INDUSTRIAL
MATHEMATICS
INSTITUTE

2000:11

Nonseparable Walsh type functions on $\mathbf{R}^d$

M. Nielsen

IMI
Preprint Series

Department of Mathematics
University of South Carolina

NONSEPARABLE WALSH TYPE FUNCTIONS ON $\mathbb{R}^d$

M. NIELSEN

Abstract. We introduce wavelet packets in the setting of a multiresolution analysis of $L^2(\mathbb{R}^d)$ generated by an arbitrary dilation matrix A satisfying $|\det A| = 2$, and note that all the analysis and algorithms of wavelet packets in the standard one dimensional case can be generalized to this multidimensional setting.

Then we consider the wavelet packets associated with a multiresolution analysis with scaling function given by the characteristic function of some set (called a tile) in $\mathbb{R}^d$. We call such wavelet packets generalized Walsh functions, and prove that the new functions share two major convergence properties with the Walsh system defined on $[0, 1)$. The functions constitute a Schauder basis for $L^p(\mathbb{R}^d)$, $1 < p < \infty$, and the expansion of L^p-functions converge pointwise almost everywhere. We also introduce some compactly supported wavelet packets in $\mathbb{R}^2$ of class $C^r(\mathbb{R}^2)$, $1 \leq r < \infty$, modeled after the generalized Walsh function and having the same convergence properties.

Introduction

Wavelet analysis was originally introduced in order to improve seismic signal processing by switching from short-time Fourier analysis to new algorithms better suited to detect and analyze abrupt changes in signals. It corresponds to a decomposition of phase space in which the trade-off between time and frequency localization has been chosen to provide better and better time localization at high frequencies in return for poor frequency localization. This makes the analysis well adapted to the study of transient phenomena and has proven a very successful approach to many problems in signal processing, numerical analysis, and quantum mechanics. Nevertheless, for stationary signals wavelet analysis is outperformed by short-time Fourier analysis. Wavelet packets were introduced by R. Coifman, Y. Meyer, and M. V. Wickerhauser ([3]) to improve the poor frequency localization of wavelet bases at high frequencies and thereby provide a more efficient decomposition of signals containing both transient and stationary components.

So far most work on wavelet packets has been done in one dimension or using separable wavelet packets in higher dimensions (i.e. tensor products of one dimensional wavelet packets). However, separable wavelet and wavelet packet bases both have several drawbacks for the application to fields like image analysis since they impose an unavoidable line structure on the plane. For example, the zero set of a separable wavelet packet at high frequencies will contain a large number (same order of magnitude as the frequency) of horizontal and vertical lines that may create artifacts in the reconstructed image. Another potential problem is in the Fourier domain where separable two-dimensional wavelet packets have four characteristic peaks making it hard to selectively localize a unique frequency. R. Coifman and F. Meyer introduced the so-called Brushlets in [8] to remove the "uncertainty" in frequency localization, however the Brushlets are essentially Fourier transforms of smooth local trigonometric bases and are therefore no longer functions associated with a multiresolution structure.

The aim of this paper is twofold. In Section 1 we introduce wavelet packets associated with the class of multiresolution analyses of $\mathbb{R}^d$ for which there are associated wavelet bases generated by only one wavelet. Such functions provide the same large number of orthonormal bases as

RESEARCH SUPPORTED BY NSF GRANT NO. DMS-9872890.

wavelet packets in one-dimension do and should provide a good platform for doing image analysis using the well known "best basis" algorithm of Coifman and Wickerhauser. Moreover, since the functions are nonseparable there is a possibility to design the wavelet packets in such a way as to avoid some of the line artifact associated with product systems mentioned above. We will investigate the usefulness of the nonseparable wavelet packets for the application to image analysis in a forthcoming paper.

Then we consider one special case of the wavelet packet construction in detail. This special case can be considered the multidimensional generalization of the Walsh system on $[0, 1)$. We prove that this multidimensional generalization share the two most important convergence properties of the classical Walsh system; the new system is a Schauder basis for $L^p(\mathbb{R}^d)$, $1 < p < \infty$, and the expansion of every L^p-function in the system converges pointwise a.e. This will be done in Section 3. In section 2 we present some facts about generalized Haar multiresolution analyses associated with a dilation matrix with determinant ± 2 that we need for section 3.

The generalized Walsh functions are not continuous and they are generally supported on irregular fractal like sets, which may limits their usefulness for image analysis. We consider a class of smoother wavelet packets with the same convergence properties as the generalized Walsh Functions in Section 4. Similar types of functions in a one dimensional setting were considered in [12].

1. Non-stationary Wavelet Packets For A General Dilation Matrix

We begin by recalling some facts about multiresolution analyses associated with a general dilation matrix that we will use later in this section to define the wavelet packets we have in mind. The reader can find a more extensive discussion of the topic in [19].

Let A be a $(d \times d)$-matrix such that $A : \mathbb{Z}^d \to \mathbb{Z}^d$. If the eigenvalues of A all have absolute value strictly greater than 1 then we call A a dilation matrix. We can define a multiresolution analysis associated with such A:

Definition 1.1. A multiresolution associated with a dilation matrix A is a sequence of closed subspaces $(V_j)_{j\in\mathbb{Z}}$ of $L^2(\mathbb{R}^d)$ satisfying

(i) $V_j \subset V_{j+1}, \quad \forall j \in \mathbb{Z}$,
(ii) $\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R}^d) \quad \text{and} \quad \bigcap_{j\in\mathbb{Z}} V_j = \{0\}$,
(iii) $f \in V_j \Leftrightarrow f(Ax) \in V_{j+1}, \quad \forall j \in \mathbb{Z}$,
(iv) there exists a function $\phi \in V_0$ called a scaling function such that the system $\{\phi(\cdot - \gamma)\}_{\gamma\in\mathbb{Z}^d}$ is an orthonormal basis for V_0.

The wavelet spaces W_j associated with such a multiresolution analysis are given by $W_j = V_{j+1} \cap V_j^{\perp}$, and one can easily check that $f \in W_j \Leftrightarrow f(A\cdot) \in W_{j+1}$ and $L^2(\mathbb{R}^d) = \bigoplus_{j\in\mathbb{Z}} W_j$. A family of wavelets associated with the multiresolution analysis is a collection of s functions $\{\Psi^r\}_{r=1}^s$ for which $\{\Psi^r(\cdot - \gamma)\}_{r=1,\gamma\in\mathbb{Z}^d}^s$ is an orthonormal basis for W_0. Suppose $|\det A| = q$. It turns out that the number s of wavelets needed to generate such a basis for W_0 is exactly $q - 1$. This makes the case $|\det A| = 2$ especially interesting since the wavelet basis is generated by only one function just as in the one-dimensional case. We will use the notation P_{V_j} and P_{W_j} to denote the orthogonal projections onto the closed spaces V_j and W_j, respectively. One can show that P_{V_j} and P_{W_j} extend to bounded operators on $L^p(\mathbb{R}^d)$, $1 < p < \infty$, provided that the scaling function has a minimum of decay at infinity, see eg. [19].

Let $\{V_j\}_{j\in\mathbb{Z}}$ be a multiresolution analysis of $L^2(\mathbb{R}^d)$ associated with a dilation matrix A satisfying $|\det A| = 2$. Suppose (Φ, Ψ) is an associated scaling function/wavelet pair. Then there exist $2\pi\mathbb{Z}^d$-periodic functions m_0 and m_1 such that

$$\hat{\Phi}(\xi) = m_0(D\xi)\hat{\Phi}(D\xi)$$
$$\hat{\Psi}(\xi) = m_1(D\xi)\hat{\Phi}(D\xi),$$

with $D = (A^*)^{-1}$. Since $|\det A| = 2$ we can find $\Gamma \in \mathbb{Z}^d$ such that $\Gamma + \mathbb{Z}^d/A^*\mathbb{Z}^d$ satisfies $\mathbb{Z}^d = A^*\mathbb{Z}^d \cup (\Gamma + A^*\mathbb{Z}^d)$. Then it is easy to check that the matrix

$$\begin{bmatrix} m_0(\xi) & m_0(\xi + 2\pi D\Gamma) \\ m_1(\xi) & m_1(\xi + 2\pi D\Gamma) \end{bmatrix}$$

is unitary for a.a. $\xi \in \mathbb{R}^d$. This observation leads to the following definition. We let A and Γ be related as above.

Definition 1.2. Let m_0 and m_1 be $2\pi\mathbb{Z}^d$ periodic functions for which

$$\begin{bmatrix} m_0(\xi) & m_0(\xi + 2\pi D\Gamma) \\ m_1(\xi) & m_1(\xi + 2\pi D\Gamma) \end{bmatrix}$$

is unitary a.e. Then we call (m_0, m_1) a pair of orthogonal quadrature filters associated with (A, Γ).

We can now define the natural generalization of wavelet packets to the setting of a multiresolution analysis associated with a dilation matrix A with $|\det A| = 2$.

