathcal{P}$ that satisfies

$$\inf_{g\in\overline{\mathbb{V}}_{\mathcal{P}}^{r_1,r_2}} \|f-g\|_{L_2([0,T)\times\Omega)} \le c\varepsilon \quad\text{and}\quad \#\mathcal{P}\le\varepsilon^{-1/s}, \tag{1}$$

and $|f|_{\overline{\mathbb{A}}_s}$ is equivalent to the infimum of all constants c that satisfy (1).

Our main result is stated in terms of Besov spaces, which will be defined in the next section, and reads as follows.

Main Result 1. *Let $0<s_i<r_i$, $i=1,2$, $0<q_1\le\infty$, $1\le q_2\le\infty$ with $s_1>\big(\frac{1}{q_1}-\frac{1}{2}\big)_+$ and $s_2>n\big(\frac{1}{q_2}-\frac{1}{2}\big)_+$. Then*

$$B^{s_1}_{q_1,q_1}([0,T),L_2(\Omega))\cap L_2([0,T),B^{s_2}_{q_2,q_2}(\Omega))\subset\overline{\mathbb{A}}_s \quad \textit{for}\quad s=\frac{1}{\frac{1}{s_1}+\frac{n}{s_2}}.$$

This result is a consequence of Theorem 25, where, given $f\in B^{s_1}_{q_1,q_1}([0,T),L_2(\Omega)\cap L_2([0,T),B^{s_2}_{q_2,q_2}(\Omega))$, and $\varepsilon>0$ we construct a time-space partition $\mathcal{P}$ that satisfies

$$\#\mathcal{P}\le c_1\varepsilon^{-\left(\frac{1}{s_1}+\frac{n}{s_2}\right)}$$

and a function $F\in\overline{\mathbb{V}}_{\mathcal{P}}^{r_1,r_2}$ such that

$$\|f-F\|_{L_2([0,T)\times\Omega)}\le c_2\,\varepsilon\left[|f|_{B^{s_1}_{q_1,q_1}([0,T),L_2(\Omega))}+\|f\|_{L_2([0,T),B^{s_2}_{q_2,q_2}(\Omega))}\right].$$

Here $B^s_{p,q}(I,X)$ denote Besov spaces of X-valued functions with respective semi-norms $|\cdot|_{B^s_{p,q}(I,X)}$, cf. Section 2.2. It is worth noting that in order to determine the largest spaces, integrability powers $0<p<1$ must be considered. This makes some proofs more complicated than if we were to consider only $p\ge1$.

Our construction is performed in two steps. The first one uses a Greedy algorithm to obtain the partition of the time domain, resorting in a Whitney-type estimate for vector-valued functions. That is, we interpret functions in $L_2([0,T)\times\Omega)$ as functions from $[0,T)$ into $L_2(\Omega)$ as is customary in the study of evolutionary PDE, and develop a nonlinear approximation theory for this situation, by revisiting and extending some results from Storozhenko and Oswald [Sto77, OS78]. This is presented in Section 3, after defining Besov spaces of vector-valued functions in Section 2. In Section 4 we revisit the known results for the stationary case and perform the aforementioned first step by applying the Greedy algorithm to vector-valued functions. In Section 5 we combine those two results and prove our main result. We end this article presenting some discussion and comparison of the approximation classes for space-time discretizations.

We finally mention that we will use $A \lesssim B$ inside some statements, proofs and reasonings in order to denote $A \leq cB$ with a constant c that depends on the parameters indicated in the corresponding statement. As usual, $A \simeq B$ means $A \lesssim B$ and $B \lesssim A$.

2. Besov spaces of vector-valued functions

The goal of this section is to define and understand some properties of Besov spaces of functions from a real interval I into a Banach space. From now on, we let X be a separable Banach space with norm $\|\cdot\|_X$.

We first introduce the moduli of smoothness and state and prove some of their properties, which are analogous to those corresponding to the case of real-valued functions. Afterwards we define the corresponding Besov spaces and state and prove some embeddings.

2.1. Moduli of smoothness of vector-valued functions on an interval.

We start this section by providing new definitions of moduli of smoothness for vector-valued functions, which are analogous to the ones already known for real-valued functions, and stating and proving some of their basic properties.

It is worth mentioning that there is a forerunner regarding moduli of smoothness and Whitney-type estimates of vector-valued functions, cf. [DF90]. However, our definition (which is an immediate generalization of the classical moduli of real-valued functions) differs from the one given in [DF90] (which is more elaborate and tricky). In particular, in [DF90] a duality approach is used between the given Banach space and its dual in order to reduce the definitions and results for abstract functions to real ones. But there is a price to pay: the results are restricted to the set of bounded functions. Therefore even classical Banach spaces like $L_p(I, X)$ cannot be considered entirely.

Given $0 < p \leq \infty$, a real interval $I = [a, b)$ with $|I| = b - a$, and a function $f : I \to X$, we say that $f \in L_p(I, X)$ if f is measurable and $\|f\|_{L_p(I,X)} := \Big(\int_I \|f(t)\|_X^p \mathrm{d}t\Big)^{1/p} < \infty$ if $p < \infty$ and $\|f\|_{L_\infty(I,X)} = \operatorname{esssup}_{t\in I} \|f(t)\|_X$. For such a function f, $r \in \mathbb{N}$ and $0 < |h| < \frac{|I|}{r}$, the r-th order difference $\Delta_h^r f : I_{rh} \to X$ is defined as

$$\Delta_h^r f(t) = \sum_{i=0}^{r} \binom{r}{i} (-1)^{r-i} f(t + ih), \qquad t \in I_{rh} := \{t \in I : t + rh \in I\},$$

which clearly satisfies $\Delta_h^r f = \Delta_h \Delta_h^{r-1} f$ and $\|\Delta_h^r f\|_{L_p(I_{rh})}^{\min\{1,p\}} \leq 2\|\Delta_h^{r-1} f\|_{L_p(I_{(r-1)h})}^{\min\{1,p\}}$, understanding that $\Delta_h f = \Delta_h^1 f$ and $\Delta_h^0 f = f$.

The modulus of smoothness is defined as

$$\omega_r(f,I,u)_p := \sup_{0<|h|\le u} \|\Delta_h^r f\|_{L_p(I_{rh},X)} = \sup_{0<h\le u} \|\Delta_h^r f\|_{L_p(I_{rh},X)}, \qquad u>0, \tag{2}$$

which is clearly increasing as a function of u, and the *averaged* modulus of smoothness is defined, for $u>0$, as

$$w_r(f,I,u)_p := \left(\frac{1}{2u}\int_{-u}^{u} \|\Delta_h^r f\|_{L_p(I_{rh},X)}^p \,\mathrm{d}h\right)^{\frac{1}{p}} = \left(\frac{1}{u}\int_{0}^{u} \|\Delta_h^r f\|_{L_p(I_{rh},X)}^p \,\mathrm{d}h\right)^{\frac{1}{p}}. \tag{3}$$

The well-known definitions for $f:\Omega\to\mathbb{R}$, with Ω a domain of $\mathbb{R}^n$, $n\ge 1$, are as follows. For $h\in\mathbb{R}^n$, the domain of $\Delta_h^r f$ is the set $\Omega_{rh} := \{x\in\Omega : x, x+h,\dots,x+rh\in\Omega\}$, and the moduli of smoothness $\omega_r(f,\Omega,u)_p$, $w_r(t,\Omega,u)_p$ are defined for $u>0$ via

$$\omega_r(f,\Omega,u)_p := \sup_{0<|h|\le u} \|\Delta_h^r f\|_{L_p(\Omega_{rh})}, \tag{4}$$

$$w_r(f,\Omega,u)_p := \left(\frac{1}{(2u)^n}\int_{[-u,u]^n} \|\Delta_h^r f\|_{L_p(\Omega_{rh})}^p \,\mathrm{d}h\right)^{\frac{1}{p}}.$$

As a consequence of the fact that $\Delta_{mh}^1 f(x) = \sum_{i=0}^{m-1} \Delta_h^1 f(x+ih)$, for $m\in\mathbb{N}$, we can prove by induction $\|\Delta_{mh}^r f\|_{L_p(A_{rmh})} \le m^r \|\Delta_h^r f\|_{L_p(A_{rh})}$, for $A=I$ or $A=\Omega$ (for details see [PP87, Sect. 3.1]). As an immediate consequence of this,

$$\omega_r(f,A,mu)_p^{\min\{1,p\}} \le m^r \omega_r(f,A,u)_p^{\min\{1,p\}}, \quad u>0. \tag{5}$$

From the properties stated above, we have

$$w_{r+1}(f,A,u)_p^{\min\{1,p\}} \le 2 w_r(f,A,u)_p^{\min\{1,p\}}. \tag{6}$$

Finally, we notice that if $f:[a,b)\to X$ and $\hat f:[0,1)\to X$ with $\hat f(t) = f(a+t(b-a))$, then, for $u>0$

$$\begin{aligned} \omega_r(f,[a,b),u)_p &= (b-a)^{1/p}\,\omega_r(\hat f,[0,1),(b-a)^{-1}u)_p,\\ w_r(f,[a,b),u)_p &= (b-a)^{1/p}\,w_r(\hat f,[0,1),(b-a)^{-1}u)_p. \end{aligned} \tag{7}$$

Now we prove that the two moduli of smoothness w_r and ω_r as defined above in (2) and (3) are equivalent. This result is well-known and proved for real-valued functions in [DL93, Lem. 6.5.1]. The proof for vector-valued functions is analogous and we sketch it here for completeness.

Lemma 1. *Given $0<p<\infty$ and $r\in\mathbb{N}$ the two definitions of moduli of smoothness $w_r(\cdot,\cdot,\cdot)_p$ and $\omega_r(\cdot,\cdot,\cdot)_p$ are equivalent, more precisely*

$$w_r(f,I,u)_p \le \omega_r(f,I,u)_p \le c\,w_r(f,I,u)_p,$$

for all $f\in L_p(I,X)$, $I=[a,b)$ and $0<u<|I|/r$, where the constant c depends only on r and p, but is otherwise independent of f, I, and u.

Proof. The fact that $w_r(f,I,u)_p \le \omega_r(f,I,u)_p$ is obvious. Therefore, it remains to prove the converse inequality. We prove the result for the reference situation of $I=[0,1)$, the general case follows by scaling using (7).

We use the reproducing formula

$$\Delta_h^r f(t) = \sum_{l=1}^{r} (-1)^l \binom{r}{l} \left[\Delta_{ls}^r f(t+lh) - \Delta_{h+ls}^r f(t)\right], \tag{8}$$

which holds if $t \in [0, 1 - rh]$ and

$$t + lh + rls \leq 1 \quad \text{and} \quad t + rh + rls \leq 1.$$

This together yields the range $t \in [0, 1 - rh - r^2 s]$. Formula (8) is proved by induction, starting with the observation that

$$\begin{aligned} \Delta_h^1 f(t) &= f(t+h) - f(t) \\ &= f(t+h) - f(t+h+ls) + f(t+h+ls) - f(t) \\ &= -\big[\Delta_{ls}^1 f(t+h) - \Delta_{h+ls}^1 f(t)\big]. \end{aligned}$$

We now consider $0 < h \leq u \leq \frac{1}{4r}$ and $0 \leq t \leq \frac{1}{2}$. This gives us the upper bound $s < \frac{1}{4r^2}$. Integrating formula (8) yields

$$\int_0^{1/2} \|\Delta_h^r f(t)\|_X^p \mathrm{d}t \lesssim \sum_{l=1}^r \int_0^{1/2} \|\Delta_{ls}^r f(t+lh)\|_X^p \mathrm{d}t + \int_0^{1/2} \|\Delta_{h+ls}^r f(t)\|_X^p \mathrm{d}t.$$

Thus, setting $I_- := [0, 1/2]$ and averaging over $s \in [0, u]$ gives

$$\begin{aligned} &\|\Delta_h^r f\|_{L_p(I_-,X)}^p \qquad (9) \\ &\quad \lesssim \sum_{l=1}^r \frac{1}{u} \left[\int_0^u \int_{I_-} \|\Delta_{ls}^r f(t+lh)\|_X^p \mathrm{d}t\mathrm{d}s + \int_0^u \int_{I_-} \|\Delta_{h+ls}^r f(t)\|_X^p \mathrm{d}t\mathrm{d}s \right] \\ &\quad = \sum_{l=1}^r \frac{1}{lu} \left[\int_0^{lu} \int_{I_-} \|\Delta_{h'}^r f(t+lh)\|_X^p \mathrm{d}t\mathrm{d}h' + \int_h^{h+lu} \int_{I_-} \|\Delta_{h'}^r f(t)\|_X^p \mathrm{d}t\mathrm{d}h' \right] \\ &\quad \lesssim \sum_{l=1}^r \frac{1}{(r+1)u} \int_0^{(r+1)u} \|\Delta_{h'}^r f\|_{L_p(I,X)}^p \mathrm{d}h' \\ &\quad \leq w_r(f, I, (r+1)u)_p^p, \qquad (10) \end{aligned}$$

where in the second step we used the substitution $h' := ls$ in the first and $h' := h + ls$ in the second integral. By symmetry, we also have that $\|\Delta_{-h}^r f\|_{L_p(I_+,X)}^p \leq w_r(f, I, (r+1)u)_p^p$ with $I_+ = [1/2, 1]$. Taking the supremum w.r.t. $0 < h \leq u$ on both sides we arrive at

$$\omega_r(f, I, u)_p \lesssim w_r(f, I, (r+1)u)_p$$

Using (5) we obtain

$$\omega_r(f, I, (r+1)u)_p \lesssim \omega_r(f, I, u)_p \lesssim w_r(f, I, (r+1)u)_p,$$

which completes the proof. □

2.2. **Besov spaces and embeddings.** Using the generalized modulus of smoothness defined in the previous subsection, we introduce the Besov spaces $B_{p,q}^s(I, X)$, $s > 0$, $0 < p, q \leq \infty$, which contain all functions $f \in L_p(I, X)$ such that for $r := \lfloor s \rfloor + 1$ the quasi-seminorm

$$\begin{aligned} |f|_{B_{p,q}^s(I,X)} &:= \left(\int_0^{|I|/r} \left[u^{-s} \omega_r(f, I, u)_p\right]^q \frac{\mathrm{d}u}{u} \right)^{1/q} < \infty, \qquad 0 < q < \infty, \\ |f|_{B_{p,\infty}^s(I,X)} &:= \sup_{0<u<|I|/r} u^{-s} \omega_r(f, I, u)_p < \infty. \end{aligned} \qquad (11)$$

Moreover, a quasi-norm for $B^s_{p,q}(I,X)$ is given by

$$\|f\|_{B^s_{p,q}(I,X)} := \|f\|_{L_p(I,X)} + |f|_{B^s_{p,q}(I,X)}, \tag{12}$$

which is a norm whenever $1 \le p, q \le \infty$.