Definition 1.3. Let $\{(m_0^{(p)}, m_1^{(p)})\}_{p=1}^{\infty}$ be a sequence of orthogonal quadrature filters associated with (A, Γ). We define the basic nonstationary wavelet packets $\{w_n\}_{n=0}^{\infty}$ by $w_0 = \Phi$, $w_1 = \Psi$, and for $2^k \le n < 2^{k+1}$ with binary expansion $n = \sum_{j=1}^{k+1} \varepsilon_j 2^{j-1}$, we let

$$\hat{w}_n(\xi) = \left[\prod_{j=1}^{k+1} m_{\varepsilon_j}^{(k-j+2)}(D^j\xi) \right] \hat{\Phi}(D^{k+1}\xi).$$

Let us state two most important facts about the wavelet packets from the above definition. The two theorems below show how to extract orthonormal bases from the wavelet packet construction above, and thus gives us some new (and hopefully useful) tools to signal and image processing. We have included a sketch of the proofs for convenience. However, the reader should notice that everything works exactly as in the one-dimensional case, only the multiresolution structure matters.

Theorem 1.4. *The basic wavelet packets*

$$\{w_n(x-k) | 0 \le n < 2^j, k \in \mathbb{Z}^d\}$$

form a basis for V_j. *Furthermore,*

$$\{w_n(x-k) | n \in \mathbb{N}_0, k \in \mathbb{Z}^d\}$$

form an orthonormal basis for $L^2(\mathbb{R}^d)$.

Proof. Let $\Omega_n = \overline{\text{Span}}\{w_n(\cdot - k)\}_{k\in\mathbb{Z}^d}$, and define $\delta f(x) = \sqrt{2} f(Ax)$. Using the QMF-condition it is not hard to verify that $\delta\Omega_n = \Omega_{2n} \oplus \Omega_{2n+1}$ (see eg. [19, p. 112]). Thus,

$$\begin{aligned}
\delta\Omega_0 \ominus \Omega_0 &= \Omega_1 \\
\delta^2\Omega_0 \ominus \delta\Omega_0 &= \delta\Omega_1 = \Omega_2 \oplus \Omega_3 \\
\delta^3\Omega_0 \ominus \delta^2\Omega_0 &= \delta\Omega_2 \oplus \delta\Omega_3 = \Omega_4 \oplus \Omega_5 \oplus \Omega_6 \oplus \Omega_7 \\
&\vdots \\
\delta^k\Omega_0 \ominus \delta^{k-1}\Omega_0 &= \Omega_{2^{k-1}} \oplus \Omega_{2^{k-1}+1} \oplus \cdots \oplus \Omega_{2^k-1}.
\end{aligned}$$

By telescoping the above equalities we finally get the wanted result

$$\delta^k\Omega_0 \equiv \delta^k V_0 = V_k = \Omega_0 \oplus \Omega_1 \oplus \cdots \oplus \Omega_{2^k-1},$$

and $\cup_{k\geq 0} V_k$ is dense in $L^2(\mathbb{R}^d)$ by the definition of a multiresolution analysis. ■

The above theorem can be generalized considerably. The following construction gives us a whole library of orthonormal bases each with different time-frequency properties.

Theorem 1.5. *Let $\{w_n\}$ be a family of non-stationary wavelet packets associated with the dilation matrix A. For every partition P of $\mathbb{N}_0$ into sets of the form $I_{nj} = \{n2^j, \ldots, (n+1)2^j - 1\}$ with $n, j \in \mathbb{N}_0$, the family*

$$\{2^{j/2} w_n(A^j \cdot - k)\}_{k\in\mathbb{Z}^d, I_{nj}\in P}$$

is an orthonormal basis for $L^2(\mathbb{R}^d)$.

Proof. An argument similar to the one in Theorem 1.4 shows that

$$\delta^k\Omega_n = \Omega_{2^k n} \oplus \Omega_{2^k n+1} \oplus \cdots \oplus \Omega_{2^k(n+1)-1}.$$

Moreover, the functions $\{2^{j/2} w_n(A^j \cdot - q)\}_{q\in\mathbb{Z}^d}$ span the space $\delta^j\Omega_n$ and

$$\sum_{I_{nj}\in P} \delta^j\Omega_n = \bigoplus_{q\geq 0} \Omega_q = L^2(\mathbb{R}^d),$$

which proves the theorem. ■

Our focus in the remainder of this paper will be on a special case of the above construction that can be considered the natural generalization of the Walsh system on $[0, 1)$ and on an associated class of smooth non-stationary wavelet packets. The Walsh function will be associated with dilation matrices that admit a Haar type multiresolution analysis and thus a generalization of the Haar wavelet. We derive some properties of generalized Haar wavelets in L^p below.

2. Generalized Haar Functions

Let A be a $(d\times d)$-dilation matrix with $|\det A| = 2$. We are interested in the case where there is an associated multiresolution analysis generated by a scaling function given by the characteristic function of a set $Q \subset \mathbb{R}^d$, called a tile. For general A and $d > 3$ there is no guarantee that such a set Q exists, see [7, 6], so we have to restrict our construction to dilation matrices A which admit such a tile. The situation is better for $1 \leq d \leq 3$ since it can be proved that a tile always exists ([7, 6]). For the remainder of this paper we assume that our A is such that an associated tile Q exists.

The set Q has many nice properties under the action of A. One can in fact show that $AQ = Q \cup (Q+\Gamma_Q)$ for some $\Gamma_Q \in \mathbb{Z}^d$ and we always have $|Q| = 1$, see [19]. Hence $Q = A^{-1}Q \cup A^{-1}(Q+\Gamma_Q)$ and

$$\hat{\chi}_Q(\xi) = m_0(D\xi)\hat{\chi}_Q(D\xi), \tag{2.1}$$

where $m_0(\xi) = \frac{1}{2} + \frac{1}{2}e^{-i\langle \Gamma_Q, \xi\rangle}$. Also, note that $|A^{-1}Q| = \frac{1}{2}$, so A^{-1} splits Q into two sub-tiles of equal measure. We let

$$\mathcal{D}_0 = \{\Omega : \Omega = A^{-j}(Q+\gamma), \gamma \in \mathbb{Z}^d, \text{ and } \Omega \subset Q\}$$

denote the collection of Q-dyadic sets. Note that two Q-dyadic sets Q_1 and Q_2 with $|Q_1| \le |Q_2|$ share the following important property of the dyadic sets on $[0,1)$, namely either $Q_1 \cap Q_2 = \emptyset$ or $Q_1 \subset Q_2$. We also need the unrestricted collection of Q-dyadic sets given by

$$\mathcal{D} = \{\Omega : \Omega = A^{-j}(Q+\gamma), \gamma \in \mathbb{Z}^d, j \in \mathbb{Z}\}.$$

With this setup we can define the natural generalization of the Haar function on $[0,1)$.

Definition 2.1. With Q and Γ_Q as above, we define the generalized Haar function by

$$H(x) = \chi_{A^{-1}Q}(x) - \chi_{A^{-1}(Q+\Gamma_Q)}(x).$$

The Haar system on Q is given by

$$\{\chi_Q\} \cup \{2^{j/2}H(A^jx - k) | j \ge 0, k \in \mathbb{Z}^d, \text{ and } \operatorname{supp}(H(A^jx-k)) \subset Q\}.$$

There is a unique way to index the Haar functions by $\mathcal{D}_0$. For $\Omega \in \mathcal{D}_0$ we simply let H_Ω denote the generalized Haar function (normalized in $L^2(Q)$) with support equal to Ω.

One would suspect that the generalized Walsh functions form an unconditional basis for $L^p(Q)$, $1 < p < \infty$, and this is exactly the conclusion of the following Theorem. We give a proof based on Burkholder's L^p-inequality for martingales just to stress the connection between generalized Haar multiresolution analyses and probability theory.

Theorem 2.2. *Let $\{H_\Omega\}_{\Omega \in \mathcal{D}_0}$ be the generalized Haar system associated with the tile Q. Then $\{H_\Omega\}_{\Omega\in\mathcal{D}_0}$ constitutes an unconditional basis for $L^p(Q)$, $1 < p < \infty$.*

Proof: Let us first verify that the system is dense in $L^p(Q)$, $1 < p < \infty$. Let

$$K_n(x,y) = \sum_{|I|=2^{-n}} H_\Omega(x)H_\Omega(y)$$

be the kernel of the projection onto V_n. We have, for $y \in I$, $|I| = 2^{-n}$,

$$\int_Q |K_n(x,y)|\,dx = |H_\Omega(y)| \int_Q 2^{n/2}\chi_Q(A^n x)\,dx = |H_\Omega(y)|2^{n/2}2^{-n} = 1,$$

and similarly, for $x \in I$,

$$\int_Q |K_n(x,y)|\,dy = |H_\Omega(x)|2^{n/2}2^{-n} = 1.$$

Hence, by standard estimates, the projection onto V_n is bounded on $L^p(Q)$, $1 < p < \infty$. Now, each V_n is spanned by a finite number of Haar functions and χ_Q so it suffices to show that $P_nf \to f$ in $L^p(Q)$-norm as $n \to \infty$ for every $f \in L^\infty(Q)$ since such functions are dense in $L^p(Q)$, $1 < p < \infty$. Let $f \in L^\infty(Q)$, and suppose $2 < p < \infty$. We have, for $p^{-1} = \alpha/2 + (1-\alpha)/(p+1)$, using the generalized Hölder inequality,

$$\|f - P_nf\|_p \le \|f - P_nf\|_2^\alpha \|f - P_nf\|_{p+1}^{1-\alpha}.$$

Hence, $\|f - P_nf\|_p \to 0$ since $0 < \alpha < 1$ and $\|f - P_nf\|_{p+1}$ is bounded by a multiple of $\|f\|_{p+1}$. The case $1 < p < 2$ can be handled the same way. To prove that the system is unconditional, we build the following regular martingale on the probability space (Q, dx). Write $\mathcal{D}_0 = \{\Omega_0, \Omega_1, \dots\}$ in such a way that $|\Omega_n| \geq |\Omega_{n+1}|$, $n \geq 0$. Let $\mathcal{B}_0$ be the σ-algebra generated by $\Omega_0 = Q$ and $\emptyset$. Suppose $\mathcal{B}_n$ has been defined, then we let $\mathcal{B}_{n+1}$ be the smallest σ-algebra generated by $\mathcal{B}_n$ and Ω_{n+1}. Let $f \in L^p(Q)$. It is easy to check that the expectation $E^{\mathcal{B}_n} f$ is given by the projection onto $\text{span}\{\chi_{\Omega_0}, \chi_{\Omega_1}, \dots, \chi_{\Omega_n}\}$, so $f_n = E^{\mathcal{B}_n} f$ is indeed a regular martingale w.r.t. $\{\mathcal{B}_n\}_{n=0}^\infty$ and it follows from Burkholder's theorem that the martingale difference sequence $\{f_{n+1} - f_n\}_{n=0}^\infty$ converges unconditionally in $L^p(Q)$, $1 < p < \infty$. However, $\{f_{2n} - f_{2n-1}\}$ are just the partial sums of the expansion of f in the generalized Haar wavelets and the result follows. ∎

3. Generalized Walsh Functions

The Walsh system on $[0, 1)$ is the system of basic wavelet packets associated with the Haar multiresolution analysis, and using the setup introduced in the previous section we can use the same scheme to obtain a natural generalization of the Walsh system to higher dimensional domains.

Let $m_0(\xi) = \frac{1}{2} + \frac{1}{2} e^{-i\langle \Gamma_Q, \xi \rangle}$ be the low-pass for a generalized Haar wavelet as defined by (2.1). We define the associated high-pass Haar filter by $m_1(\xi) = \frac{1}{2} - \frac{1}{2} e^{-i\langle \Gamma_Q, \xi \rangle}$. We have the following definition of the generalized Walsh functions.

Definition 3.1. The generalized Walsh function $\{\mathcal{W}_n\}_{n=0}^\infty$ are the basic wavelet packets generated by the Haar low-pass and high-pass filters starting from the Haar scaling function and wavelet.

Remark 3.2. The generalized Walsh functions can also be defined recursively by letting $\mathcal{W}_0(x) = \chi_Q(x)$ and then we define $\{\mathcal{W}_n\}_{n=1}^\infty$ recursively by

$$\mathcal{W}_{2n+\varepsilon}(x) = \mathcal{W}_n(Ax) + (-1)^\varepsilon \mathcal{W}_n(Ax - \Gamma_Q), \qquad \varepsilon = 0, 1.$$

The third possible definition is to view the generalized Walsh system as the product system on the probability space (Q, dx) defined by the generalized Rademacher functions. The generalized Rademacher functions are obtained by letting

$$r_0(x) = \sum_{k \in \mathbb{Z}^d} H(x - k) \in L^\infty(\mathbb{R}^d),$$

where H is the Haar function of Definition 2.1, and then we define $r_n(x) = r_0(A^n x)$. Then for $n \in \mathbb{N}_0$ with binary expansion $n = \sum_{j=0}^\infty \varepsilon_j 2^j$ we have

$$\mathcal{W}_n(x) = \chi_Q(x) \prod_{j=0}^\infty r_j(x)^{\varepsilon_j},$$

which can be proved easily by induction. Notice that an easy consequence of this definition is that

$$\mathcal{W}_n(x)\mathcal{W}_m(x) = \mathcal{W}_{n \oplus m}(x), \tag{3.1}$$

where $\oplus$ is the bitwise "exclusive or" operator.

Remark 3.3. From an abstract point of view there is no difference between the Walsh system and the Generalized Walsh system, they are both realizations of the characters for the group $2^{\mathbb{N}}$, however the probability spaces in which the realizations live are very different which makes the functions adapted to analysis of different types of objects.

The first thing we want to check is that the generalized Walsh system constitutes a Schauder basis for $L^p(Q)$, for $1 < p < \infty$. This will be the content of Theorem 3.6. But first let us first recall some important facts about the classical Walsh system on $[0,1)$. The system is defined recursively on $[0,1)$ by letting $W_0 = \chi_{[0,1)}$ and

$$W_{2n+\varepsilon}(x) = W_n(2x) + (-1)^\varepsilon W_n(2x-1), \qquad \varepsilon = 0,1.$$

Clearly, this is a special case of our new construction with $d = 1$. One important fact we need is that for $2^J \le n < 2^{J+1}$ we have

$$W_n(x) = \sum_{s=0}^{2^J-1} W_{n-2^J}(s2^{-J}) W_1(2^J x - s)$$

The $2^J \times 2^J$-matrix defined by $(\mathcal{H}_J)_{i,j} = 2^{-J} W_i(j2^{-J})$ for $i,j = 0,1,\dots,2^J-1$, is called the Hadamard matrix of order 2^J. The proof of this fact can be found in [13], and we will in fact prove a more general statement in Section 4. The following lemma about the generalized Haar functions is elementary and we leave the proof to the reader.

Lemma 3.4. *Suppose $F \subset \mathcal{D}_0$ is a finite subset for which $f = \sum_{\Omega\in F} c_\Omega H_\Omega \in W_j$. Then*

$$\|f\|_p = 2^{j(1/2-1/p)} \left(\sum_{\Omega\in F} |c_\Omega|^p\right)^{1/p}.$$

From this simple Lemma, and from the fact that the classical Walsh system is a Schauder basic for $L^p[0,1)$, $1 < p < \infty$, we can deduce the following property of the Hadamard matrix

Lemma 3.5. *Let $\mathcal{H}_n$ be the $2^n \times 2^n$ Hadamard matrix, and let D^n_m be the $2^n \times 2^n$ diagonal matrix with m $1's$ in the upper left corner and zeros everywhere else. Then there exits a constant C independent of m and n such that*

$$\|\mathcal{H}_n D^n_m \mathcal{H}^\star_n\|_{\ell^p\to\ell^p} \le C.$$

Proof. Given $\{c_j\}_{j=1}^{2^n} \subset \mathbb{C}$ we form $f = \sum_{j=2^n}^{2^{n+1}} c_{j-2^n+1} W_j$ and $f_m = \sum_{j=2^n}^{2^n+m} c_{j-2^n+1} W_j$, where $\{W_j\}_n$ the Walsh system on $[0,1)$. We have, by the Schauder basis properties of the Walsh system,

$$\|f_m\|_p \le C\|f\|_p,$$

with C independent of m and n. Recall that the Hadamard matrix $\mathcal{H}_n$ is the change of basis matrix between the Walsh basis for W_n and the Haar basis for the same space. Hence, by Lemma 3.4

$$\|f\|_p = 2^{n(1/2-j/p)} \|\mathcal{H}_n[(c_j)]\|_{\ell^p} \quad \text{ and } \|f_m\|_p = 2^{n(1/2-1/p)} \|\mathcal{H}_n D^n_m \mathcal{H}^\star_n[\mathcal{H}_n(c_j)]\|_{\ell^p},$$

and we conclude that

$$\|\mathcal{H}_n D^n_m \mathcal{H}^\star_n\|_{\ell^p\to\ell^p} \le C.$$

■

We notice that for $2^J \le n < 2^{J+1}$ the wavelet packet $\mathcal{W}_n$ is given as a sum of exactly 2^J wavelets in W_J with the expansion coefficients given by the procedure outlined in Definition 1.3. The coefficients of the generalized Haar low-pass and high-pass filters are the same as in the one-dimensional case so we deduce that there is an ordering of the generalized Haar functions $\{H_\Omega\}_{\Omega\in\mathcal{D}_0,|\Omega|=2^{-J}}$ such that the wavelet packets $\{\mathcal{W}_n\}_{n=2^J}^{2^{J+1}-1}$ is given by the Hadamard transform of the Haar functions w.r.t. this ordering. We can now state and prove the following result.