Remark 2. One can replace the integral $\int_0^{|I|/r}$ by $\int_0^1$ if $|I| < \infty$ and still get an equivalent norm. More precisely,

$$\int_0^{|I|/r} [u^{-s}\omega_r(f,I,u)_p]^q \frac{\mathrm{d}u}{u} \simeq \int_0^1 [u^{-s}\omega_r(f,I,u)_p]^q \frac{\mathrm{d}u}{u}$$

with equivalence constants that depend only on s, r, p, q, but are otherwise independent of f and $|I|$ as $|I| \to 0$.

We prove this claim for $0 < q < \infty$, the case $q = \infty$ is analogous. If $|I|/r < 1$ then, on the one hand, $\int_0^{|I|/r}[u^{-s}\omega_r(f,I,u)_p]^q \frac{\mathrm{d}u}{u} \le \int_0^1[u^{-s}\omega_r(f,I,u)_p]^q \frac{\mathrm{d}u}{u}$. On the other hand, $\omega_r(f,I,u)_p = \omega_r(f,I,|I|/r)_p$, when $u \ge |I|/r$. Therefore, using (5) and the monotonicity of $\omega_r(f,I,\cdot)_p$,

$$\begin{aligned}
\int_{|I|/r}^1 [u^{-s}\omega_r(f,I,u)_p]^q \frac{\mathrm{d}u}{u} &= \omega_r(f,I,|I|/r)_p^q \int_{|I|/r}^1 u^{-sq-1}\mathrm{d}u \\
&\lesssim \omega_r(f,I,|I|/(2r))_p^q \,(|I|/r)^{-sq} \\
&\lesssim \omega_r(f,I,|I|/(2r))_p^q \int_{|I|/(2r)}^{|I|/r} u^{-sq-1}\mathrm{d}u \\
&\le \int_{|I|/(2r)}^{|I|/r} [u^{-s}\omega_r(f,I,u)_p]^q \frac{\mathrm{d}u}{u},
\end{aligned}$$

which yields the second inequality for the case $|I|/r < 1$.

If $|I|/r > 1$, trivially $\int_0^1[u^{-s}\omega_r(f,I,u)_p]^q \frac{\mathrm{d}u}{u} \le \int_0^{|I|/r}[u^{-s}\omega_r(f,I,u)_p]^q \frac{\mathrm{d}u}{u}$. Besides, using again (5) and the monotonicity of $\omega_r(f,I,\cdot)_p$,

$$\omega_r\left(f,I,\frac{1}{2}\right)_p \le \omega_r(f,I,u)_p \le \omega_r\left(f,I,\frac{|I|}{r}\right)_p \lesssim |I|^r\omega_r\left(f,I,\frac{1}{2}\right)_p, \quad \frac{1}{2} \le u \le |I|/r.$$

Hence,

$$\int_1^{|I|/r} [u^{-s}\omega_r(f,I,u)_p]^q \frac{\mathrm{d}u}{u} \lesssim |I|^{rq}\omega_r\left(f,I,\frac{1}{2}\right)_p^q \lesssim |I|^{rq}\int_{\frac{1}{2}}^1 [u^{-s}\omega_r(f,I,u)_p]^q \frac{\mathrm{d}u}{u},$$

which proves the claim for the case $|I|/r > 1$.

Remark 3. Our definition for the Besov spaces above is in good agreement with the standard case: When $f : \Omega \to \mathbb{R}$, with Ω a domain of $\mathbb{R}^n$, the usual Besov spaces $B^s_{p,q}(\Omega)$ are defined as those subspaces containing all functions $f \in L_p(\Omega)$ for which

$$|f|_{B^s_{p,q}(\Omega)} := \left(\int_0^{\mathrm{diam}(\Omega)} \left[u^{-s}\omega_r(f,\Omega,u)_p\right]^q \frac{\mathrm{d}u}{u}\right)^{1/q} < \infty \tag{13}$$

(with the usual modification if $q = \infty$) and $r = \lfloor s \rfloor + 1$. Here the modulus of smoothness involved is the usual one given in (4). The space $B^s_{p,q}(\Omega)$ is then quasi-normed via $\|f\|_{B^s_{p,q}(\Omega)} := \|f\|_{L_p(\Omega)} + |f|_{B^s_{p,q}(\Omega)}$. For more information on these spaces we refer to [DL93, Tri83].

Later on it will be useful for us to discretize the quasi-seminorm (11) as follows.

Lemma 4. *The quasi-seminorm (11) for $B^s_{p,q}(I,X)$ is equivalent to*

$$\begin{aligned} |f|^*_{B^s_{p,q}(I,X)} &:= \left(\sum_{k=0}^{\infty} \left[2^{ks} \omega_r(f,I,2^{-k})_p \right]^q \right)^{1/q}, \qquad 0 < q < \infty, \\ |f|^*_{B^s_{p,\infty}(I,X)} &:= \sup_{k \geq 0} 2^{ks} \omega_r(f,I,2^{-k})_p, \end{aligned} \tag{14}$$

with constants of equivalence independent of f and I as $|I| \to 0$.

Proof. The proof follows along the lines of the standard case, which may be found in [DL93, p. 56]. Using (5) with $m = 2$ and the monotonicity of $\omega_r(f,I,\cdot)$ we see that for $u \in [2^{-k-1}, 2^{-k}]$ it holds

$$\begin{aligned} 2^{-r} \left(2^{ks} \omega_r(f,I,2^{-k})_p \right)^{\min\{1,p\}} &\leq \left(u^{-s} \omega_r(f,I,u) \right)_p^{\min\{1,p\}} \\ &\leq \left(2^{(k+1)s} \omega_r(f,I,2^{-k})_p \right)^{\min\{1,p\}}. \end{aligned}$$

Raising all terms of the inequality to the power $\frac{1}{\min\{1,p\}}$ we obtain

$$u^{-s} \omega_r(f,I,u)_p \simeq 2^{ks} \omega_r(f,I,2^{-k})_p \qquad \text{for} \qquad u \in [2^{-k-1}, 2^{-k}].$$

Hence, since $\int_{2^{-k-1}}^{2^{-k}} \frac{\mathrm{d}u}{u} = \ln 2 \simeq 1$ we get

$$\left(\int_{2^{-k-1}}^{2^{-k}} [u^{-s} \omega_r(f,I,u)_p]^q \frac{\mathrm{d}u}{u} \right)^{1/q} \simeq 2^{ks} \omega_r(f,I,2^{-k})_p.$$

This completes the proof for $0 < q < \infty$ taking into account Remark 2, after adding all terms for $k \geq 0$. The case $q = \infty$ is analogous. □

2.2.1. *Embedding results.* Before we provide some embeddings for the scale $B^s_{p,q}(I,X)$ needed later on, let us briefly recall what is known concerning the Besov spaces $B^s_{p,q}(\Omega)$.

Proposition 5. *Let $s > 0$ and $0 < p, q \leq \infty$.*

(i) *Let $0 < \varepsilon < s$, $0 < \nu \leq \infty$, and $q \leq \vartheta \leq \infty$, then*

$$B^s_{p,q}(\Omega) \hookrightarrow B^{s-\varepsilon}_{p,\nu}(\Omega) \qquad \text{and} \qquad B^s_{p,q}(\Omega) \hookrightarrow B^s_{p,\vartheta}(\Omega).$$

(ii) *(Sobolev-type embedding) Let $0 < \sigma < s$ and $p < \tau$ be such that*

$$s - \frac{n}{p} \geq \sigma - \frac{n}{\tau}, \tag{15}$$

then

$$B^s_{p,q}(\Omega) \hookrightarrow B^\sigma_{\tau,\vartheta}(\Omega), \tag{16}$$

where $0 < \vartheta \leq \infty$ and, additionally, $q \leq \vartheta$ if an equality holds in (15). Moreover, in the limiting case when $\sigma = 0$ and ϑ is such that

$$s - \frac{n}{p} \geq -\frac{n}{\vartheta}, \tag{17}$$

we have

$$B^s_{p,q}(\Omega) \hookrightarrow L_\vartheta(\Omega), \tag{18}$$

where again $q \leq \vartheta$ if an equality holds in (17).

(iii) *If the domain $\Omega \subset \mathbb{R}^n$ is bounded, then for $\tau \leq p$ we have the embedding*

$$B^s_{p,q}(\Omega) \hookrightarrow B^s_{\tau,q}(\Omega). \tag{19}$$

Remark 6. (i) The above results can be found in [DL93, § 2.10, 12.8], [HS09, Thm. 1.15], and [BS88].
In the interpolation diagram aside we have illustrated the area of possible embeddings of a fixed original space $B^s_{p,q}(\Omega)$ into spaces $B^{\sigma_1}_{\tau_1,\nu_1}(\Omega)$ and $B^{\sigma_2}_{\tau_2,\nu_2}(\Omega)$. The lighter shaded area corresponds to the additional embeddings we have if the underlying domain Ω is bounded.

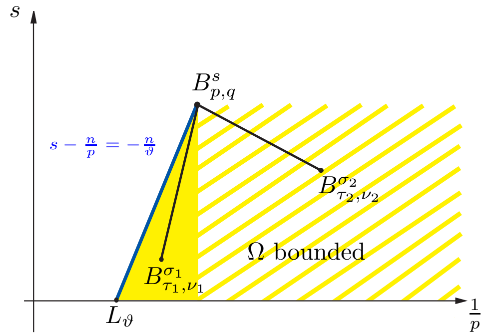


FIGURE 2. Embeddings for $B^s_{p,q}(\Omega)$

(ii) In the non-limiting case (corresponding to the strict inequality in (15) and (17)) the embeddings in Proposition 5 are known to be compact. In particular, for $\alpha > 0$ and $p < \tau$, the embeddings $B^{s+\alpha}_{p,p}(\Omega) \hookrightarrow B^s_{\tau,\tau}(\Omega)$ $(s > 0)$ and $B^\alpha_{p,p}(\Omega) \hookrightarrow L_\tau(\Omega)$ $(s = 0)$ are compact if, and only if,

$$\alpha - \frac{n}{p} > -\frac{n}{\tau}.$$

For the scale $B^s_{p,q}(I, X)$ there are counterparts of the embeddings from Proposition 5.

Proposition 7. *Assume $s > 0$ and $0 < p, q \leq \infty$.*

(i) *Let $0 < \varepsilon < s$, $0 < \nu \leq \infty$, and $q \leq \vartheta \leq \infty$, then*

$$B^s_{p,q}(I, X) \hookrightarrow B^{s-\varepsilon}_{p,\nu}(I, X) \qquad \text{and} \qquad B^s_{p,q}(I, X) \hookrightarrow B^s_{p,\vartheta}(I, X).$$

(ii) *If the time interval I is bounded, then for $\tau \leq p$ we have the embedding*

$$B^s_{p,q}(I, X) \hookrightarrow B^s_{\tau,q}(I, X). \tag{20}$$

Proof. The embeddings in (i) and (ii) can be proven as in the standard case, using the discrete version of the seminorm for Besov spaces, i.e.,

$$|f|_{B^s_{p,q}(I,X)} \simeq |f|^*_{B^s_{p,q}(I,X)} = \left(\sum_{k=0}^{\infty} [2^{ks} \omega_r(f, 2^{-k})_p]^q \right)^{\frac{1}{q}}, \quad 0 < q < \infty,$$

with the analogous one for $q = \infty$. Indeed, the second embedding in (i) is just a consequence of the monotonicity of the ℓ_q sequence spaces, i.e., $\ell_q \hookrightarrow \ell_\vartheta$ for $q \leq \vartheta$. The first embedding for $q \leq \nu$ is also clear since $2^{k(s-\varepsilon)} \leq 2^{ks}$. If $\nu < q$ one uses Hölder's inequality with $\frac{q}{\nu} > 1$, which gives the desired result.
Moreover, (ii) follows immediately since for $\tau \leq p$ and $|I| < \infty$ we have $L_p(I, X) \hookrightarrow L_\tau(I, X)$. □

Remark 8. The counterpart of the limiting embedding (18) in Prop. 5(ii) is derived in Corollary 21 as an application of our generalized Whitney's estimate presented in Proposition 20. Moreover, the Sobolev-type embeddings as stated in Prop. 5(ii), formula (16), should also hold. The proof in the standard case, cf. [DL93, § 12.8], involves spline representations for Besov spaces, which we have not provided for our generalized setting so far. This is out of the scope of the present paper.

3. Jackson- and Whitney-type theorems for vector-valued functions

In this section we prove Jackson- and Whitney-type theorems for functions defined on an interval, but valued on a Banach space. Some proofs are rather technical, and analogous to the ones presented for scalar-valued functions in [Sto77, OS78].

Let us mention that regarding Jackson's theorem there is a proof for $1 \le p \le \infty$, which is based on the K-functional method of interpolation [PP87, §3.5] and seems extendable to vector-valued functions. There is an alternative proof in [PP87, Thm. 7.1], which holds for $0 < p \le \infty$ and avoids all the technicalities from [Sto77, OS78]. However, it is based on a contradiction argument and does not work in the vector-valued case, or at least we could not generalize it to the infinite-dimensional setting.

The proof of Whitney's theorem that we present below in Section 3.2 follows the steps from [DeV98, Sect. 6.1]. In order to do it, we need an equivalence of L_p-norms for vector-valued polynomials, which is contained in Lemma 17 and Corollary 19. After proving Whitney's estimate in $B^s_{q,q}(I,X) \cap L_p(I,X)$ in Proposition 20 we obtain the embedding $B^s_{q,q}(I,X) \subset L_p(I,X)$, and arrive at Whitney's estimate in $B^s_{q,q}(I,X)$.

3.1. **Jackson's estimate.** The goal of this section is to prove a Jackson-type estimate, which is stated below in Theorem 9 and requires some definitions.

Given a separable Banach space X, $r \in \mathbb{N}$, and an interval $I = [a,b)$, we denote by $\mathbb{V}^r_{I,X}$ the space of X-valued polynomials of order r w.r.t. time, which we define as follows:

$$\mathbb{V}^r_{I,X} := \left\{ P : I \to X, P(t) = \sum_{j=1}^{r} \ell^r_j(t) P_j : \ P_j \in X, \ t \in I \right\}, \tag{21}$$

with ℓ^r_j the usual (scalar-valued) Lagrange basis functions

$$\ell^r_j(t) = \prod_{i \ne j} \frac{t - t_i}{t_j - t_i} \qquad \text{for } t_j = a + (j-1)\frac{b-a}{r-1}, \qquad j = 1, 2, \dots, r. \tag{22}$$

Notice that any basis for the space Π^r of scalar-valued polynomials in $\mathbb{R}$, such as $1, t, t^2, \dots, t^{r-1}$, leads to the same space $\mathbb{V}^r_{I,X}$.