Theorem 3.6. *Let $\{\mathcal{W}_n\}_{n=0}^\infty$ be a generalized Walsh system. Then $\{\mathcal{W}_n\}_{n=0}^\infty$ form a Schauder basis for $L^p(Q)$, $1 < p < \infty$.*

Proof: The generalized Walsh system is dense in $L^p(Q)$ since it is possible to write every Haar wavelet H_I as a finite linear combination of generalized Walsh functions, and the Haar system is dense in $L^p(Q)$ by Theorem 2.2. So, given $f_n = \sum_{j=0}^{n-1} c_n \mathcal{W}_j$ for some sequence $\{c_j\} \subset \mathbb{C}$, it suffices to prove that there exists a constant C such that $\|f_m\|_p \leq C\|f_n\|_p$ whenever $m \leq n$. Define $s, k \geq 0$ by $m = 2^s + k$, $k < 2^s$, and write $f_m = f_{2^s} + (f_m - f_{2^s})$. Clearly, $f_{2^s} = P_{V_s} f_n$ so $\|f_{2^s}\|_p \leq C\|f_n\|_p$ by Theorem 2.2. All that remains is to bound $f_m - f_{2^s} \in W_s$. Let $M_s = [\langle \mathcal{W}_j, H_I \rangle]_{j=2^s, H_I \in W_s}^{2^{s+1}}$ be the change of basis matrix from the generalized Walsh basis for W_s to the Haar basis for W_s. There exists a ordering of the Haar functions $\{H_\Omega\}_{|\Omega|=2^{-j}}$ such that the change of basis matrix is given by the Hadamard Transform, and the coefficients of $f_m - f_{2^s}$ in the Haar basis are thus given by,

$$M_s D_m^s M_s^\star [M_s (c_j)_{j=2^s}^{2^{s+1}-1}],$$

where D_m^s is the $2^s \times 2^s$ diagonal matrix with m 1's in the upper left corner and zeros everywhere else. By Lemma 3.5, $\|M_s D_m^s M_s^\star [M_s (c_j)_{j=2^s}^{2^{s+1}-1}]\|_{\ell^p} \leq C\|M_s D_{2^s}^s M_s^\star [M_s (c_j)_{j=2^s}^{2^{s+1}-1}]\|_{\ell^p}$ with C a constant independent of m and s. Hence, from Lemma 3.4 we deduce that

$$\|f_m - f_{2^s}\|_p \leq C\|P_{W_s} f_n\|_p \leq C_1 \|f_n\|_p,$$

and we are done. ∎

For technical reasons we will need the following special class of dilation matrices.

Definition 3.7. Let A be a $(d \times d)$-dilation matrix with $|\det A| = 2$. We say that A is almost isotropic if there exist integers s, t such that $A^s = 2^t I_d$, where I_d is the $(d \times d)$ identity matrix.

Remark 3.8. One example of an almost isotropic dilation matrix is the quincunx dilation

$$A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix},$$

which satisfies $A^8 = 16 I_2$. This example shows that the structure of the multiresolution analysis associated with an almost isotropic dilation matrix can be significantly different from the pure isotropic case.

Remark 3.9. Any 2×2 dilation matrix with determinant ± 2 is almost isotropic, which follows from the characterization of such matrices given in [1].

Fix a Haar multiresolution analysis associated with a $(d \times d)$-dilation matrix A with $|\det A| = 2$. Let Q be a tile associate with this matrix, and let $\{\mathcal{W}_n\}_n$ be the associated Walsh functions. The following operator will be fundamental in our study of the metric properties of the Walsh wavelet packet library.

Definition 3.10. The Carleson operator G for the wavelet packet system $\{w_n\}_n$ is defined by

$$(Gf)(x) = \sup_{N \geq 0} \left| \sum_{n=0}^{N} \sum_{k \in \mathbb{Z}^d : |k| \leq N} \langle f, w_n(\cdot - k) \rangle w_n(x - k) \right|,$$

for $f \in L^p(Q)$, $1 < p < \infty$.

The Carleson operator picks out the partial sum with the worst pointwise behavior at each point $x \in Q$. It is clearly not a priori obvious that the operator for a given system is finite at any point for general functions f, but the following will be proved in Appendix A. We remind the reader that an operator T mapping $L^p(\mathbb{R}^d)$ into the set of measurable functions is of strong type (p,p) if T is sub-linear and satisfies $\|Tf\|_p \leq C_p \|f\|_p$ for some finite constant C_p.

Theorem 3.11. *The Carleson operator associated with any generalized Walsh system generated by an almost isotropic dilation matrix is of strong type (p,p), $1<p<\infty$.*

Remark 3.12. There are several proofs of this fact for the one dimensional Walsh system, see eg. [2, 14]. The proof we outline in the appendix is based a technique introduced by C. Thiele in [18].

The Corollary below follows by standard arguments from Theorem 3.11.

Corollary 3.13. *The Walsh wavelet packet expansion of any $f\in L^p(Q)$, $1<p<\infty$, converges a.e.*

4. Smooth Walsh Type Functions

The expansion of L^p functions in the generalized Walsh functions works well as we have seen in the previous section, however the basis functions are not continuous which can be a problem for certain applications. The aim of this section is to introduce smooth analogues of the generalized Walsh system with the same nice L^p-properties. Let us define the class of functions we have in mind.

Definition 4.1. Let $\{\mathcal{W}_n^S\}_{n\geq 0,k\in\mathbb{Z}}$ be a family of non-stationary wavelet packets constructed by using a family $\{(m_0^{(p)},m_1^{(p)})\}_{p=1}^\infty$ of finite filters in Definition 1. If there exists a constant $J\in\mathbb{N}$ such that $(m_0^{(p)},m_1^{(p)})$ is the Haar low-pass and high-pass filter, respectively, for every $p\geq J$ and w_1 has compact support then we call $\{\mathcal{W}_n^S\}_{n\geq 0}$ a family of Walsh type wavelet packets.

We have to state and prove a few technical lemmas before we can attack the main result stated in Theorem 4.10 below. The lemmas below are well known results in the one-dimensional isotropic case, and we just have to tweak the proofs a little bit to make them work for almost isotropic dilation in $\mathbb{R}^d$. The techniques used should be well know to the reader, so we will only give the outlines of the proofs. Further details on the techniques can be found in [9, 10, 19].

Lemma 4.2. *Let A be an almost isotropic $(d\times d)$-dilation matrix, and let $f^i\in C^1(\mathbb{R}^d)\cap L^2(\mathbb{R}^d)$, $i=1,2$, be two functions for which*

$$|f^i(x)|,|\partial/\partial x_i f^j(x)|\leq C(1+|x|)^{-d+\varepsilon},\qquad i=1,2,\dots,d, j=1,2,$$

for some constant C. Suppose $\{f^i_{j,k}\equiv 2^{j/2}f^2(A^j\cdot-k)\}_{j\in\mathbb{Z},k\in\mathbb{Z}^d}$ is an orthonormal system for $i=1,2$, and let $\varepsilon\in\ell^\infty(\mathbb{Z}\times\mathbb{Z}^d)$ with $\|\varepsilon\|_{\ell^\infty}\leq 1$. Then the operator $T:L^2(\mathbb{R}^d)\to L^2(\mathbb{R}^d)$ defined by

$$Tg=\sum_{j\in\mathbb{Z},k\in\mathbb{Z}^d}\varepsilon_{j,k}\langle g,f^1_{j,k}\rangle f^2_{j,k}.$$

can be extended to a bounded operator on $L^p(\mathbb{R}^d)$, $1<p<\infty$, with bound independent of ε.

Proof: Fix the nonnegative integers s,t such that $A^s=2^tI_d$, and take any finite sequence $\varepsilon\in\mathbb{Z}\times\mathbb{Z}^d$ with $\|\varepsilon\|_{\ell^\infty}\leq 1$. We can write any integer j as $j=us+r$ with $u\in\mathbb{Z}$ and $0\leq r<s$. Hence

$$\begin{aligned}Tg&=\sum_{j\in\mathbb{Z},k\in\mathbb{Z}^d}\varepsilon_{j,k}\langle g,f^1_{j,k}\rangle f^2_{j,k}\\&=\sum_{r=0}^{s-1}\sum_{u\in\mathbb{Z},k\in\mathbb{Z}^d}\varepsilon_{us+r,k}\langle g,2^{j/2}f^1(2^{tu}A^r\cdot-k)\rangle 2^{j/2}f^2(2^{tu}A^rx-k)\end{aligned}$$

It follows that

$$\|Tg\|_p \leq C \sum_{r=0}^{s-1} \left\| \sum_{u\in\mathbb{Z},k\in\mathbb{Z}^d} \varepsilon_{us+r,k} \langle g, 2^{tdu/2} f^1(2^{tu}A^r \cdot -k)\rangle 2^{tdu/2} f^2(2^{tu}A^r x - k) \right\|_p,$$

where we have used that $j = tdu + r$. Now, each term on the right can be shown to be associated with a Calderón-Zygmund operator using a straightforward modification of well known estimates, see eg. [19, 10], using the decay of f^i and $\partial/\partial x_i f^j$. ■

The following Lemma generalizes Lemma 12 in [11].

Lemma 4.3. *Let Ψ be a wavelet associated with an almost isotropic $(d \times d)$-dilation matrix A, and let H be a generalized Haar wavelet for the same dilation. Suppose $\Psi \in C^1(\mathbb{R}^d)$ satisfies*

$$|\Psi(x)|, |\partial/\partial x_i \Psi(x)| \leq C(1+|x|)^{-d+\varepsilon}, \qquad i = 1,2,\ldots,d,$$

for some constant C. Then the wavelet systems generated by Ψ and H, respectively, are equivalent unconditional bases for $L^p(\mathbb{R}^d)$, $1 < p < \infty$.

Proof. We can use the same technique as in proof presented on pages 166-167 of [10]. The kernel for the operator P mapping one system onto the other is given by

$$K_\varepsilon(x,y) = \sum_{j\in\mathbb{Z},k\in\mathbb{Z}^d} \varepsilon_{j,k} 2^j H(A^j x - k)\Psi(A^j y - k).$$

$K_\varepsilon(x,y)$ is smooth in the y-variable and we can use the same argument as in Lemma 4.2 to show that P is bounded on $L^p(\mathbb{R}^d)$, $1 < p \leq 2$. All that remains is to prove that P^* is bounded from $L^1(\mathbb{R}^d)$ into $L^1_{\text{weak}}(\mathbb{R}^d)$. To do this, we take $f \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ and make a Calderon-Zygmund decomposition of f at level $\alpha > 0$ with the twist that the decomposition be based not on dyadic d-cubes but on the Q-dyadic sets in $\mathcal{D}$. There is no problem making this type of decomposition following the outline in e.g. [4, Chap. 9] since for a.a. $x \in \mathbb{R}^d$ there is a sequence $\{Q_j\}_{j=1}^\infty \subset \mathcal{D}$ with $|Q_j| = 2^{-j}$ for which the Lebesgue theorem of differentiation holds. This due to the fact that A is almost isotropic (the eccentricity of the sets in $\mathcal{D}$ is uniformly bounded). With this slightly modified Calderón-Zygmund decomposition in hand we can complete the proof of the Lemma by following [10, p. 167]. ■

We now use the Lemmas presented above to obtain the first interesting conclusion about the Walsh type wavelet packets, the generalized Walsh type wavelet packets are equivalent to the Walsh functions in $L^p(\mathbb{R}^d)$, $1 < p < \infty$.

Lemma 4.4. *Let $\{\mathcal{W}_n\}_{n=0}^\infty$ be a generalized Walsh systems and $\{\mathcal{W}_n^S\}_{n=0}^\infty$ a Walsh type system associated with the same almost isotropic $d \times d$-dilation matrix. Suppose $\mathcal{W}_0^S \in C^1(\mathbb{R}^d)$ and*

$$|\mathcal{W}_0^S(x)|, |\partial/\partial x_i \mathcal{W}_0^S(x)| \leq C(1+|x|)^{-d+\varepsilon}, \qquad i = 1,2,\ldots,d,$$

for some constants $C, \varepsilon > 0$. Then there exists an isomorphism $P : L^p(\mathbb{R}^d) \to L^p(\mathbb{R}^d)$, $1 < p < \infty$, for which $P\mathcal{W}_n(\cdot - k) = \mathcal{W}_n^S(\cdot - k)$.

Proof. Let K be the scale from which only the Haar filters are used to generate the Walsh type wavelet packets. Let $\{V_j\}$ be the Haar MRA associated with the generalized Walsh functions. Since P_{V_K} is bounded on $L^p(\mathbb{R}^d)$ it suffices to prove that PP_{V_K} and $P(1 - P_{V_K})$ are bounded. One can easily check that PP_{V_K} is bounded by brute force estimates on the kernel using that only 2^K different functions (and their integer translates) are involved.

We turn to $P(1 - P_{V_K})$. Let $T : L^p(\mathbb{R}^d) \to L^p(\mathbb{R}^d)$ be one of the isomorphism from Lemma 4.3 mapping the generalized Haar system onto some $C^1(\mathbb{R}^d)$ wavelet system generated by the wavelet Ψ. We use the map T to define an intermediary system $\{\mathcal{W}_n^I(x-k)\}_{n=1,k\in\mathbb{Z}^d}^\infty$ defined by $\mathcal{W}_n^I(x-k) = T\mathcal{W}_n(x-k)$. The new system is clearly equivalent to the generalized Walsh system. Let $v_{j,k}^n = 2^{j/2}\mathcal{W}_n^S(A^j\cdot -k)$ and $g_{j,k}^n = 2^{j/2}\mathcal{W}_n^I(A^j\cdot -k)$. Notice that

$$\{g_{j,k}^n\}_{2^K\leq n<2^{K+1},(j,k)\in\mathbb{Z}\times\mathbb{Z}^d} \quad \text{and} \quad \{v_{j,k}^n\}_{2^K\leq n<2^{K+1},(j,k)\in\mathbb{Z}\times\mathbb{Z}^d}$$

are both orthonormal bases for $L^2(\mathbb{R}^d)$. It follows from lemma 4.2 that there is an isomorphism $U : L^p(\mathbb{R}^d) \to L^p(\mathbb{R}^d)$ for which

$$Ug_{j,k}^n = v_{j,k}^n, \qquad 2^K \leq n < 2^{K+1}, (j,k) \in \mathbb{Z}\times\mathbb{Z}^d.$$

Let $n \geq 2^{N+1}$. We expand $\mathcal{W}_n^S(x-k)$ to get

$$\mathcal{W}_n^S(x-k) = \sum_{s\in F} c_{n,s} v_{K,s}^{\tilde{n}}(x-k), \tag{4.1}$$

with $2^K \leq \tilde{n} < 2^{K+1}$ and $F \subset \mathbb{Z}^d$ a finite set (depending on n). The coefficients $c_{n,s}$ depend only on n and the Haar filter. Thus, $\mathcal{W}_n^I(x-k)$ has the same expansion:

$$\mathcal{W}_n^I(x-k) = \sum_{s\in F} c_{n,s} g_{K,s}^{\tilde{n}}(x-k). \tag{4.2}$$

We conclude that $U\mathcal{W}_n^I(x-k) = \mathcal{W}_n^S(x-k)$ for $n \geq 2^{K+1}$ and $k \in \mathbb{Z}^d$, i.e. the isomorphism $UT : L^p(\mathbb{R}^d) \to L^p(\mathbb{R}^d)$, $1 < p < \infty$, maps $\mathcal{W}_n(x-k)$ onto $\mathcal{W}_n^S(x-k)$ for $n \geq 2^{K+1}$. This completes the proof of the Lemma. ■

Remark 4.5. The previous Lemma shows that the generalized Walsh type functions do form a Schauder basis for $L^p(\mathbb{R}^d)$, $1 < p < \infty$. However, the system is bound to fail as a basis for $L^1(\mathbb{R}^d)$ since the functions are uniformly bounded.

Lemma 4.6. *Let A be an almost isotropic $(d\times d)$-dilation matrix associated with an MRA $\{V_j\}$ with scaling function Φ satisfying*

$$|\Phi(x)| \leq C(1+|x|)^{-n-\varepsilon},$$

for some $\varepsilon > 0$. Then the Carleson operator, $f \to \sup_j |P_{V_j}f(x)|$, associated with the projections onto V_j is of strong type (p,p), $1 < p < \infty$.

Proof. By assumption, $A^s = 2^t I_d$ for some $s, t \in \mathbb{N}$, and for $j \in \mathbb{Z}$ we write $j = su + r$ with $0 \leq r < s$. Then the kernel of the projection onto V_j can be written as

$$\begin{aligned} K_j(x,y) &= \sum_{k\in\mathbb{Z}^d} 2^j \Phi(A^j x - k)\overline{\Phi(A^j y - k)} \\ &= 2^r \sum_{k\in\mathbb{Z}^d} 2^{tdu}\Phi(2^{tu}A^r x - k)\overline{\Phi(2^{tu}A^r y - k)}, \end{aligned}$$

where we have used that $s = td$. From this and standard estimates we deduce that

$$|K_j(x,y)| \leq C2^{tdu}(1+2^{tu}|x-y|)^{-d-\varepsilon},$$

with C is a constant independent of j. But then it follows from [15, p. 62] that, for $f \in L^p(\mathbb{R}^d)$,

$$|P_j f(x)| = \left|\int_{\mathbb{R}^d} K_j(x,y)f(y)\,dy\right| \leq CMf(x),$$

where M is the Hardy-Littlewood maximal operator. Hence, $\sup_j |P_{V_j}f(x)| \leq CMf(x)$ and we are done. ■

Remark 4.7. The idea of using the maximal function to bound the scaling space projections is due to T. Tao [17].