The main result of this section is the following.

Theorem 9 (Jackson's Theorem)**.** *Let $0 < p \le \infty$ and $r \in \mathbb{N}$. Then there exists a constant $c > 0$ such that for any interval I and every $f \in L_p(I,X)$, there exists a vector-valued polynomial $P_r \in \mathbb{V}^r_{I,X}$, which satisfies*

$$\|f - P_r\|^p_{L_p(I,X)} \le c\, w_r(f, I, h)^p_p \qquad \text{with} \qquad h = \frac{|I|}{2r}. \tag{23}$$

In other words, there exist $a_0, a_1, \dots, a_{r-1} \in X$ such that, if $P_r(t) = a_0 + a_1 t + \dots + a_{r-1} t^{r-1}$, then (23) holds.

Due to the homogeneity (5) and the equivalence of Lemma 1, Jackson's estimate can also be stated as:

$$E_r(f,I)_p := \inf_{P_r \in \mathbb{V}^r_{I,X}} \|f - P_r\|^p_{L_p(I,X)} \le c\, w_r(f, I, |I|)^p_p, \qquad \forall f \in L_p(I,X). \tag{24}$$

In order to prove this estimate, we need several auxiliary lemmas, which are rather technical, and analogous to the ones proved for scalar-valued functions in [Sto77, OS78]. We generalize them to our setting. The basic idea is to first study periodic functions and their higher order differences, and then relate them to differences of the functions we are actually interested in.

Let $f : [a,b) \to X$ be an X-valued function and f^* denote its periodic continuation with period $d := b-a$, i.e.,

$$f^*(t) = f(t-\ell d), \quad \text{where } \ell \in \mathbb{Z} \text{ is such that } t-\ell d \in [a,b).$$

Moreover, for $0<p<\infty$ and $k \in \mathbb{N}$ consider the integrals

$$I^*_{p,k}(h) := \int_a^b \|\Delta_h^k f^*(t)\|_X^p \mathrm{d}t = \int_0^d \|\Delta_h^k f^*(t)\|_X^p \mathrm{d}t, \tag{25}$$

$$I_{p,k}(h) := \int_a^{b-kh} \|\Delta_h^k f(t)\|_X^p \mathrm{d}t. \tag{26}$$

Note that we do not emphasize on the fact that the expressions $I^*_{p,k}(h)$ and $I_{p,k}(h)$ also depend on the functions f and f^*, respectively, since it will always be clear from the context which function we deal with.

We start with the following result showing how the best approximation of some function $f \in L_p(I,X)$ by a constant $a_0 \in X$ can be bounded using first differences of its periodic continuation f^*.

Lemma 10. *Let $0<p<\infty$ and $f \in L_p(I,X)$. There exists $a_0 \in X$ such that*

$$\|f-a_0\|^p_{L_p(I,X)} \le \frac{1}{d}\int_0^d I^*_{p,1}(y)\mathrm{d}y.$$

Proof. We show how to construct $a_0 \in X$ satisfying the desired inequality. Let f^* denote the d-periodic continuation of f. We make the following easy observation,

$$\begin{aligned}\inf_{a_0\in X}\|f-a_0\|^p_{L_p(I,X)} &= \inf_{a_0\in X}\int_0^d \|f^*(t)-a_0\|_X^p\mathrm{d}t \\ &= \inf_{a_0\in X}\int_0^d \|f^*(t+y)-a_0\|_X^p\mathrm{d}t \\ &\le \int_0^d \|f^*(t+y)-f^*(y)\|_X^p\mathrm{d}t, \quad \text{for any } y\in[0,d).\end{aligned}$$

Now using the fact that f^* is d-periodic and the left-hand side does not depend on y, integration from 0 to d w.r.t. y yields

$$\begin{aligned}\inf_{a_0\in X}\|f-a_0\|^p_{L_p(I,X)} &\le \frac{1}{d}\int_0^d\int_0^d \|f^*(t+y)-f^*(y)\|_X^p\mathrm{d}t\mathrm{d}y \\ &= \frac{1}{d}\int_a^b\int_a^b \|f(t)-f(y)\|_X^p\mathrm{d}t\mathrm{d}y = \frac{1}{d}\int_a^b g(y)\mathrm{d}y,\end{aligned}$$

where in the last line we put $g(y) := \int_a^b \|f(t)-f(y)\|_X^p\mathrm{d}t$. Note that the set S defined as

$$S := \left\{z\in[a,b):\ g(z) \le \frac{1}{d}\int_a^b g(y)\mathrm{d}y\right\},$$

is non-empty. Therefore, taking $z\in S$ and putting $a_0 := f(z)$ we obtain

$$\|f-a_0\|^p_{L_p(I,X)} = \int_a^b \|f(t)-f(z)\|_X^p\mathrm{d}t = g(z)$$

$$
\begin{aligned}
&\leq \frac{1}{d}\int_a^b\int_a^b \|f(t)-f(y)\|_X^p \mathrm{d}t\mathrm{d}y \\
&= \frac{1}{d}\int_0^d\int_0^d \|f^*(t+y)-f^*(y)\|_X^p \mathrm{d}t\mathrm{d}y \\
&= \frac{1}{d}\int_0^d\int_0^d \|f^*(t+y)-f^*(y)\|_X^p \mathrm{d}y\mathrm{d}t \\
&= \frac{1}{d}\int_0^d I_{p,1}^*(t)\mathrm{d}t,
\end{aligned}
$$

which shows that $a_0 := f(z)$ with $z \in S$ yields the assertion. □

The following lemma shows that we can bound integrals of lower order differences of periodic functions with integrals involving higher order differences.

Lemma 11. *Let $0 < p < \infty$ and $k \in \mathbb{N}$. Then we have the following relation*

$$\int_0^d I_{p,k}^*(y)\mathrm{d}y \leq c\int_0^d I_{p,k+1}^*(y)\mathrm{d}y,$$

with the constant $c > 0$ only depending on k and p, but otherwise independent of the function f and the interval $[a, b)$.

Proof. We make use of the following identity

$$\Delta_{2y}^k f^*(t) - 2^k\Delta_y^k f^*(t) = \sum_{i=1}^k \binom{k}{i}\sum_{m=0}^{i-1}\Delta_y^{k+1} f^*(t+my), \tag{27}$$

which can be found in [Tim63, Sect. 3.3.2]. Let $0 < p < 1$. In this case we know that $|\cdot|^p$ is subadditive. This and integration from 0 to d w.r.t. t in (27) leads to

$$
\begin{aligned}
2^{kp}\int_0^d \|\Delta_y^k f^*(t)\|_X^p \mathrm{d}t - \int_0^d \|\Delta_{2y}^k f^*(t)\|_X^p \mathrm{d}t& \\
\leq \sum_{i=1}^k \binom{k}{i}\sum_{m=0}^{i-1}\int_0^d \|\Delta_y^{k+1} f^*(t+my)\|_X^p \mathrm{d}t&.
\end{aligned}
$$

Now integrating once more from 0 to d w.r.t. y and using the definition of $I_{p,k}^*$ gives

$$2^{kp}\int_0^d I_{p,k}^*(y)\mathrm{d}y - \int_0^d I_{p,k}^*(2y)\mathrm{d}y \leq \sum_{i=1}^k \binom{k}{i} i\int_0^d I_{p,k+1}^*(y)\mathrm{d}y. \tag{28}$$

Since $I_{p,k}^*(y)$ is d-periodic, we have the identity

$$\int_0^d I_{p,k}^*(2y)\mathrm{d}y = \frac{1}{2}\int_0^{2d} I_{p,k}^*(y')\mathrm{d}y' = \int_0^d I_{p,k}^*(y)\mathrm{d}y.$$

Inserting this in (28) we obtain

$$\left(2^{kp}-1\right)\int_0^d I_{p,k}^*(y)\mathrm{d}y \leq c_{k,p}\int_0^d I_{p,k+1}^*(y)\mathrm{d}y, \tag{29}$$

which gives the desired estimate in the case $0 < p < 1$. When $1 \leq p < \infty$ we proceed with (27) as follows: We add $2^k\Delta_y^k f^*(t)$ on both sides of (27) and integrate from 0 to d w.r.t. t and from 0 to d w.r.t. y afterwards in the L_p-norm. This gives (28) but with the integrals to the power $\frac{1}{p}$. We proceed as before and end up with (29) to the power $\frac{1}{p}$, which proves the asserted estimate. $\square$

The following lemma shows how to bound integrals of higher order differences of the periodic extension of a function by integrals of higher order differences of the original function plus first order differences.

Lemma 12. *Let $f \in L_p(I, X)$, where $0 < p < \infty$. Then for any $k \in \mathbb{N}$ it holds*

$$\int_0^{\frac{d}{k}} I^*_{p,k}(y)\mathrm{d}y \le 2\int_0^{\frac{d}{k}} I_{p,k}(y)\mathrm{d}y + c\, d\, I_{p,1}\left(\frac{d}{k}\right).$$

with the constant $c > 0$ only depending on k and p, but otherwise independent of the function f and the interval $[a, b)$.

Proof. By definition $\Delta^k_y f^*(t) = \sum_{i=0}^k (-1)^{k-i}\binom{k}{i} f^*(t + iy)$ and the fact that f^* is the d-periodic continuation of f, i.e., $f = f^*$ on $[a, b)$ and $f^*(t) = f(t - d)$ for some $t \in [b, b + d)$, we express $I^*_{p,k}$ in terms of the values of f as follows:

$$\begin{aligned} I^*_{p,k}(y) &= \int_a^b \|\Delta^k_y f^*(t)\|^p_X \mathrm{d}t \\ &= \int_a^{b-ky} \|\Delta^k_y f(t)\|^p_X \mathrm{d}t + \sum_{j=1}^k \int_{b-jy}^{b-(j-1)y} \|S_j\|^p_X \mathrm{d}t, \qquad 0 \le y \le \frac{d}{k}, \end{aligned} \tag{30}$$

where

$$S_j(t) = \sum_{i=0}^{j-1} (-1)^{k-i}\binom{k}{i} f(t + iy) + \sum_{i=j}^{k} (-1)^{k-i}\binom{k}{i} f(t + iy - d).$$

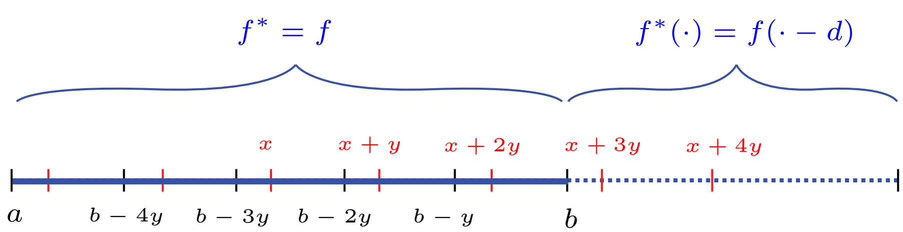


FIGURE 3. Express f^* via f with $j = 3$

Now we transform S_j as follows: we augment the first term of the first sum and the last of the second sum in order to obtain the value of the k-th difference of f at the point t with step $y - \frac{d}{k}$. This yields

$$\begin{aligned} S_j &= \sum_{i=0}^{k} (-1)^{k-i}\binom{k}{i} f\left(t + i\left(y - \frac{d}{k}\right)\right) \quad \begin{matrix}\nearrow i = 0 \text{ first sum} \\ \searrow i = k \text{ second sum}\end{matrix} && (=: T_1) \\ &+ \sum_{i=1}^{j-1} (-1)^{k-i}\binom{k}{i} \left[f(t + iy) - f\left(t + i\left(y - \frac{d}{k}\right)\right)\right] && (=: T_2(j)) \\ &+ \sum_{i=j}^{k-1} (-1)^{k-i}\binom{k}{i} \left[f(t + iy - d) - f\left(t + i\left(y - \frac{d}{k}\right)\right)\right] && (=: T_3(j)) \end{aligned}$$

$$=: T_1 + T_2(j) + T_3(j).$$

Since T_1 does not depend on j,

$$\sum_{j=1}^{k} \int_{b-jy}^{b-(j-1)y} \|T_1\|_X^p \,\mathrm{d}t = \int_{b-ky}^{b} \|T_1\|_X^p \mathrm{d}t = \int_{b-ky}^{b} \|\Delta^k_{y-\frac{d}{k}} f(t)\|_X^p \,\mathrm{d}t$$
$$= \int_{a}^{a+ky} \|\Delta^k_{\frac{d}{k}-y} f(t)\|_X^p \mathrm{d}t = I_{p,k}\left(\frac{d}{k} - y\right), \tag{31}$$

where in the third step we changed the step $y - \frac{d}{k}$ involving the k-th difference of f into $\frac{d}{k} - y$ in order to obtain a nonnegative step. We now estimate the sum

$$\sum_{j=1}^{k} \int_{b-jy}^{b-(j-1)y} \|T_2(j)\|_X^p + \|T_3(j)\|_X^p \mathrm{d}t$$
$$\lesssim \sum_{j=1}^{k} \int_{b-jy}^{b-(j-1)y} \left\{ \sum_{i=1}^{j-1} \binom{k}{i}^p \left\| f(t+iy) - f\left(t + i\Big(y - \frac{d}{k}\Big)\right) \right\|_X^p \right.$$
$$\left. + \sum_{i=j}^{k-1} \binom{k}{i}^p \left\| f(t+iy-d) - f\left(t + i\Big(y - \frac{d}{k}\Big)\right) \right\|_X^p \right\} \mathrm{d}t$$

(change summation $\sum_{j=1}^{k} \sum_{i=1}^{j-1} = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k}$ and $\sum_{j=1}^{k} \sum_{i=j}^{k-1} = \sum_{i=1}^{k-1} \sum_{j=1}^{i}$)

$$= \sum_{i=1}^{k-1} \binom{k}{i}^p \int_{b-ky}^{b-iy} \left\| f(t+iy) - f\left(t + i\Big(y - \frac{d}{k}\Big)\right) \right\|_X^p \mathrm{d}t$$
$$+ \sum_{i=1}^{k-1} \binom{k}{i}^p \int_{b-iy}^{b} \left\| f(t+iy-d) - f\left(t + i\Big(y - \frac{d}{k}\Big)\right) \right\|_X^p \mathrm{d}t$$

(1st integral: Substitution $t' = t + i\left(y - \frac{d}{k}\right)$; reverse sum $i \mapsto k - i$)

(2nd integral: Substitution $t'' = t + iy - d$)

$$= \sum_{i=1}^{k-1} \binom{k}{i}^p \int_{\frac{k-i}{k}a + \frac{i}{k}b - iy}^{\frac{k-i}{k}a + \frac{i}{k}b} \left\| f\left(t' + (k-i)\frac{d}{k}\right) - f(t') \right\|_X^p \mathrm{d}t'$$
$$+ \sum_{i=1}^{k-1} \binom{k}{i}^p \int_{a}^{a+iy} \left\| f(t'') - f\left(t'' + (k-i)\frac{d}{k}\right) \right\|_X^p \mathrm{d}t''. \tag{32}$$

Now (30), (31), and (32) yield

$$I^*_{p,k}(y) \leq I_{p,k}(y) + I_{p,k}\left(\frac{d}{k} - y\right)$$
$$+ \sum_{i=1}^{k-1} \binom{k}{i}^p \int_{\frac{k-i}{k}a + \frac{i}{k}b - iy}^{\frac{k-i}{k}a + \frac{i}{k}b} \left\| f\left(t' + (k-i)\frac{d}{k}\right) - f(t') \right\|_X^p \mathrm{d}t'$$
$$+ \sum_{i=1}^{k-1} \binom{k}{i}^p \int_{a}^{a+iy} \left\| f(t'') - f\left(t'' + (k-i)\frac{d}{k}\right) \right\|_X^p \mathrm{d}t''.$$

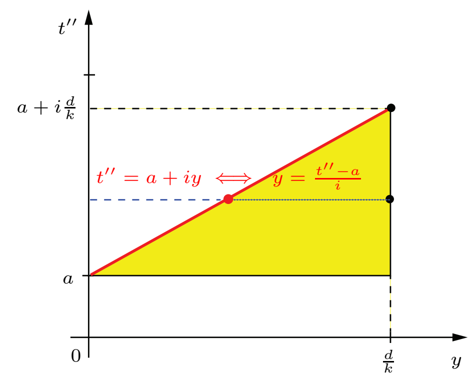
Integrating from 0 to $\frac{d}{k}$ w.r.t. y gives

$$\int_0^{\frac{d}{k}} I^*_{p,k}(y)\mathrm{d}y \le \int_0^{\frac{d}{k}} I_{p,k}(y)\mathrm{d}y + \int_0^{\frac{d}{k}} I_{p,k}\left(\frac{d}{k}-y\right)\mathrm{d}y + c\left\{\sum_{i=1}^{k-1}\int_0^{\frac{d}{k}}\int_{\frac{k-i}{k}a+\frac{i}{k}b-iy}^{\frac{k-i}{k}a+\frac{i}{k}b}\left\|f\left(t'+(k-i)\frac{d}{k}\right)-f\left(t'\right)\right\|_X^p \mathrm{d}t'\mathrm{d}y + \sum_{i=1}^{k-1}\int_0^{\frac{d}{k}}\int_a^{a+iy}\left\|f(t'')-f\left(t''+(k-i)\frac{d}{k}\right)\right\|_X^p \mathrm{d}t''\mathrm{d}y\right\}.$$

We change the order of integration in the double integrals. For the second integral this yields

$$\int_0^{\frac{d}{k}}\int_a^{a+iy}(\ldots)\mathrm{d}t''\mathrm{d}y \longrightarrow \int_a^{a+i\frac{d}{k}}\int_{\frac{t''-a}{i}}^{\frac{d}{k}}(\ldots)\mathrm{d}y\mathrm{d}t''.$$

Similarly for the first one. Moreover, observing that the integrand in both cases does not depend on y we obtain



$$\begin{aligned}\int_0^{\frac{d}{k}} I^*_{p,k}(y)\mathrm{d}y &\le 2\int_0^{\frac{d}{k}} I_{p,k}(y)\mathrm{d}y + c\left\{\sum_{i=1}^{k-1}\int_a^{a+i\frac{d}{k}}\left(\frac{t'}{i}-\frac{a}{i}\right)\left\|f\left(t'+(k-i)\frac{d}{k}\right)-f\left(t'\right)\right\|_X^p\mathrm{d}t'\right.\\ &\quad\left.+\sum_{i=1}^{k-1}\int_a^{a+i\frac{d}{k}}\left(\frac{d}{k}-\frac{t''-a}{i}\right)\left\|f(t'')-f\left(t''+(k-i)\frac{d}{k}\right)\right\|_X^p\mathrm{d}t''\right\}\\ &= 2\int_0^{\frac{d}{k}} I_{p,k}(y)\mathrm{d}y + c\frac{d}{k}\sum_{i=1}^{k-1}\int_a^{a+i\frac{d}{k}}\left\|f(t)-f\left(t+(k-i)\frac{d}{k}\right)\right\|_X^p\mathrm{d}t. \qquad (33)\end{aligned}$$

Using a telescopic sum we see that

$$\left\|f(t)-f\left(t+(k-i)\frac{d}{k}\right)\right\|_X^p \lesssim \sum_{j=1}^{k-i}\left\|f\left(t+(j-1)\frac{d}{k}\right)-f\left(t+j\frac{d}{k}\right)\right\|_X^p$$

and for $i=1,2,\ldots,k-1$,

$$\begin{aligned}\sum_{j=1}^{k-i}\int_a^{a+i\frac{d}{k}}&\left\|f\left(t+(j-1)\frac{d}{k}\right)-f\left(t+j\frac{d}{k}\right)\right\|_X^p\mathrm{d}t\\ &=\sum_{j=1}^{k-i}\int_{a+\frac{j-1}{k}d}^{a+\frac{i+j-1}{k}d}\left\|f\left(t'\right)-f\left(t'+\frac{d}{k}\right)\right\|_X^p\mathrm{d}t'\\ &\le (k-i)\int_a^{a+\frac{k-1}{k}d}\left\|f\left(t'\right)-f\left(t'+\frac{d}{k}\right)\right\|_X^p\mathrm{d}t', \qquad (34)\end{aligned}$$

where in the second step we used a change of variables $t' := t + (j-1)\frac{d}{k}$. Inserting (34) into (33) finally gives

$$\begin{aligned}\int_0^{\frac{d}{k}} I^*_{p,k}(y)\mathrm{d}y &\le 2\int_0^{\frac{d}{k}} I_{p,k}(y)\mathrm{d}y + d\, c_{k,p}\int_a^{b-\frac{d}{k}} \left\| f(t') - f\left(t' + \frac{d}{k}\right)\right\|_X^p \mathrm{d}t' \\ &= 2\int_0^{\frac{d}{k}} I_{p,k}(y)\mathrm{d}y + c_{k,p}\, d\, I_{p,1}\left(\frac{d}{k}\right),\end{aligned}$$

which completes the proof. □

The previous lemmas give the following result, which shows that we can bound the best approximation of a function $f \in L_p(I,X)$ by a constant $a_0 \in X$ with the help of integrals of higher order differences and first order differences of f.

Lemma 13. *Let $I = [a,b)$, $0 < p < \infty$, and $m \in \mathbb{N}$. There exists a constant $c = c_{m,p}$, such that for every $f \in L_p(I,X)$ there exists $a_0 \in X$ satisfying, for $h = \frac{b-a}{m}$,*

$$\begin{aligned}\|f - a_0\|^p_{L_p(I,X)} &\le c\left[\frac{1}{h}\int_0^h \int_a^{b-ms} \|\Delta_s^m f(t)\|_X^p \mathrm{d}t\mathrm{d}s + \int_a^{b-h} \|\Delta_h f(t)\|_X^p \mathrm{d}t\right] \quad (35) \\ &= c\left[w_m(f,I,h)_p^p + \|\Delta_h f(t)\|^p_{L_p(I_h,X)}\right].\end{aligned}$$

Remark 14. Note that the second term with the first order differences in (35) is crucial: If f is a polynomial of degree $m-1$ the first integral on the right-hand side vanishes but the left-hand side might not.

Proof. We first notice that, by induction, we can easily check that

$$\Delta^m_{my} f^*(t) = \sum_{i_m=0}^{m-1} \cdots \sum_{i_1=0}^{m-1} \Delta_y^m f^*(t + i_1 y + \cdots + i_m y),$$

so that $I^*_{p,m}(my) \lesssim I^*_{p,m}(y)$. Taking $a_0 \in X$ as constructed in Lemma 10 and using Lemmas 11 and 12 with $k = m$, setting $h = \frac{d}{m}$, we obtain

$$\begin{aligned}\|f - a_0\|^p_{L_p(I,X)} &\lesssim \frac{1}{d}\int_0^d I^*_{p,1}(y)\mathrm{d}y \lesssim \frac{1}{d}\int_0^d I^*_{p,m}(y)\mathrm{d}y \\ &\lesssim \frac{1}{d}\int_0^d I^*_{p,m}\left(\frac{y}{m}\right)\mathrm{d}y = \frac{m}{d}\int_0^{\frac{d}{m}} I^*_{p,m}(y')\mathrm{d}y' \\ &\le \frac{m}{d}\left[2\int_0^{\frac{d}{m}} I_{p,m}(y)\mathrm{d}y + c_{p,m} d I_{p,1}\left(\frac{d}{m}\right)\right] \\ &= c'_{p,m}\frac{1}{h}\left[\int_0^h \int_a^{b-my} \|\Delta_y^m f\|_X^p \mathrm{d}t\mathrm{d}y + \int_a^{b-h} \|\Delta_h f(t)\|_X^p \mathrm{d}t\right],\end{aligned}$$

which is the desired result. □

Finally, a repeated application of Lemma 13 now allows us to establish Jackson's inequality.

Proof of Theorem 9. We assume $I = [0,1)$. The general case follows by scaling, using (7). Let $h = \frac{1}{2r}$, $f \in L_p(I,X)$, and denote the approximant $a_0 \in X$ from

Lemma 13 by $M(f,I) := a_0$. Now define the coefficients $a_0,\dots,a_{r-1}$ recursively as follows:

$$\begin{aligned}
a_{r-1} &= M\left(\Delta_h^{r-1} f, [0,1-(r-1)h]\right) \frac{1}{h^{r-1}} \frac{1}{(r-1)!},\\
f_1(t) &= f(t) - a_{r-1} t^{r-1},\\
a_{r-2} &= M(\Delta_h^{r-2} f_1, [0,1-(r-2)h]) \frac{1}{h^{r-2}} \frac{1}{(r-2)!},\\
f_2(t) &= f_1(t) - a_{r-2} t^{r-2} = f(t) - (a_{r-1} t^{r-1} + a_{r-2} t^{r-2}),\\
&\vdots\\
a_2 &= M(\Delta_h^2 f_{r-3}, [0,1-2h]) \frac{1}{h^2} \frac{1}{2!},\\
f_{r-2}(t) &= f_{r-3}(t) - a_2 t^2 = f(t) - (a_{r-1} t^{r-1} + a_{r-2} t^{r-2} + \ldots + a_2 t^2),\\
a_1 &= M(\Delta_h^1 f_{r-2}, [0,1-h]) \frac{1}{h},\\
f_{r-1}(t) &= f_{r-2}(t) - a_1 t = f(t) - (a_{r-1} t^{r-1} + a_{r-2} t^{r-2} + \ldots + a_1 t),\\
a_0 &= M(f_{r-1}, [0,1]).
\end{aligned}$$

With

$$P_r(t) = \sum_{k=0}^{r-1} a_k t^k = a_0 + a_1 t + \ldots + a_{r-1} t^{r-1}$$

we compute

$$\begin{aligned}
\|f - P_r\|^p_{L_p(I,X)} &= \|f_{r-1} - a_0\|^p_{L_p(I,X)} = \|f_{r-1} - M(f_{r-1},[0,1])\|^p_{L_p(I,X)}\\
&\lesssim w_{2r}\left(f_{r-1}, I, \frac{1}{2r}\right)^p_p + \|\Delta_h f_{r-1}\|^p_{L_p(I_h,X)}\\
&\qquad \text{(which follows from applying Lem. 13 with } m = 2r)\\
&= w_{2r}(f,I,h)^p_p + \|\Delta_h f_{r-2} - a_1 h\|^p_{L_p(I_h,X)}\\
&= w_{2r}(f,I,h)^p_p + \|\Delta_h f_{r-2} - M(\Delta_h f_{r-2},[0,1-h])\|^p_{L_p(I_h,X)}\\
&\lesssim w_{2r}(f,I,h)^p_p + w_{2r-1}(\Delta_h f_{r-2},[0,1-h],h)^p_p + \|\Delta_h^2 f_{r-2}\|^p_{L_p(I_{2h},X)}\\
&\qquad \text{(which follows from applying Lem. 13 with } m = 2r-1)\\
&\lesssim w_{2r-1}(f,I,h)^p_p + \|\Delta_h^2 f_{r-3} - a_2 2! h^2\|^p_{L_p(I_{2h},X)}\\
&\qquad \text{(we used (6))}\\
&= w_{2r-1}(f,I,h)^p_p + \|\Delta_h^2 f_{r-3} - M(\Delta_h^2 f_{r-3},[0,1-2h])\|^p_{L_p(I_{2h},X)}\\
&\lesssim w_{2r-2}(f,I,h)^p_p + \|\Delta_h^3 f_{r-4} - a_3 3! h^3\|^p_{L_p(I_{3h},X)}\\
&\qquad \text{(which follows from applying Lem. 13 with } m = 2r-2)\\
&\vdots
\end{aligned}$$

$$\lesssim w_{r+2}(f,I,h)_p^p + \|\Delta_h^{r-1} f - a_{r-1}(r-1)! h^{r-1}\|_{L_p(I_{(r-1)h},X)}^p$$
$$\lesssim w_{r+1}(f,I,h)_p^p + \|\Delta_h^r f\|_{L_p(I_{rh},X)}^p$$
(which follows from applying Lem. 13 with $m = r+1$)
$$\leq w_{r+1}(f,I,h)_p^p + w_r(f,I,h)_p^p$$
$$\lesssim w_r(f,I,h)_p^p,$$

which proves the theorem. □

Remark 15. Note that Theorem 9 also holds for $p = \infty$: if we extend (25) by

$$I_{\infty,k}^*(h) := \sup_{t\in[a,b)} \|\Delta_h^k f^*(t)\|_X = \sup_{t\in[0,d)} \|\Delta_h^k f^*(t)\|_X,$$

and similarly (26), then Lemmas 10–13 can be extended to the case when $p = \infty$ by obvious modifications in the proofs, i.e., mostly replacing the integrals by suprema. In this case the factors '$\frac{1}{d}$' and 'd' in Lemmas 10 and 12, respectively, disappear.