Note that there is exactly 2^J values of $k \in \mathbb{Z}^d$ for which the function $\chi_Q(A^J x - k)$ has support contained in Q. Let $F_J \subset \mathbb{Z}^d$ denote the set of such k's. We let $Q_k^J = \operatorname{supp}\{\chi_Q(A^J x - k)\}$, $k \in F_J$.

Lemma 4.8. *Let $f_1 \in L^2(\mathbb{R}^d)$, and define $\{f_n\}_{n\geq 2}$ recursively by*

$$f_{2n+\varepsilon}(x) = f_n(Ax) + (-1)^\varepsilon f_n(Ax - \Gamma), \qquad \varepsilon = 0, 1.$$

Then for $n,\ J \in \mathbb{N}$, $2^J \leq n < 2^{J+1}$, we have

$$f_n(x) = \sum_{k \in F_J} \left(|Q_k^J|^{-1} \int_{Q_k^J} \mathcal{W}_{n-2^J}(\omega)\, d\omega \right) f_1(A^J x - k).$$

Proof. Clearly, it suffices to prove that

$$\mathcal{W}_n(x) = \sum_{k \in F_J} \left(|Q_k^J|^{-1} \int_{Q_k^J} \mathcal{W}_{n-2^J}(\omega)\, d\omega \right) \mathcal{W}_1(A^J x - k).$$

However, since $2^J \leq n < 2^{J+1}$, it follows from (3.1) that $\mathcal{W}_n(x) = \mathcal{W}_{n-2^J}(x)\mathcal{W}_{2^J}(x)$. Then the result follows from the fact that each $\mathcal{W}_{n-2^J}(x)$, $2^J \leq n < 2^{J+1}$, is constant on each set Q_k^J and $\operatorname{supp}\{\mathcal{W}_1(A^J x - k)\} = Q_k^J$. ■

Remark 4.9. We will use the notation $f(Q_k^J)$ to denote the average

$$|Q_k^J|^{-1} \int_{Q_k^J} f(\omega)\, d\omega.$$

We can state the main result about generalized Walsh type wavelet packets.

Theorem 4.10. *Let L be the Carleson operator for a basic Walsh type wavelet packet system $\{\mathcal{W}_n^S\}_n$ with $\mathcal{W}_0 \in C^1(\mathbb{R}^d)$. Then L is of strong type (p,p), $1 < p < \infty$.*

Proof. Let us begin by reducing the problem. Choose $N \in \mathbb{N}$ such that $\operatorname{supp}(\mathcal{W}_n^S) \subset [-N, N]^d$ for $n \geq 0$. Fix $p \in (1, \infty)$ and take any

$$f(x) = \sum_{n\geq 0, k \in \mathbb{Z}^d} c_{n,k} \mathcal{W}_n^S(x - k) \in L^p(\mathbb{R}^d).$$

Define

$$f_k(x) = \sum_{n\geq 0} c_{n,k} \mathcal{W}_n^S(x-k), \qquad g_k(x) = \sum_{n \geq 0} c_{n,k} \mathcal{W}_n(x-k).$$

We have $\|f_k\|_p \simeq \|g_k\|_p$, with bounds independent of k, by Lemma 4.4. Note that for $q \in \mathbb{Z}^d$,

$$|\{x \in q + [0,1)^d : |Lf(x)| > \alpha\}| \leq \frac{C}{\alpha^p} \sum_{|k-q| \leq (N+1)^d} \int |Lf_k(x)|^p\, dx,$$

so (using the Marcinkiewicz interpolation theorem) it suffices to prove that $\|Lf_k\|_p \leq C\|f_k\|_p$, where C is a constant independent of k, since

$$\sum_{q \in \mathbb{Z}^d} \sum_{|k-q| \leq (N+1)^d} \|f_k\|_p^p \leq 2^d (N+1)^d \sum_{k \in \mathbb{Z}^d} \|f_k\|_p^p \leq C 2^d (N+1)^d \sum_{k\in\mathbb{Z}^d} \|g_k\|_p^p \leq \tilde{C} 2^d (N+1)^d \|f\|_p^p.$$

We can, w.l.o.g., assume that $k = 0$. Let $K \in \mathbb{N}$ be the scale from which only the Haar filter is used to generate the wavelet packets $\{\mathcal{W}_n^S\}_{n \geq 2^{K+1}}$. Let $m \in \mathbb{N}$ and suppose $2^J \leq m < 2^{J+1}$ for some $J > K + 1$. Clearly, for each $x \in \mathbb{R}^d$,

$$\sum_{n=0}^{m} c_{n,0}\mathcal{W}_n^S(x) = \sum_{n=0}^{2^{K+1}-1} c_{n,0}\mathcal{W}_n^S(x) + \sum_{n=2^{K+1}}^{2^J-1} c_{n,0}\mathcal{W}_n^S(x) + \sum_{n=2^J}^{m} c_{n,0}\mathcal{W}_n^S(x),$$

so we have

(4.3)

$$\sup_{m\geq 1}\left|\sum_{n=0}^{m} c_{n,0}\mathcal{W}_n^S(x)\right| \leq \sup_{1\leq m<2^{K+1}}\left|\sum_{n=0}^{m} c_{n,0}\mathcal{W}_n^S(x)\right| + \sup_{J>K+1}\left|\sum_{n=2^{K+1}}^{2^J+1} c_{n,0}\mathcal{W}_n^S(x)\right| + \sup_{J>K+1}(M_J f_0)(x),$$

where

$$(M_J f_0)(x) = \sup_{2^J\leq m<2^{J+1}}\left|\sum_{n=2^J}^{m} c_{n,0}\mathcal{W}_n^S(x)\right|.$$

We use brute force to estimated the first term of (4.3)

$$\begin{aligned}\sup_{0<m<2^{K+1}}\left|\sum_{n=0}^{m} c_{n,0}\mathcal{W}_n^S(x)\right| &\leq \sum_{n=0}^{2^{K+1}-1} |c_{n,0}|\|\mathcal{W}_n^S(x)\|_\infty \chi_{[-N,N]^d}(x)\\ &\leq \|f_0\|_p \sum_{n=0}^{2^{K+1}-1} \|\mathcal{W}_n^S\|_{p'}\|\mathcal{W}_n^S(x)\|_\infty \chi_{[-N,N]^d}(x).\end{aligned}$$

The second term of (4.3) satisfies

$$\left\|\sup_{J>K+1}\left|\sum_{n=2^{K+1}}^{2^J+1} c_{n,0}\mathcal{W}_n^S(x)\right|\right\|_p \leq C\|f_0\|_p$$

by Lemma 4.6 since

$$\sum_{n=2^{K+1}}^{2^J+1} c_{n,0}\mathcal{W}_n^S(x) = P_{V_K}f_0(x) - P_{V_J}f_0(x)$$

so

$$\sup_{J>K+1}\left|\sum_{n=2^{K+1}}^{2^J+1} c_{n,0}\mathcal{W}_n^S(x)\right| \leq 2\sup_J |P_{V_J}f_0(x)|.$$

The challenge is to prove that the third term is of type (p,p). Note that

$$(M_J f_0)(x) \leq \sum_{j=0}^{2^K-1} (M_J^j f_0)(x),$$

where

$$(M_J^j f_0)(x) = \sup_{2^J+j2^{J-K}\leq m<2^J+(j+1)2^{J-K}}\left|\sum_{n=2^J+j2^{J-K}}^{m} c_{n,0}\mathcal{W}_n^S(x)\right|,$$

so it suffices to prove that

$$\|\sup_{J>K+1}(M_J^j f_0)\|_p \leq C\|f_0\|_p$$

for $j = 0, 1, \dots 2^K - 1$. Fix $J > K + 1$, $0 \le j < 2^K - 1$, and $2^J + j2^{J-K} \le m < 2^J + (j+1)2^{J-K}$. We have, using Lemma 4.8,

$$\Big|\sum_{n=2^J+j2^{J-K}}^{m} c_{n,0}\mathcal{W}_n^S(x)\Big| = \Big|\sum_{s\in F_{J-K}}\Big\{\sum_{n=2^J+j2^{J-K}}^{m} c_{n,0}\mathcal{W}_{n-2^J-j2^{J-K}}(Q_s^{J-K})\Big\}\mathcal{W}^S_{2^K+j}(A^{J-K}x - s)\Big|.$$

Define

$$F_m(t) = \sum_{n=2^J+j2^{J-K}}^{m} c_{n,0}\mathcal{W}_{n-2^J-j2^{J-K}}(t), \quad \text{and} \quad F(t) = \sup_{m<2^J+(j+1)2^{J-K}} |F_m(t)|.$$