3.2. **Whitney's estimate.** Having established Jackson's estimate (24) in Theorem 9 we now proceed to prove Whitney's estimate.

Theorem 16 (Generalized Whitney's theorem)**.** *Let $0 < p, q \leq \infty$, $r \in \mathbb{N}$, and $s > 0$. If $(1/q - 1/p)_+ \leq s < r$ then there exists a constant $c > 0$ which depends only on p, q, r such that*

$$E_r(f,I)_p = \inf_{P\in\mathbb{V}_{I,X}^r} \|f - P\|_{L_p(I,X)} \leq c|I|^{s+\frac{1}{p}-\frac{1}{q}} |f|_{B_{q,q}^s(I,X)}, \tag{36}$$

for all $f \in B_{q,q}^s(I,X)$ and for any finite interval I.

Since this involves the L_p-norm on the left-hand side and an L_q-norm on the right-hand side, we first deal with the problem of how to switch from p-norms to q-norms for vector-valued polynomials. Using this together with the Jackson estimate, the fact that according to Lemma 4 we can express the quasi-norm of the Besov spaces $B_{p,q}^s(I,X)$ as a discrete summation instead of integrals yields Whitney's estimate.

Lemma 17. *Let $0 < p < \infty$ and $I = [0,1]$. On $\mathbb{V}_{I,X}^r$ the quasi-norm*

$$\|P\|_p := \left(\int_0^1 \|P(t)\|_X^p \mathrm{d}t\right)^{1/p}$$

is equivalent to the norm

$$\|P\|_* := \max_{j=1,\dots,r} \|P_j\|_X, \qquad P_j = P(t_j), \quad t_j = \frac{j-1}{r-1}, \quad j = 1,\dots,r.$$

The constants involved in the equivalence depend on r and p, but are otherwise independent of $P \in \mathbb{V}_{I,X}^r$.

Remark 18. At first sight, it may seem that this lemma is obvious, because it looks like an equivalence of quasi-norms in a finite-dimensional space. But this is not the case, since the space $\mathbb{V}_{I,X}^r$ is not finite-dimensional, when X is an arbitrary Banach space.
With slight modifications in the proof, Lemma 17 also holds for $p = \infty$ and the quasi-norm $\|P\|_\infty = \sup_{t\in[0,1]} \|P(t)\|_X$.

Proof. Let $\{\ell_j\}_{j=1}^r$ denote the Lagrange basis of Π^r corresponding to the equally spaced nodes $t_j = \frac{j-1}{r-1}$, $j = 1, \dots, r$ on $[0,1]$, i.e.,

$$\ell_j(t) = \prod_{i\neq j} \frac{t-t_i}{t_j-t_i}, \quad \text{so that} \quad \ell_j(t_i) = \delta_{ij} \quad \text{and} \quad P = \sum_{j=1}^r P_j \ell_j, \text{ if } P \in \mathbb{V}_{I,X}^r.$$

Obviously, for $P \in \mathbb{V}_{I,X}^r$,

$$\begin{aligned} \|P\|_p = \left(\int_0^1 \|P(t)\|_X^p \mathrm{d}t\right)^{1/p} &= \left(\int_0^1 \Big\| \sum_{j=1}^r \ell_j(t) P_j \Big\|_X^p \mathrm{d}t\right)^{1/p} \\ &\le c_{p,r} \sum_{j=1}^r \left(\int_0^1 \ell_j^p(t) \|P_j\|_X^p \mathrm{d}t\right)^{1/p} \le c_{p,r} \max_{j=1,\dots,r} \|P_j\|_X = c_{p,r}\|P\|_*. \end{aligned}$$

Let now $P = \sum_{j=1}^r P_j\ell_j \in \mathbb{V}_{I,X}^r$ and let i be such that $\|P_i\|_X = \max_j \|P_j\|_X = \|P\|_*$. Then, for each $t \in I$, we have

$$\begin{aligned} \|P(t)\|_X = \Big\| \sum_{j=1}^r \ell_j(t) P_j \Big\|_X &\ge |\ell_i(t)|\|P_i\|_X - \sum_{j\neq i} |\ell_j(t)|\|P_j\|_X \\ &\ge \|P\|_* \Big(|\ell_i(t)| - \sum_{j\neq i} |\ell_j(t)| \Big). \end{aligned} \tag{37}$$

Since at the point $t = t_i$ we have $\ell_i(t_i) = 1$ and $\ell_j(t_i) = 0$ for all $j \neq i$, there exists $\delta > 0$ such that

$$|t - t_i| < \delta \quad \Longrightarrow \quad |\ell_i(t)| > \frac{3}{4} > \frac{1}{4} > \sum_{j\neq i} |\ell_j(t)|;$$

notice that $\delta > 0$ can be chosen independent of i, but will depend on r. Hence,

$$|\ell_i(t)| - \sum_{i\neq j} |\ell_j(t)| > \frac{1}{2}.$$

Hence, (37) gives us

$$\|P(t)\|_X \ge \frac{1}{2}\|P\|_* \qquad \text{for} \qquad |t - t_i| < \delta.$$

Raising to the power p and averaging over the interval $(t_i - \delta, t_i + \delta) \cap I$ yields

$$\|P\|_* \le \left(\frac{2^p}{\delta} \int_{(t_i-\delta,t_i+\delta)\cap I} \|P(t)\|_X^p \mathrm{d}t \right)^{1/p} \le \bar{c}_{p,r} \left(\int_I \|P(t)\|_X^p \mathrm{d}t \right)^{1/p} = \bar{c}_{p,r}\|P\|_p,$$

and the assertion follows. □

By a scaling argument we obtain from the previous Lemma the following equivalence of $L_p(I,X)$ norms in $\mathbb{V}_{I,X}^r$ on an arbitrary interval I. The proof is very simple and is thus omitted.

Corollary 19. *Let $0 < p, q \le \infty$ and $r \in \mathbb{N}$. Then there exists a constant $c > 0$ which depends only on p, q, r such that on any finite interval I,*

$$\|P\|_{L_p(I,X)} \le c|I|^{1/p-1/q} \|P\|_{L_q(I,X)}, \qquad \forall P \in \mathbb{V}_{I,X}^r. \tag{38}$$

Following the steps from [DeV98, Sec. 6.1] we can now prove Whitney's estimate in $B^s_{q,q}(I,X) \cap L_p(I,X)$.

Proposition 20. *Let $0 < p, q \le \infty$, $r \in \mathbb{N}$, and $s > 0$. If $(1/q - 1/p)_+ \le s < r$ then there exists a constant $c > 0$ which depends only on p, q, r such that*

$$E_r(f,I)_p := \inf_{P \in \mathbb{V}^r_{I,X}} \|f - P\|_{L_p(I,X)} \le c|I|^{s+\frac{1}{p}-\frac{1}{q}} |f|_{B^s_{q,q}(I,X)}, \tag{39}$$

for all $f \in B^s_{q,q}(I,X) \cap L_p(I,X)$ and for any finite interval I.

Proof. Since $E_{r+1}(f,I)_p \le E_r(f,I)_p$, it is sufficient to prove the result in the case $r = \lfloor s \rfloor + 1$, and by scaling it is sufficient to consider $I = [0,1)$. Also, since $E_r(f,I)_p \le E_r(f,I)_q$ when $p < q$, it is sufficient to consider the case $q \le p$.

Let D_k for $k = 0,1,2,\dots$ denote the following dyadic partitions of I:

$$D_k := \{I^j_k := 2^{-k}[j-1, j),\ j = 1,\dots,2^k\}.$$

We let S_k denote a piecewise polynomial function of order r on the partition D_k satisfying the Jackson estimate (24) with p replaced by q, in each sub-interval, i.e.,

$$\|f - S_k\|_{L_q(I^j_k,X)} \lesssim w_r(f, I^j_k, 2^{-k})_q, \qquad j = 1,2,\dots,2^k, \quad k = 0,1,\dots,$$

whence $S_0 \in \mathbb{V}^r_{I,X}$.

Then, on the one hand, we have

$$\|f - S_k\|^q_{L_q(I,X)} = \sum_{j=1}^{2^k} \|f - S_k\|^q_{L_q(I^j_k,X)} \lesssim \sum_{j=1}^{2^k} w_r(f, I^j_k, 2^{-k})^q_q.$$

Denoting $\tilde{I}^j_k = \left(I^j_k\right)_{rh}$ we obtain

$$\begin{aligned}
\|f - S_k\|^q_{L_q(I,X)} &\lesssim \frac{1}{2^{-k}} \int_0^{2^{-k}} \sum_{j=1}^{2^k} \|\Delta^r_h f\|^q_{L_q(\tilde{I}^j_k,X)} \mathrm{d}h \\
&= \frac{1}{2^{-k}} \int_0^{2^{-k}} \sum_{j=1}^{2^k} \int_{\tilde{I}^j_k} \|\Delta^r_h f(t)\|^q_X \mathrm{d}t\mathrm{d}h \\
&\le \frac{1}{2^{-k}} \int_0^{2^{-k}} \int_{[0,1-rh]} \|\Delta^r_h f(t)\|^q_X \mathrm{d}t\mathrm{d}h \\
&= \frac{1}{2^{-k}} \int_0^{2^{-k}} \|\Delta^r_h f\|^q_{L_q([0,1-rh],X)} \mathrm{d}h = w_r(f, I, 2^{-k})^q_q.
\end{aligned} \tag{40}$$

On the other hand, using (38) in each subinterval I^j_{k+1}, we have

$$\begin{aligned}
\|S_k - S_{k+1}\|^p_{L_p(I,X)} &= \sum_{j=1}^{2^{k+1}} \|S_k - S_{k+1}\|^p_{L_p(I^j_{k+1},X)} \\
&\lesssim 2^{-k\left(1-\frac{p}{q}\right)} \sum_{j=1}^{2^{k+1}} \|S_k - S_{k+1}\|^p_{L_q(I^j_{k+1},X)} \\
&\lesssim 2^{-k\left(1-\frac{p}{q}\right)} \left(\sum_{j=1}^{2^{k+1}} \|S_k - S_{k+1}\|^q_{L_q(I^j_{k+1},X)} \right)^{p/q}
\end{aligned}$$

$$= 2^{-k\left(1-\frac{p}{q}\right)}\|S_k - S_{k+1}\|^p_{L_q(I,X)}, \tag{41}$$

where in the second to last line we used the fact that $\ell_{q/p} \hookrightarrow \ell_1$ for $q \leq p$. This yields for $\overline{p} = \min\{1, p\}$,

$$\|S_k - S_{k+1}\|^{\overline{p}}_{L_p(I,X)} \lesssim 2^{-k\left(\frac{1}{p}-\frac{1}{q}\right)\overline{p}}\|S_k - S_{k+1}\|^{\overline{p}}_{L_q(I,X)}. \tag{42}$$

But then using (40), (42), and the assumption that $f \in L_p(I, X)$, we obtain

$$\begin{aligned}
E_r(f, I)^{\overline{p}}_p \leq \|f - S_0\|^{\overline{p}}_{L_p(I,X)} &\leq \sum_{k=0}^{\infty} \|S_k - S_{k+1}\|^{\overline{p}}_{L_p(I,X)} \\
&\lesssim \sum_{k=0}^{\infty} 2^{-k\left(\frac{1}{p}-\frac{1}{q}\right)\overline{p}}\|S_k - S_{k+1}\|^{\overline{p}}_{L_q(I,X)} \\
&\leq \sum_{k=0}^{\infty} 2^{-k\left(\frac{1}{p}-\frac{1}{q}\right)\overline{p}} \left(\|S_k - f\|^{\overline{p}}_{L_q(I,X)} + \|f - S_{k+1}\|^{\overline{p}}_{L_q(I,X)}\right) \\
&\lesssim \sum_{k=0}^{\infty} 2^{-k\left(\frac{1}{p}-\frac{1}{q}\right)\overline{p}}\|f - S_k\|^{\overline{p}}_{L_q(I,X)} \\
&\lesssim \sum_{k=0}^{\infty} 2^{-k\left(\frac{1}{p}-\frac{1}{q}\right)\overline{p}} w_r(f, I, 2^{-k})^{\overline{p}}_q \\
&= \sum_{k=0}^{\infty} 2^{-k\left(\left(\frac{1}{p}-\frac{1}{q}\right)+s\right)\overline{p}} 2^{ks\overline{p}} w_r(f, I, 2^{-k})^{\overline{p}}_q \\
&\leq \sum_{k=0}^{\infty} 2^{-k\delta\overline{p}} 2^{ks\overline{p}} w_r(f, I, 2^{-k})^{\overline{p}}_q,
\end{aligned} \tag{43}$$

where $\delta := \left(\frac{1}{p} - \frac{1}{q}\right) + s \geq 0$ due to our assumption $s \geq (1/q - 1/p)_+$. In (43) we proceed as follows: if $q < \overline{p}$ we make use of the embedding $\ell_q \hookrightarrow \ell_{\overline{p}}$ together with the fact that $2^{-k\delta\overline{p}} \leq 1$ and for $q > \overline{p}$ we apply Hölder's inequality with $\frac{q}{\overline{p}} > 1$. This finally gives

$$\|f - S_0\|_{L_p(I,X)} \leq \left(\sum_{k=0}^{\infty} 2^{ksq} w_r(f, 2^{-k}, I)^q_q\right)^{1/q} \simeq |f|_{B^s_{q,q}(I,X)}. \tag{44}$$

The assertion thus follows by recalling that $S_0 \in \mathbb{V}^r_{I,X}$. □

As a consequence of the previous theorem we have that under the same assumptions $B^s_{q,q}(I, X)$ is embedded into $L_p(I, X)$.

Corollary 21. *Let $0 < p, q \leq \infty$, $r \in \mathbb{N}$, and $s > 0$. If $(1/q - 1/p)_+ \leq s$ then $B^s_{q,q}(I, X)$ is embedded into $L_p(I, X)$ and there exists a constant $c > 0$ which depends only on p, q, r, and s such that*

$$\|f\|_{L_p(I,X)} \leq c\|f\|_{B^s_{q,q}(I,X)},$$

for all $f \in B^s_{q,q}(I, X)$ and for any finite interval I.