The following estimate follows easily

$$\Big|\sum_{n=2^J+j2^{J-K}}^{m} c_{n,0}\mathcal{W}_n^S(x)\Big| \le \sum_{s\in F_{J-K}} F(Q_s^{J-K})|\mathcal{W}^S_{2^K+j}(A^{J-K}x - s)|.$$

Then using the fact that $\mathrm{supp}(\mathcal{W}^S_{2^K+j}) \subset [-N, N]^d$ we obtain the following estimate

$$\Big|\sum_{n=2^J+j2^{J-K}}^{m} c_{n,0}\mathcal{W}_n^S(x)\Big| \le \|\mathcal{W}^S_{2^K+j}\|_\infty \sum_{s\in F_{J-K}\cap S_{J-K}(x)} F(Q_s^{J-K}),$$

where $S_{J-K}(x) = A^{J-K}x + [-N-1, N+1]^d \subset \mathbb{R}^d$. Notice that $S_{J-K}(x) \cap F_{J-K}$ contains at most $2^d(N+1)^d$ points. We need an estimate of F that does not depend on J. Note that for k, $0 \le k < 2^{J-K}$, using (3.1),

$$\mathcal{W}_{2^J+j2^{J-K}}(\omega)\mathcal{W}_k(\omega) = \mathcal{W}_{2^J+j2^{J-K}+k}(\omega),$$

since the binary expansions of $2^J + j2^{J-K}$ and of k have no 1's in common. Hence,

$$|F_m(\omega)| = |\mathcal{W}_{2^J+j2^{J-K}}(\omega)F_m(\omega)| = \Big|\sum_{n=2^J+j2^{J-K}}^{m} c_{n,0}\mathcal{W}_n(\omega)\Big|,$$

so $F(\omega) \le 2(Gg_0)(\omega)$, with G the Carleson operator for the generalized Walsh system. Thus,

$$\Big|\sum_{n=2^J+j2^{J-K}}^{m} c_{n,0}\mathcal{W}_n^S(x)\Big| \le 2\|\mathcal{W}^S_{2^K+j}\|_\infty \sum_{s\in F_{J-K}\cap S_{J-K}(x)} |Q_s^{J-K}|^{-1}\int_{Q_s^{J-K}} Gg_0(\omega)\,d\omega.$$

We let $Q_s^\star$ be the smallest dyadic d-cube centered at x containing Q_s^{J-K}. Note that $|Q_s^\star| \le C2^d(N+1)^d|Q_s^{J-K}|$. We have

$$\begin{aligned}\Big|\sum_{n=2^J+j2^{J-K}}^{m} c_{n,0}\mathcal{W}_n^S(x)\Big| &\le 2\|\mathcal{W}^S_{2^K+j}\|_\infty \sum_{s\in F_{J-K}\cap S_{J-K}(x)} |Q_s^{J-K}|^{-1}\int_{Q_s^\star} (Gg_0)(t)\,dt \\ &\le C\|\mathcal{W}^S_{2^K+j}\|_\infty 2^{2d}(N+1)^{2d}(MGg_0)(x), \end{aligned} \tag{4.4}$$

where M is the maximal operator of Hardy and Littlewood. The righthand side of (4.4) does not depend on m nor J so we may conclude that

$$\sup_{J>K+1}(M_J^j f_0)(x) \le C\|\mathcal{W}^S_{2^K+j}\|_\infty 2^{2d}(N+1)^{2d}(MGg_0)(x), \quad \text{a.e.}$$

and thus, since M and G are both of strong type (p,p) (see Theorem 3.11),

$$\|\sup_{J>K+1}(M_J^j f_0)\|_p \le C\|g_0\|_p \le C_1\|f_0\|_p, \qquad j = 0, 1, \dots 2^K - 1,$$

and we are done. ■

The pointwise convergence result now follows by a standard argument (see [5])

Corollary 4.11. *The Walsh type wavelet packet expansion of any $f \in L^p(\mathbb{R}^d)$, $1 < p < \infty$, converges a.e.*

The basic Walsh type wavelet packets is only one out of an infinite number of the possible Walsh type wavelet packet bases given by Theorem 1.5 and it is interesting to know if we have the same convergence properties for other bases in the library. Fortunately, it turns out that we can generalize the above Corollary to any basis in the library, and the key to this result is the possibility of decomposing the partial sum operator for a given wavelet packet system in the basic wavelet packets. In fact, the proof below shows that the basis wavelet packets always have the *worst* metric properties of all the bases in the library.

Corollary 4.12. *Let $\mathcal{P} = \{I_{n,j}\}$ be a partition of $\mathbb{N}_0$ as in Theorem 1.5. Let $f \in L^p(\mathbb{R}^d)$, $1 < p < \infty$. Define the partial sum operator for the Walsh type wavelet packet system associated with $\mathcal{P}$ by*

$$S_N f(x) = \sum_{I_{n,j} \in \mathcal{P}: n \cdot j \leq N, k \in \mathbb{Z}^d} \langle f, 2^{j/2} \mathcal{W}_n^S(A^j \cdot - k)\rangle 2^{j/2} \mathcal{W}_n^S(A^j x - k).$$

We have $S_N f(x) \to f$ in $L^p(\mathbb{R}^d)$-norm and pointwise a.e.

Proof. Consider $S_N f(x)$. By the proof of Theorem 1.5 there is an $\tilde{N} \leq N$ such that

$$S_N f(x) = \sum_{n=0}^{\tilde{N}} \sum_{k \in \mathbb{Z}^d} \langle f, \mathcal{W}_n^S(\,\cdot - k)\rangle \mathcal{W}_n^S(x - k).$$

From this we obtain the pointwise bound $|S_N f(x)| \leq Lf(x)$, where L is the Carleson operator for the Walsh type system. Thus, the Carleson operator for the wavelet packet system given by $\mathcal{P}$, $\sup_N |S_N(f)(x)|$, is bounded pointwise by $Lf(x)$ and is thus of strong type (p,p), $1 < p < \infty$. Both claims of the Corollary follow easily from this fact. ■

Remark 4.13. In one dimension, the above Corollary generalizes the results obtained by the author in [12].

5. Periodic Wavelet Packets

The process of 1-periodization works well for one-dimensional wavelet and wavelet packets due to the fact that the one-dimensional multiresolution structure is based on integer shifts. The same is true for the general multiresolution structure in Definition 1.1 so it should be no surprise to the reader that we can periodize the nonseparable wavelet packets and obtain the same useful results as in the one dimensional case. We will just state the results and leave the easy details to the reader.

Let $\{\mathcal{W}_n\}_n$ be a wavelet packet system in $\mathbb{R}^d$ for which each $\mathcal{W}_n \in L^1(\mathbb{R}^d)$. For the wavelet packet $\mathcal{W}_{n,j,k}(x) = 2^{j/2}\mathcal{W}_n(A^j(x-k))$ we can define the associated periodized wavelet packet by

$$\mathcal{W}_{n,j,k}^{\mathrm{per}}(x) = \chi_\Sigma(x) 2^{j/2} \sum_{\gamma \in \mathbb{Z}^d} \mathcal{W}_n(A^j(x-\gamma) - k),$$

where Σ is any tile of $\mathbb{R}^d$ such as Q itself or the fundamental domain $[0,1)^d$. One can easily verify that Theorem 1.5 is still true with the obvious modification that the space Ω_n be defined as the closed span of $\{\mathcal{W}_{n,0,k}^{\mathrm{per}} | k \in \mathbb{Z}^d\}$. Also, notice that the dimension of $\mathrm{span}\{\mathcal{W}_{n,j,k}^{\mathrm{per}} | k \in \mathbb{Z}^d\}$ is exactly 2^j. For periodic Walsh type wavelet packets we obtain the periodic analog of Theorem 4.10.

Corollary 5.1. *Consider a system of periodic Walsh type wavelet packets* $\{\mathcal{W}_{n,0,0}^{per}\}_n$ *for which* $\mathcal{W}_0 \in C^1(\mathbb{R}^d)$. *Let* $f \in L^p(\Sigma)$, $1 < p < \infty$. *Then*

$$\sum_{n=0}^{N} \langle f, \mathcal{W}_{n,0,0}^{per} \rangle \mathcal{W}_{n,0,0}^{per}(x) \longrightarrow f, \qquad \text{as } N \to \infty,$$

in $L^p(\Sigma)$*-norm and pointwise a.e.*

Remark 5.2. The result can be proved by using the compact support of the aperiodic Walsh type wavelet packets to bound the Carleson operator for the periodic system by the Carleson operator for the aperiodic system.

Remark 5.3. The periodic version of the one-dimensional Walsh system is the system itself, so this case is not that interesting. However, for higher dimensional Walsh systems periodization has the advantage that it can transform the fundamental domain from the potentially complicated fractal tile Q to a less complicated fundamental domain such as $[0,1)^d$.