Proof. Let $f \in B^s_{q,q}(I, X) \cap L_p(I, X)$, $r \in \mathbb{N}$, $r > s$, and let S_0 be as in the proof of Theorem 16. Then,

$$\|f\|_{L_p(I,X)} \lesssim \|f - S_0\|_{L_p(I,X)} + \|S_0\|_{L_p(I,X)} \lesssim |f|_{B^s_{q,q}(I,X)} + \|S_0\|_{L_q(I,X)}.$$

Since $S_0 \in \mathbb{V}^r_{I,X}$ was chosen satisfying Jackson estimate (24) with p replaced by q,
$\|S_0\|_{L_q(I,X)} \lesssim \|f - S_0\|_{L_q(I,X)} + \|f\|_{L_q(I,X)} \lesssim w_r(f, I, 1)_q + \|f\|_{L_q(I,X)} \lesssim \|f\|_{L_q(I,X)}$.
Therefore, for all $f \in B^s_{q,q}(I, X) \cap L_p(I, X)$,

$$\|f\|_{L_p(I,X)} \lesssim |f|_{B^s_{q,q}(I,X)} + \|f\|_{L_q(I,X)} \lesssim \|f\|_{B^s_{q,q}(I,X)}.$$

Finally, since $B^s_{q,q}(I, X) \cap L_p(I, X)$ is dense in $B^s_{q,q}(I, X)$ the assertion follows. □

The generalized Whitney's theorem, Theorem 16, is now a consequence of Proposition 20 and Corollary 21.

4. Adaptive approximation in one variable

4.1. **The stationary case.** Given a polyhedral space domain $\Omega \subset \mathbb{R}^n$, $n \in \mathbb{N}$, we let $\mathbb{T}(\mathcal{T}_0)$ denote the set of all triangulations $\mathcal{T}$ (partitions into simplices) that are obtained by successive application of the bisection routine of [Ste08] from a properly labeled initial triangulation $\mathcal{T}_0$ of Ω. If $n = 1$, $\mathbb{T}(\{0 < T\})$ denotes the set of all partitions of $\Omega = [0, T)$ into sub-intervals that may be obtained by successive bisection of $\mathcal{T}_0 = \{[0, T)\}$. For simplicity, the one-dimensional partition $\{[0 = t_0, t_1), [t_1, t_2), \dots, [t_{N-1}, t_N = T)\}$ will be usually denoted by $\{0 = t_0 < t_1 < \cdots < t_N = T\}$. Whenever we write $\mathcal{T}_* = \text{Refine}(\mathcal{T}, \mathcal{M})$, we understand that $\mathcal{M} \subset \mathcal{T}$ and $\mathcal{T}_*$ is the refinement of $\mathcal{T}$ obtained by the bisection routine of [Ste08]. In the one-dimensional case, we understand that $\mathcal{T}_*$ is obtained by the sole replacement in $\mathcal{T}$ of each element $T = [a, b) \in \mathcal{M}$ by its children $[a, \frac{a+b}{2})$, $[\frac{a+b}{2}, b)$.

Therefore, the following complexity bound holds:

> Let $\mathcal{T}_0$, $\mathcal{T}_1$, $\mathcal{T}_2$, $\dots$, be a sequence of partitions in $\mathbb{T}(\mathcal{T}_0)$ obtained by successive calls of $\mathcal{T}_{k+1} = \text{Refine}(\mathcal{T}_k, \mathcal{M}_k)$, with $\mathcal{M}_k \subset \mathcal{T}_k$ the set of *marked* elements. Then, there exists a constant C that depends on the initial triangulation $\mathcal{T}_0$ such that

$$\#\mathcal{T}_k - \#\mathcal{T}_0 \le C \sum_{j=0}^{k-1} \#\mathcal{M}_j, \qquad k = 1, 2, \dots. \tag{45}$$

For $\mathcal{T} \in \mathbb{T}(\mathcal{T}_0)$, recall that $\mathbb{V}^r_{\mathcal{T}}$ is the finite element space of continuous piecewise polynomials of order r, i.e.,

$$\mathbb{V}^r_{\mathcal{T}} := \{g \in C(\overline{\Omega}) : \ g\big|_T \in \Pi^r \text{ for all } T \in \mathcal{T}\},$$

where Π^r denotes the set of polynomials of total degree (strictly) less than r. The underlying domain Ω and its dimension are implicitly indicated by the partition $\mathcal{T}$, which will sometimes correspond to a time interval $[0, T)$ and sometimes to an n-dimensional space domain.

Approximation Classes. Let X be a quasi-Banach space on the polyhedral bounded Lipschitz domain $\Omega \subset \mathbb{R}^n$ with quasi-norm $\|\cdot\|_X$. Let $\mathcal{T}_0$ be a triangulation of Ω, properly labeled so that (45) holds, and assume further that $\mathbb{V}^r_{\mathcal{T}} \subset X$ for $\mathcal{T} \in \mathbb{T}(\mathcal{T}_0)$. In this context, for $f \in X$, the best N-term approximation error is given by

$$\sigma_N(f) = \inf_{|\mathcal{T}| \le N} \inf_{g \in \mathbb{V}^r_{\mathcal{T}}} \|f - g\|_X.$$

For $s > 0$ we define the approximation class $\mathbb{A}_s(X)$ as the set of those functions in X whose best N-term approximation error is of order N^{-s}, i.e.,

$$\mathbb{A}_s(X) := \{f \in X : \ \exists c > 0 \text{ such that } \sigma_N(f) \le cN^{-s}, \ \forall N \in \mathbb{N}\}.$$

Equivalently, we can define $\mathbb{A}_s(X)$ through a semi-quasi-norm as follows:

$$\mathbb{A}_s(X) := \{f \in X : \ |f|_{\mathbb{A}_s(X)} < \infty\} \quad \text{with} \quad |f|_{\mathbb{A}_s(X)} := \sup_{N\in\mathbb{N}} N^s \sigma_N(f).$$

Alternatively, this definition is equivalent to saying that $f \in \mathbb{A}_s(X)$ if there is a constant c such that for all $\varepsilon > 0$, there exists a mesh $\mathcal{T}$ that satisfies

$$\inf_{g\in\mathbb{V}^r_{\mathcal{T}}} \|f-g\|_X \le c\varepsilon \quad \text{and} \quad |\mathcal{T}| \le \varepsilon^{-1/s}, \tag{46}$$

and $|f|_{\mathbb{A}_s(X)}$ is equivalent to the infimum of all constants c that satisfy (46).

We use the following result from [GM14, Thm. 2.2, Cor. 2.3], which is the high-order analog to the one presented in [BDDP02] for linear finite elements ($r = 2$).

Theorem 22. *Let $X = B^\alpha_{p,p}(\Omega)$, $0 < p < \infty$, $0 < \alpha < \min\{r, 1 + \frac{1}{p}\}$ or $X = L_p(\Omega)$ if $\alpha = 0$. If $f \in B^{s+\alpha}_{\tau,\tau}(\Omega)$ with $s > 0$, $0 < \frac{1}{\tau} < \frac{s}{n} + \frac{1}{p}$, and $s + \alpha < r$, then*

$$B^{\alpha+s}_{\tau,\tau}(\Omega) \subset \mathbb{A}_{s/n}(B^\alpha_{p,p}(\Omega)) \qquad (\alpha > 0), \tag{47}$$

$$B^{s}_{\tau,\tau}(\Omega) \subset \mathbb{A}_{s/n}(L_p(\Omega)) \qquad (\alpha = 0). \tag{48}$$

In particular, if $p = 2$ and $\alpha = 0$ we have the following result.

Corollary 23. *Let $X = L_2(\Omega)$, $r \in \mathbb{N}$, $0 < s < r$, and $0 < \frac{1}{\tau} < \frac{s}{n} + \frac{1}{2}$. Then there exists a constant $C = C(r, s, \tau, \Omega, \mathbb{T})$ such that, for every $\varepsilon > 0$ there exists $\mathcal{T} \in \mathbb{T}(\mathcal{T}_0)$ and $g \in \mathbb{V}^r_{\mathcal{T}}$ such that*

$$\|f-g\|_X \le \varepsilon\, |f|_{B^s_{\tau,\tau}(\Omega)} \qquad \text{and} \qquad |\mathcal{T}| \lesssim \varepsilon^{-n/s}.$$

4.2. **Greedy algorithm.** Theorem 22, or equivalently Corollary 23, is proved with the help of a so called *Greedy algorithm*. In order to make this article self-contained, we present it here and use it to build a quasi-optimal partition of $[0, T)$ to approximate a vector-valued function in $L_p([0, T), X)$. This, in turn, is an intermediate tool for constructing the optimal time-space partition.

In the rest of this section we consider the following framework. We let X denote a Banach space, $r \in \mathbb{N}$ denotes the polynomial *order* with respect to time, and for an interval I, recall the definition of $\mathbb{V}^r_{I,X}$ from (21):

$$\mathbb{V}^r_{I,X} := \Big\{P(t) = \sum_{j=0}^{r-1} a_j t^j, \ a_j \in X, \ t \in I\Big\} \subset L_p(I, X),$$

i.e., the tensor product space $\Pi^r \otimes X$ on the time slice $I \times \Omega$. For a partition $\mathcal{T} = \{0 = t_0 < t_1 < \cdots < t_N = T\}$ of the time interval $[0, T)$, we consider the following corresponding (abstract) finite element space:

$$\mathbb{V}^r_{\mathcal{T},X} = \{P \in L_p([0, T), X) : \ P_{|I} \in \mathbb{V}^r_{I,X}, \ I \in \mathcal{T}\}.$$

Recall the definition of the best approximation error $E_r(f, I)_p$ associated with an interval $I \subset [0, T)$, i.e.,

$$E_r(f, I)_p = \inf_{P_I\in\mathbb{V}^r_{I,X}} \|f - P_I\|_{L_p(I,X)}, \tag{49}$$

so that

$$\inf_{g\in\mathbb{V}^r_{\mathcal{T},X}} \|f-g\|_{L_p([0,T),X)} = \left(\sum_{I\in\mathcal{T}} E_r(f,I)_p^p\right)^{1/p}.$$

An algorithm approximating the solution with a parameter $\delta>0$ reads as follows:

Algorithm 1 *Greedy* algorithm

```
function GREEDY(f,δ)
    Let T_0 = {0 < T} = {[0,T)}.
    k = 0
    while M_k := {I ∈ T_k : E_r(f,I)_p > δ} ≠ ∅ do
        Let T_{k+1} = REFINE(T_k, M_k)
        k ← k + 1
    end while
end function
```

4.3. **Semi-discretization in time.** Concerning the error when approximating a vector-valued function with piecewise polynomials with respect to time, we have the following result.

Theorem 24 (**Time discretization**)**.** *Let X be a separable Banach space, let $s>0$, $0<p,q\le\infty$, and $\left(\frac{1}{q}-\frac{1}{p}\right)_+\le s<r$, with $r\in\mathbb{N}$. Then, if $f\in B^s_{q,q}([0,T),X)$ and $\varepsilon>0$, there exists $\delta>0$ such that* GREEDY*(f,δ) terminates in finitely many steps and the generated partition $\mathcal{T}$ satisfies*

$$\#\mathcal{T}\le c_1\,\varepsilon^{-1/s}, \tag{50}$$

where the constant $c_1>0$ depends on p, q, and s but not on f. Moreover, there exists $P\in\mathbb{V}^r_{\mathcal{T},X}$ satisfying

$$\|f-P\|_{L_p([0,T),X)}\le c_2\,\varepsilon\,|f|_{B^s_{q,q}([0,T),X)}\le c_3\,(\#\mathcal{T})^{-s}|f|_{B^s_{q,q}([0,T),X)}, \tag{51}$$

with $c_2,c_3>0$ depending on p, q, and s but not on f.

Proof. Let $\varepsilon>0$ be given and let $\delta=\varepsilon^{\frac{s+1/p}{s}}|f|_{B^s_{q,q}([0,T),X)}$. Using Whitney's estimate (36) we see that the error $E_r(f,I)_p$ associated with an interval I satisfies

$$E_r(f,I)_p=\inf_{P_I\in\mathbb{V}^r_{I,X}}\|f-P_I\|_{L_p(I,X)}\lesssim|I|^{s+\frac{1}{p}-\frac{1}{q}}|f|_{B^s_{q,q}(I,X)}. \tag{52}$$

Since $s+\frac{1}{p}-\frac{1}{q}>0$ the right-hand side goes to zero as $|I|$ goes to zero, which shows that the Greedy algorithm terminates in a finite number of steps K.

We now bound the number of elements of $\mathcal{T}:=\mathcal{T}_K$ as follows. Initially, $\mathcal{T}_0=\{[0,T)\}$, therefore, $\#\mathcal{T}_0=1$. In each iteration of the **while**-loop, $\#\mathcal{M}_k$ elements are *marked* for refinement. If $\overline{\mathcal{M}}=\bigcup_{k=0}^{K-1}\mathcal{M}_k$ is the union of all marked elements in a certain step of the algorithm, then, due to (45), the resulting final partition $\mathcal{T}$ satisfies $\#\mathcal{T}\lesssim 1+\#\overline{\mathcal{M}}\lesssim\#\overline{\mathcal{M}}$. We see that estimating $\#\mathcal{T}$ is comparable with estimating $\#\overline{\mathcal{M}}$. In order to count the number of elements in $\overline{\mathcal{M}}$ observe that $\#\overline{\mathcal{M}}=\sum_{k=0}^{\infty}\#\mathcal{M}^k$, with

$$\mathcal{M}_k=\left\{I\in\overline{\mathcal{M}}:|I|=T/2^k\right\}\quad\text{if}\quad 0\le k\le K-1\qquad\text{and}\qquad\mathcal{M}_k=\emptyset\quad\text{if}\quad k\ge K.$$

On the one hand, since our time interval $[0,T)$ is finite, we obtain the upper bound

$$\#\mathcal{M}_k \le 2^k, \qquad k \in \mathbb{N}_0.$$

On the other hand, if $I \in \mathcal{M}_k$ from steps 4 and 6 of the Greedy algorithm and formula (52), we have

$$\delta < E_r(f,I)_p \lesssim \left(\frac{1}{2^k}\right)^{s+\frac{1}{p}-\frac{1}{q}} |f|_{B^s_{q,q}(I,X)}, \quad \text{so that} \quad \delta^q \lesssim \left(\frac{1}{2^k}\right)^{sq+\frac{q}{p}-1} |f|^q_{B^s_{q,q}(I,X)}.$$

This implies

$$\delta^q \#\mathcal{M}_k = \sum_{I\in\mathcal{M}_k} \delta^q \lesssim \left(\frac{1}{2^k}\right)^{sq+\frac{q}{p}-1} \sum_{I\in\mathcal{M}_k} |f|^q_{B^s_{q,q}(I,X)} \le \left(\frac{1}{2^k}\right)^{sq+\frac{q}{p}-1} |f|^q_{B^s_{q,q}([0,T),X)},$$

i.e.,

$$\#\mathcal{M}_k \lesssim \min\left\{2^k, \frac{1}{\delta^q}\left(\frac{1}{2^k}\right)^{sq+\frac{q}{p}-1} |f|^q_{B^s_{q,q}([0,T),X)}\right\}.$$

The first term corresponds to an increasing geometric series, the second to a decreasing one. Setting $k_0 := \min\left\{k \in \mathbb{N}_0 : \frac{1}{\delta^q}\left(\frac{1}{2^k}\right)^{sq+\frac{q}{p}-1} |f|^q_{B^s_{q,q}([0,T),X)} < 2^k\right\}$ we obtain

$$\begin{aligned}\#\overline{\mathcal{M}} = \sum_{k=0}^{\infty} \#\mathcal{M}_k &\le \sum_{k=0}^{k_0-1} 2^k + \sum_{k=k_0}^{\infty} \frac{1}{\delta^q}\left(\frac{1}{2^k}\right)^{sq+\frac{q}{p}-1} |f|^q_{B^s_{q,q}([0,T),X)} \\ &\lesssim 2^{k_0} + \frac{1}{\delta^q}\left(\frac{1}{2^{k_0}}\right)^{sq+\frac{q}{p}-1} |f|^q_{B^s_{q,q}([0,T),X)} \lesssim 2^{k_0}. \end{aligned} \tag{53}$$

In order to estimate 2^{k_0} we observe that

$$2^{k_0-1} \le \frac{1}{\delta^q}\left(\frac{1}{2^{k_0}}\right)^{sq+\frac{q}{p}-1} |f|^q_{B^s_{q,q}([0,T),X)} < 2^{k_0},$$

$$2^{k_0(s+\frac{1}{p})-\frac{1}{q}} \le \frac{1}{\delta}|f|_{B^s_{q,q}([0,T),X)} < 2^{k_0(s+\frac{1}{p})}.$$

We see that

$$2^{k_0} \le \left(\frac{1}{\delta}\right)^{\frac{1}{s+1/p}} |f|^{\frac{1}{s+1/p}}_{B^s_{q,q}([0,T),X)}, \tag{54}$$

therefore, from (53) and (54) we get

$$\#\mathcal{T} \lesssim \#\overline{\mathcal{M}} \lesssim \left(\frac{1}{\delta}\right)^{\frac{1}{s+1/p}} |f|^{\frac{1}{s+1/p}}_{B^s_{q,q}([0,T),X)}, \quad \text{i.e.,} \quad \delta \lesssim (\#\mathcal{T})^{-(s+\frac{1}{p})}|f|_{B^s_{q,q}([0,T),X)},$$

and (50) follows after recalling that $\delta = \varepsilon^{\frac{s+1/p}{s}}|f|_{B^s_{q,q}([0,T),X)}$.

Finally, for each $I \in \mathcal{T}$ we let $P_I \in \mathbb{V}^r_{I,X}$ satisfy $\|f - P_I\|_{L_p(I,X)} \le 2E_r(f,I)_p$ and let $P(t) = \sum_{I\in\mathcal{T}} \chi_I(t)P_I(t)$, $t \in [0,T)$. Hence,

$$P \in \mathbb{V}^r_{\mathcal{T},X} \quad \text{and} \quad \|f-P\|^p_{L_p([0,T),X)} \lesssim \delta^p \#\mathcal{T} \lesssim \#\mathcal{T}^{-ps}|f|^p_{B^s_{q,q}([0,T),X)},$$

and (51) follows. $\square$

5. Discretization in time and space

We now consider the error when approximating a function with piecewise polynomials with respect to time and space. In this article, we deal with the approximation in $L_2([0,T)\times\Omega) = L_2([0,T),X)$, where hereafter we let $X = L_2(\Omega)$. We restrict ourselves to this Hilbertian case in order to avoid additional technical difficulties and leave the study of more general quasi-norms, e.g. $p\neq 2$ and $X \neq L_2(\Omega)$, to a forthcoming article.

5.1. **Time marching fully discrete adaptivity.** Recall that the type of discretizations that we consider are those consisting of a partition $\{0 = t_0 < t_1 < \cdots < t_N = T\}$ of the time interval and a sequence of partitions $\mathcal{T}_1,\dots,\mathcal{T}_N \in \mathbb{T}$ of the space domain Ω, where $\mathcal{T}_i$ corresponds to the subinterval $[t_{i-1},t_i)$, $i = 1,\dots N$. The time-space partition is then given by

$$\mathcal{P} = \left(\{0 = t_0 < t_1 < \ldots < t_N = T\}, \{\mathcal{T}_1,\dots,\mathcal{T}_N\}\right), \quad \text{with } \#\mathcal{P} = \sum_{i=1}^{N} \#\mathcal{T}_i,$$

Given $r_1, r_2 \in \mathbb{N}$, the finite element space $\overline{\mathbb{V}}_{\mathcal{P}}^{r_1,r_2}$ subject to such a partition $\mathcal{P}$ is defined as

$$\overline{\mathbb{V}}_{\mathcal{P}}^{r_1,r_2} := \{G : [0,T)\times\Omega \to \mathbb{R} : G_{\big|[t_{i-1},t_i)\times\Omega} \in \Pi^{r_1}\otimes \mathbb{V}_{\mathcal{T}_i}^{r_2}, \text{ for all } i = 1,2,\dots,N\},$$

i.e., $G \in \overline{\mathbb{V}}_{\mathcal{P}}^{r_1,r_2}$ if and only if $G(t,\cdot) \in \mathbb{V}_{\mathcal{T}_i}^{r_2}$ for all $t \in [t_{i-1},t_i)$ and $G(\cdot,x)\big|_{[t_{i-1},t_i)} \in \Pi^{r_1}$ for all $x\in\Omega$ and all $i = 1,2,\dots,N$.

In order to construct an optimal approximate solution with tolerance $\varepsilon > 0$ we use the one-dimensional Greedy algorithm as described on page 24 for the (adaptive) discretization in time and an n-dimensional Greedy algorithm for (adaptive) discretizations in space. This allows us to use the results from Theorems 24 and 22, respectively. In particular, we obtain the following result.

Theorem 25 (**Approximation with fully discrete functions).** *Let* $0 < s_i < r_i$, $i = 1, 2$, $0 < q_1 \le \infty$, $1 \le q_2 \le \infty$ *with* $s_1 > \left(\frac{1}{q_1} - \frac{1}{2}\right)_+$ *and* $s_2 > n\left(\frac{1}{q_2} - \frac{1}{2}\right)_+$. *Let* $f \in B^{s_1}_{q_1,q_1}([0,T),X) \cap L_2([0,T), B^{s_2}_{q_2,q_2}(\Omega))$, *with* $X = L_2(\Omega)$. *Then, for each* $\varepsilon > 0$ *there exists a time-space partition* $\mathcal{P}$ *that satisfies*

$$\#\mathcal{P} \le c_1 \varepsilon^{-\left(\frac{1}{s_1}+\frac{n}{s_2}\right)}$$

and a function $F \in \overline{\mathbb{V}}_{\mathcal{P}}^{r_1,r_2}$ *such that*

$$\|f - F\|_{L_2([0,T),X)} \le c_2\varepsilon |\!|\!| f |\!|\!| \le c_3 (\#\mathcal{P})^{-\frac{1}{\frac{1}{s_1}+\frac{n}{s_2}}} |\!|\!| f |\!|\!|,$$

where $|\!|\!| f |\!|\!| = |f|_{B^{s_1}_{q_1,q_1}([0,T),X)} + \|f\|_{L_2([0,T),B^{s_2}_{q_2,q_2}(\Omega))}$ *and the positive constants* c_1, c_2, c_3 *depend on* q_1, q_2, *and* s_1, s_2 *but not on* f.

Remark 26. Here (with a little abuse) we use the notation

$$L_2(I, B^s_{q,q}(\Omega)) = \left\{ f : I \to B^s_{q,q}(\Omega) : \|f\|_{L_2(I,B^s_{q,q}(\Omega))} < \infty \right\}$$

with $\|f\|_{L_2(I,B^s_{q,q}(\Omega))} := \left(\int_I \|f(t)\|^2_{B^s_{q,q}(\Omega)} \mathrm{d}t\right)^{1/2}$.

The restriction $q_2 \ge 1$ in Theorem 25 can probably be removed and replaced by $q_2 > 0$. It appears here due to the fact that we require in the proof below a uniform bound of the approximants on a subinterval $I = [t_{i-1}, t_i)$, which is established in

Lemma 27. Our current proof of Lemma 27 uses Minkowski's inequality which only works if $q_2 \geq 1$. So far we were not able to find an appropriate modification for $q_2 < 1$.

Proof of Theorem 25. Given $f \in B^{s_1}_{q_1,q_1}([0,T), X) \cap L_2([0,T), B^{s_2}_{q_2,q_2}(\Omega))$ and $\varepsilon > 0$, the approximant $F \in \overline{\mathbb{V}}^{r_1,r_2}_{\mathcal{P}}$ is constructed in two steps as follows.

We first use a one-dimensional Greedy algorithm and apply the results from Theorem 24. This gives a partition of the time interval $0 = t_0 < t_1 < \ldots < t_N = T$ and an approximant $G = \sum_{i=1}^N \chi_{[t_{i-1},t_i)} G_i \in \mathbb{V}^{r_1}_{\{0<t_1<\cdots<T\},X}$ with G_i the $L_2([t_{i-1},t_i), X)$ projection of $f_{|[t_{i-1},t_i)}$ into $\mathbb{V}^{r_1}_{[t_{i-1},t_i),X}$. This partition and approximant satisfy

$$N \lesssim \varepsilon^{-1/s_1} \quad \text{and} \quad \|f - G\|_{L_2([0,T),X)} \lesssim \varepsilon |f|_{B^{s_1}_{q_1,q_1}([0,T),X)}.$$

Also, if $\{W_i^j\}_{j=1}^{r_1}$ is an orthonormal basis of $\mathbb{V}^{r_1}_{[t_{i-1},t_i),\mathbb{R}}$ then

$$G_i(t) = \sum_{j=1}^{r_1} G_i^j \, W_i^j(t), \quad \text{with} \quad G_i^j = \int_I f(t) W_i^j(t) \, \mathrm{d}t,$$

noting that the last integral is a Bochner integral in $X = L_2(\Omega)$.

We now observe that due to Lemma 27 below we have $G_i^j \in B^{s_2}_{q_2,q_2}(\Omega)$ and

$$\|G_i\|_{L_2([t_{i-1},t_i),B^{s_2}_{q_2,q_2}(\Omega))} \lesssim \|f\|_{L_2([t_{i-1},t_i),B^{s_2}_{q_2,q_2}(\Omega))}. \tag{55}$$

The second step consists in approximating each function $G_i^j \in B^{s_2}_{q_2,q_2}(\Omega)$ using the space-adaptive Greedy algorithm. Resorting to Corollary 23 we find a mesh $\mathcal{T}_i^j \in \mathbb{T}(\mathcal{T}_0)$ and a finite element function $F_i^j \in \mathbb{V}^{r_2}_{\mathcal{T}_i^j}$ with

$$\#\mathcal{T}_i^j \lesssim \varepsilon^{-\frac{n}{s_2}} \quad \text{and} \quad \|G_i^j - F_i^j\|_X \lesssim \varepsilon |G_i^j|_{B^{s_2}_{q_2,q_2}(\Omega)}.$$

Therefore, after defining $\mathcal{T}_i = \oplus_{j=1}^{r_1} \mathcal{T}_i^j$ (the overlay of the meshes [CKNS08]), we have that $F_i(t) := \sum_{j=1}^{r_1} W_i^j(t) F_i^j \in \mathbb{V}^{r_1}_{[t_{i-1},t_i),\mathbb{V}^{r_2}_{\mathcal{T}_i}}$ satisfies

$$\#\mathcal{T}_i \leq \sum_{j=1}^{r_1} \#\mathcal{T}_i^j \lesssim \varepsilon^{-\frac{n}{s_2}} \quad \text{[CKNS08, Lem. 3.7] and}$$

$$\|F_i - G_i\|_{L_2([t_{i-1},t_i),X)} \lesssim \varepsilon \|G_i\|_{L_2([t_{i-1},t_i),B^{s_2}_{q_2,q_2}(\Omega))} \lesssim \varepsilon \|f\|_{L_2([t_{i-1},t_i),B^{s_2}_{q_2,q_2}(\Omega))},$$

due to (55).

Finally, we let $\mathcal{P} = \{\{0 = t_0 < t_1 < \cdots < t_N = T\}, \{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_N\}\}$ and define $F = \sum_{i=1}^N \chi_{[t_{i-1},t_i)} F_i \in \overline{\mathbb{V}}^{r_1,r_2}_{\mathcal{P}}$, whence by the triangle inequality

$$\begin{aligned}\|f - F\|_{L_2([0,T),X)} &\leq \|f - G\|_{L_2([0,T),X)} + \|G - F\|_{L_2([0,T),X)} \\ &\lesssim \varepsilon |f|_{B^{s_1}_{q_1,q_1}([0,T),X)} + \left(\sum_{i=1}^N \|G_i - F_i\|^2_{L_2([t_{i-1},t_i),X)} \right)^{1/2} \\ &\lesssim \varepsilon \Big(|f|_{B^{s_1}_{q_1,q_1}([0,T),X)} + \|f\|_{L_2([t_{i-1},t_i),B^{s_2}_{q_2,q_2}(\Omega))} \Big)\end{aligned}$$

and

$$\#\mathcal{P} = \sum_{i=1}^N \#\mathcal{T}_i \lesssim N \, \varepsilon^{-\frac{n}{s_2}} \lesssim \varepsilon^{-\frac{1}{s_1}} \, \varepsilon^{-\frac{n}{s_2}} = \varepsilon^{-(\frac{1}{s_1} + \frac{n}{s_2})}.$$

The assertion of the theorem thus follows. □

In Theorem 25, formula (55), we required a uniform bound of the approximants G_i on a subinterval $I = [t_{i-1}, t_i)$, which is provided by the following lemma.

Lemma 27. *Given a finite interval I, let $r = r_1$, $s = s_2$, and $q = q_2$ satisfy the assumptions from Theorem 25 and assume $f \in L_2(I, B^s_{q,q}(\Omega))$. If $G \in \mathbb{V}^r_{I,X}$ is the $L_2(I, X)$ projection of $f \in L_2(I, B^s_{q,q}(\Omega))$, then*

$$\|G\|_{L_2(I,B^s_{q,q}(\Omega))} \lesssim \|f\|_{L_2(I,B^s_{q,q}(\Omega))}.$$

Proof. If $\{W^j\}_{j=1}^r$ is an orthonormal basis of $\mathbb{V}^r_{I,\mathbb{R}}$ then

$$G(t) = \sum_{j=1}^r G^j W^j(t) \quad \text{with} \quad G^j = \int_I f(t) W^j(t) \mathrm{d}t,$$

i.e.,

$$G(t)(x) = \sum_{j=1}^r G^j(x) W^j(t) \quad \text{with} \quad G^j(x) = \int_I f(t,x) W^j(t) \mathrm{d}t,$$

for almost every $x \in \Omega$. Notice first that

$$\begin{aligned}
\|G\|^2_{L_2(I,B^s_{q,q}(\Omega))} &= \int_I \|G(t)\|^2_{L_q(\Omega)} + |G|^2_{B^s_{q,q}(\Omega)} \mathrm{d}t \\
&= \int_I \Big\| \sum_{j=1}^r G^j W^j(t) \Big\|^2_{L_q(\Omega)} + \Big| \sum_{j=1}^r G^j W^j(t) \Big|^2_{B^s_{q,q}(\Omega)} \mathrm{d}t \\
&\lesssim \sum_{j=1}^r \|G^j\|^2_{L_q(\Omega)} + |G^j|^2_{B^s_{q,q}(\Omega)},
\end{aligned}$$

so that

$$\|G\|_{L_2(I,B^s_{q,q}(\Omega))} \lesssim \sum_{j=1}^r \|G^j\|_{L_q(\Omega)} + |G^j|_{B^s_{q,q}(\Omega)}. \tag{56}$$

We now bound $\|G^j\|_{L_q(\Omega)}$ and $|G^j|_{B^s_{q,q}(\Omega)}$ and focus on the case $1 \le q < \infty$, noting that the case $q = \infty$ is analogous. Since $q \ge 1$, by Minkowski's inequality, for any $j = 1, 2, \dots, r$ we have

$$\begin{aligned}
\|G^j\|_{L_q(\Omega)} &= \left(\int_\Omega \left| \int_I f(x,t) W^j(t) \mathrm{d}t \right|^q \mathrm{d}x \right)^{1/q} \\
&\le \int_I \left(\int_\Omega \left| f(x,t) W^j(t) \right|^q \mathrm{d}x \right)^{1/q} \mathrm{d}t \\
&= \int_I \left| W^j(t) \right| \|f(\cdot,t)\|_{L_q(\Omega)} \mathrm{d}t \\
&\le \left\| W^j(t) \right\|_{L_2(I)} \Big\| \|f(\cdot,t)\|_{L_q(\Omega)} \Big\|_{L_2(I)}
\end{aligned}$$

so that

$$\|G^j\|_{L_q(\Omega)} \lesssim \|f\|_{L_2(I,L_q(\Omega))}, \quad j = 1, 2, \dots, r. \tag{57}$$

We now deal with $\left|G^j\right|_{B^s_{q,q}(\Omega)}$. Observe that for any j we have

$$\left|G^j\right|_{B^s_{q,q}(\Omega)} \lesssim \left(\int_0^1 \left[u^{-s} w_r(G^j, I, u)_q \right]^q \frac{\mathrm{d}u}{u} \right)^{1/q} = \left(\int_0^1 u^{-sq} w_r(G^j, I, u)^q_q \frac{\mathrm{d}u}{u} \right)^{1/q}$$

$$= \left(\int_0^1 u^{-sq} \frac{1}{(2u)^n} \int_{|h|\le u} \left\| \Delta_h^r G^j \right\|^q_{L_q(\Omega_{rh})} \mathrm{d}h \frac{\mathrm{d}u}{u} \right)^{1/q}$$

$$= \left(\int_0^1 u^{-sq} \frac{1}{(2u)^n} \int_{|h|\le u} \int_{\Omega_{rh}} \left| \int_I \Delta_h^r f(t,x) W^j(t) \mathrm{d}t \right|^q \mathrm{d}x \mathrm{d}h \frac{\mathrm{d}u}{u} \right)^{1/q}$$

$$= \left(\int_0^1 \int_{|h|\le u} \int_{\Omega_{rh}} u^{-sq} \frac{1}{(2u)^n} \left| \int_I \Delta_h^r f(t,x) W^j(t) \mathrm{d}t \right|^q \mathrm{d}x \mathrm{d}h \frac{\mathrm{d}u}{u} \right)^{1/q}.$$

Again, by Minkowski's inequality

$$\left|G^j\right|_{B^s_{q,q}(\Omega)} \le \int_I \left(\int_0^1 \int_{|h|\le u} \int_{\Omega_{rh}} u^{-sq} \frac{1}{(2u)^n} \left| \Delta_h^r f(t,x) W^j(t) \right|^q \mathrm{d}x \mathrm{d}h \frac{\mathrm{d}u}{u} \right)^{1/q} \mathrm{d}t$$

$$= \int_I \left|W^j(t)\right| \left(\int_0^1 u^{-sq} \frac{1}{(2u)^n} \int_{|h|\le u} \int_{\Omega_{rh}} \left| \Delta_h^r f(t,x) \right|^q \mathrm{d}x \mathrm{d}h \frac{\mathrm{d}u}{u} \right)^{1/q} \mathrm{d}t$$

$$= \int_I \left|W^j(t)\right| |f(t)|_{B^s_{q,q}(\Omega)} \,\mathrm{d}t \le \left\|W^j\right\|_{L_2(I)} \|f(t)\|_{L_2(I,B^s_{q,q}(\Omega))}$$

whence

$$\left|G^j\right|_{B^s_{q,q}(\Omega)} \lesssim \|f\|_{L_2(I,B^s_{q,q}(\Omega))}. \tag{58}$$

Therefore from (57) (58) and (56) we get

$$\|G\|_{L_2(I,B^s_{q,q}(\Omega))} \lesssim \|f\|_{L_2(I,B^s_{q,q}(\Omega))} \qquad \text{if } \ 1 \le q < \infty,$$

and analogously for $q = \infty$. □

If we use the same polynomial degree in space and time in Theorem 25 the result reads as follows.

Corollary 28 (**Fully discrete with same polynomial degree**). *Let* $1 \le q \le \infty$ *and* $n\left(\frac{1}{q} - \frac{1}{2}\right)_+ < s < r \in \mathbb{N}$. *If* $f \in B^s_{q,q}([0,T), X) \cap L_2([0,T), B^s_{q,q}(\Omega))$ *with* $X = L_2(\Omega)$, *then for each* $\varepsilon > 0$ *there exists a time-space partition* $\mathcal{P}$ *that satisfies*

$$\#\mathcal{P} \le c_1 \varepsilon^{-\frac{n+1}{s}}$$

and a function $F \in \overline{\mathbb{V}}^{r,r}_{\mathcal{P}}$ *such that*

$$\|f - F\|_{L_2([0,T),X)} \le c_2 \varepsilon |||f||| \le c_3 (\#\mathcal{P})^{-\frac{s}{n+1}} |||f|||,$$

where $|||f||| = |f|_{B^s_{q,q}([0,T),X)} + \|f\|_{L_2([0,T),B^s_{q,q}(\Omega))}$ *and the positive constants* c_1, c_2, c_3 *depend on* q *and* s *but not on* f.

5.2. **Comparison with space-time finite elements.** If we were to use space-time finite elements of order r in $\mathbb{R}^{n+1}$, in order to obtain the same rate $(\#\mathcal{P})^{-\frac{s}{n+1}}$ as that indicated in Corollary 28, Corollary 23 tells us that the function f should belong to $B^s_{q,q}([0,T) \times \Omega)$ with $0 < s < r$ and $0 < \frac{1}{q} < \frac{s}{n+1} + \frac{1}{2}$. This raises the following question:

> What is the relation between the spaces
>
> $B^s_{q,q}([0,T) \times \Omega)$ and $B^s_{q_1,q_1}([0,T), L_2(\Omega)) \cap L_2([0,T), B^s_{q_2,q_2}(\Omega))$
>
> for the respective ranges of the parameters q_1, q_2, and q?

The following proposition provides a first attempt to give an answer to this question.

Proposition 29. *Let $0 < s < r$ and $0 < q_1, q_2, q < \infty$, where we additionally require that*

$$\frac{1}{q} < \frac{s}{n+1} + \frac{1}{2}, \qquad \frac{1}{q_1} < s + \frac{1}{2}, \qquad \text{and} \qquad \frac{1}{q_2} < \frac{s}{n} + \frac{1}{2}. \tag{59}$$

Then we have

$$\bigcup_{q_1,q_2} B^s_{q_1,q_1}([0,T), L_2(\Omega)) \cap L_2([0,T), B^s_{q_2,q_2}(\Omega)) \not\subset \bigcup_q B^s_{q,q}([0,T) \times \Omega), \tag{60}$$

where the union is taken over all q, q_1, q_2 according to (59).

Proof. We show that we can find functions belonging to $\bigcup_{q_1,q_2} B^s_{q_1,q_1}([0,T), L_2(\Omega)) \cap L_2([0,T), B^s_{q_2,q_2}(\Omega))$ which are not in $\bigcup_q B^s_{q,q}([0,T) \times \Omega)$. For this let us choose q_1 such that

$$\frac{1}{q} < \frac{s}{n+1} + \frac{1}{2} < \frac{1}{q_1} < s + \frac{1}{2}$$

and consider a function f which is constant with respect to the space variable x and belongs to $B^s_{q_1,q_1}([0,T))$. Clearly, by our assumptions this function is also in $L_2([0,T))$. Moreover, by or choice of q_1 we see from [HS13, Cor. 3.7] that

$$B^s_{q_1,q_1}([0,T), L_2(\Omega)) \not\hookrightarrow B^s_{q,q}([0,T) \times \Omega),$$

since $q_1 < q$, which proves the claim. Alternatively, we could choose q_2 such that

$$\frac{1}{q} < \frac{s}{n+1} + \frac{1}{2} < \frac{1}{q_2} < \frac{s}{n} + \frac{1}{2}$$

and consider a function f which is constant with respect to the time variable t and belongs to $B^s_{q_2,q_2}(\Omega)$. Then, clearly this function also belongs to $L_2(\Omega)$. By our choice of q_2 it again follows from [HS13, Cor. 3.7] that

$$B^s_{q_2,q_2}(\Omega) \not\hookrightarrow B^s_{q,q}(\Omega),$$

since $q_2 < q$. This completes the proof. ☐

Remark 30. We believe that for the above spaces under consideration we actually have the following inclusion

$$\bigcup_q B^s_{q,q}([0,T) \times \Omega) \subsetneq \bigcup_{q_1,q_2} B^s_{q_1,q_1}([0,T), L_2(\Omega)) \cap L_2([0,T), B^s_{q_2,q_2}(\Omega)),$$

where q, q_1, and q_2 are chosen according to (59). This can be interpreted in the sense that the respective solution spaces for time-stepping algorithms yielding the approximation class $\mathbb{A}_{\frac{s}{n+1}}(L_2([0,T) \times \Omega))$ are in fact larger than the corresponding solution spaces for space-time finite elements.
In this context the fact that $B^s_{q,q}([0,T) \times \Omega) \subset L_2([0,T), B^s_{q_2,q_2}(\Omega))$ should be easier to handle. However, these matters are quite technical and, therefore, this interesting question will be tackled in a future paper.

References

[BS88] C. Bennett and R. Sharpley. *Interpolation of operators.* Academic Press, Boston, 1988.

[BDDP02] P. Binev, W. Dahmen, R. DeVore, and P. Petrushev, Approximation classes for adaptive methods, *Serdica Math. J.*, 28:391–416, 2002.

[CKNS08] J.M. Cascón, Ch. Kreuzer, R.H. Nochetto, and K. Siebert. Quasi-optimal convergence rate for an adaptive finite element method. *SIAM J. Numer. Anal.*, 46:2524–2550, 2008.

[DeV98] R.A. DeVore. Nonlinear approximation. *Acta Numer.*, 7:51-150, 1998.

[DL93] R.A. DeVore and G.G. Lorentz. *Constructive approximation.* Grundlehren der Mathematischen Wissenschaften 303, Springer, Berlin, 1993.

[DVP88] R.A. DeVore and V.A. Popov. Interpolation of Besov spaces. *Trans. Amer. Math. Soc.*, 305:397–414, 1988.

[DF90] T.D. Donchev and E.M. Farkhi. Moduli of smoothness of vector-valued functions of a real variable and applications. *Numer. Funct. Anal. and Optimiz.*, 11(5&6):497–509, 1990.

[GM14] F.D. Gaspoz and P. Morin. Approximation classes for adaptive higher order finite element approximation. *Math. Comp.*, 83(289):2127–2160, 2014.

[HS09] D.D. Haroske and C. Schneider. Besov spaces with positive smoothness on $\mathbb{R}^n$, embeddings and growth envelopes. *J. Approx. Theory*, 161:723–747, 2009.

[HS13] D.D. Haroske and L. Skrzypczak. Embeddings of Besov-Morrey spaces on bounded domains. *Studia Math.*, 218:119–144, 2013.

[OS78] P. Oswald and E.A. Storozhenko. Jackson's Theorem in the spaces $L^p(\mathbb{R}^k)$, $0 < p < 1$. *Sibirskii Mathematicheskii Zhurnal*, 19(4):888–901, 1978 (translated English version).

[PP87] P.P. Petrushev and V.A. Popov. *Rational approximation of real functions.* Encyclopedia of Mathematics and its Applications 28, Cambridge University Press, Cambridge, 1987.

[Ste08] R. Stevenson. The completion of locally refined simplicial partitions created by bisection. *Math. Comp.*, 77(261):227–241, 2008.

[Sto77] E.A. Storozhenko. On the approximation by algebraic polynomials of functions of class L^p, $0 < p < 1$. *Math. USSR Izvestija*, 11(3):613–623, 1977 (translated English version).

[Tim63] A.F. Timan. *Theory of approximation of functions of a real variable.* Fizmatgiz, Moscow, 1960; English translation, Macmillan, NY, 1963.

[Tri83] H. Triebel. *Theory of Function spaces.* Monographs in Mathematics 78, Birkhäuser Verlag, Basel, 1983.

Marcelo Actis, Universidad Nacional del Litoral and CONICET, Departamento de Matemática, Facultad de Ingeniería Química, S3000AOM Santa Fe, Argentina
Phone: (+54) 342 457 1164
E-mail: `mactis@fiq.unl.edu.ar`

Pedro Morin, Universidad Nacional del Litoral and CONICET, Departamento de Matemática, Facultad de Ingeniería Química, S3000AOM Santa Fe, Argentina
Phone: (+54) 342 457 1164
E-mail: `pmorin@fiq.unl.edu.ar`

Cornelia Schneider, Friedrich-Alexander-Universität Erlangen-Nürnberg, Lehrstuhl AM3, Cauerstr. 11, 91058 Erlangen, Germany
Phone: (+49) 9131 85-67207, Fax: (+49) 9131 85-67201
E-mail: `schneider@math.fau.de`