6. Some Examples of $C^k(\mathbb{R}^2)$ Walsh Type Wavelet Packets

We have all the machinery to obtain nice nonseparable $C^k(\mathbb{R}^d)$ wavelet packets with good L^p and pointwise properties provided that we can find appropriate low-pass filters yielding compactly supported $C^k(\mathbb{R}^d)$, $k \geq 1$, scaling functions associated with the given dilation matrix A. Unfortunately, such constructions are difficult in general mainly due to the fact that not every nonnegative trigonometric polynomial of two variables admits a spectral factorization. We remind the reader that it is still an open problem whether the quincunx dilation admits a $C^1(\mathbb{R}^2)$ compactly supported scaling function. However, a construction of C^k-wavelets, $k \geq 1$, is carried out in [1] for the special case of a 2×2-dilation matrix A satisfying $A^2 = 2I_2$ such as

$$A = \begin{bmatrix} 0 & 2 \\ 1 & 0 \end{bmatrix}. \tag{6.1}$$

We can obviously use these compactly supported scaling function/wavelet pairs and the associated filters in Definition 4.1 to construct examples of nonseparable Walsh type wavelet packets of type $C^k(\mathbb{R}^2)$, for $k = 1, 2, \ldots$.

Appendix A. A Proof of Theorem 3.11

We give a proof of Theorem 3.11 based on an elegant technique introduced by C. Thiele in [18], which he used to prove the same result for the one-dimensional Walsh system. We have made some adjustments to adapted the proof to the present multidimensional setting, but a large part of combinatorics involved in the proof of Theorem 3.11 is virtually identical to the combinatorics presented in [18] so we will only state those results and refer the reader to [18] for the details.

First, some notation. Fix a generalized Walsh system $\{\mathcal{W}_n\}_n$ associated with the tile Q. The set $\mathcal{F} = Q \times \mathbb{N}_0$ is called the generalized Walsh phase plane. Let $\Omega \in \mathcal{D}_0$ (see page 2 for the definition of $\mathcal{D}_0$) and $j, n \geq 0$. Consider sets of the form

$$\Omega \times \{n2^j, n2^j + 1, \ldots, n2^{j+1} - 1\} \subset Q \times \mathbb{N}_0.$$

We call such a set a tile if $2^j|\Omega| = 1$ and a bitile if $2^j|\Omega| = 2$. We let $\mathcal{T}$ and $\mathcal{B}$ denote the collection of all tiles and bitiles, respectively. Let P be a tile or bitile. We use the notation $P = \Omega_P \times \omega_P$ to separate the time and frequency sets of P. For $E \subset \mathcal{F}$ we define the following projection operator

$$\Pi_E f(x) = \sum_{n:(x,n) \in E} \langle f, \mathcal{W}_n \rangle \mathcal{W}_n(x).$$

The Carleson operator associated with the function $b : Q \to \mathbb{N} \cap [0, 2^N]$ is defined by Π_{E_b} where $E_b = \{(x, n) \subset Q \times \mathbb{N}_0 : n < b(x)\}$. It is clear that Theorem 3.11 will follow if we can prove that Π_{E_b} is of strong type on $L^p(Q)$, $1 < p < \infty$, with bound independent of b and N (the bound will depend on p).

We define a partial ordering on $\mathcal{B}$ by saying that $P \prec P'$ if $P \cap P' \neq \emptyset$ and $\Omega_P \subset \Omega_{P'}$ (or equivalently $\omega_{P'} \subset \omega_P$).

We fix $f \in L^p(Q)$, $1 < p < \infty$. For each $P \in \mathcal{B}$ we define the associated density

$$d_P = \left[\log_2 \sup_{P \prec P'} \|\Pi_{P'} f\|_\infty \right].$$

Using the ordering of the bitiles we split $\mathcal{B}$ according to their density as follows

- $\mathcal{T}_k = \{P \in \mathcal{B} : d_P = k\}$.
- $\mathcal{T}_k^{\max} = \{$maximal bitiles in $\mathcal{T}_k$ w.r.t. the given partial ordering of $\mathcal{B}\}$.
- $\mathcal{T}_{k,i} = \{P \in \mathcal{T}_k : 2^i \leq |P' \in \mathcal{T}_k^{\max} : P \prec P'\}| < 2^{i+1}\}$
- $\mathcal{T}_{k,i}^{\max} = \{$maximal bitiles in $\mathcal{T}_{k,i}$ w.r.t. the given partial ordering of $\mathcal{B}\}$.

Each set $\mathcal{T}_{k,i}$ is called a forest, and for $R \in \mathcal{T}_{k,i}^{\max}$ we define the tree $\mathcal{T}_{k,i,R} = \{P \in \mathcal{T}_{k,i} | P \prec R\}$ and call R the tree top. One can easily check using the definition of the density d that if $P_1, P_2 \in \mathcal{T}_{k,i,R}$ and $P \in \mathcal{B}$ is such that $P_1 \prec P \prec P_2$ then $P \in \mathcal{T}_{k,i,R}$. We call a set of bitiles with this property convex.

Let $P = Q \times \{n, \ldots, n' - 1\}$ be a bitile. We split P in to an lower tile $l_P = Q \times \{n, \ldots, (n + n')/2 - 1\}$ and an upper tile $u_P = Q \times \{(n + n')/2, \ldots, n'\}$, and let E_P be the set of all points (x, n) contained in the lower tile of P, such that $(x, b(x))$ is contained in the upper tile of P.

Then we have the following combinatorial type lemma.

Lemma A.1 ([18]). 1. *The union*

$$\bigcup_{P \in \mathcal{B}} E_P$$

is a partition of E_b.

2. *Let E be a disjoint union of tiles, and let $\mathbf{p}$ be the collection of all tiles that are subsets of E. Then E is the disjoint union of the minimal (maximal) tiles in $\mathbf{p}$.*
3. *The union of a finite convex collection of bitiles can be written as a disjoint union of tiles.*
4. *Let p be a tile and E a subset of the phase plane such that $p \subset E$. If E can be written as a union of tiles, then $E \backslash p$ can be written as a union of tiles.*

We let $T_P = \Pi_{E_P}$, and from Lemma A.1.1 we obtain the finite decomposition $\Pi_{E_b} = \sum_{P \in \mathcal{B}} T_P$ (the sum is finite since b is bounded). For finite subsets $\Xi \subset \mathcal{B}$ we use the notation T_Ξ to denote the operator $\sum_{P \in \Xi} T_P$.

We note that any bitile in $\mathcal{T}_k$ is dominated by at least one maximal bitile or else we could obtain an infinite sequence of associated time intervals $\{\Omega_{P_k}\}_{k=1}^\infty \subset Q$ with $|\Omega_{P_k}| = 2^k$ which is impossible since $|Q| = 1$. The same argument shows that each bitile in $\mathcal{T}_{k,i}$ is dominated by at least one bitile in $\mathcal{T}_{k,i}^{\max}$. Thus, $\mathcal{T}_k$ is partitioned by the forests contained in it, and each forest is the union of its trees. The trees actually form a partition of of the forest, which can be deduced as follows. Suppose a bitile $P \in \mathcal{T}_{k,i}$ is smaller than the two distinct tree tops R_1 and R_2. Then $\Omega_P \subset \Omega_{R_1} \cap \Omega_{R_2} \neq \emptyset$. Notice that by the definition of $\mathcal{T}_{k,i}$ there are less than 2^{i+1} bitiles of $\mathcal{T}_k^{\max}$ greater than P, but at least 2^i of them greater than each of the tree tops, so that there must be a bitile M greater than both tree tops, which means that $\omega_M \subset \omega_{R_1} \cap \omega_{R_2} \neq \emptyset$ so R_1 and R_2 are comparable and thus equal since they are maximal. Hence the partition $\mathcal{T}_{k,i} = \cup_{R \in \mathcal{T}_{k,i}^{\max}} \mathcal{T}_{k,i,R}$ and we obtain the corresponding decomposition of the Carleson operator

$$\Pi_{E_b} = \sum_{i\geq 0, k\in\mathbb{Z}, R\in\mathcal{T}_{k,i}^{\max}} T_{\mathcal{T}_{k,i,R}}.$$

The following two Lemmas will provide the estimates on "tree operators" we need to prove the Theorem.

Lemma A.2. *For $q \in (1,\infty)$ there is a constant C_q such that for every tree $\mathcal{T}_{k,i,R}$ we have*

$$\|T_{\mathcal{T}_{k,i,R}}\|_q \leq C_q.$$

Proof. Define $\mathbf{T}_l = \{P \in \mathcal{T}_{k,i,R} | \, l_P \cap l_R = \emptyset\}$ and $\mathbf{T}_u = \mathcal{T}_{k,i,R} \backslash \mathbf{T}_l$. Clearly, $T_{\mathcal{T}_{k,i,R}} = T_{\mathbf{T}_l} + T_{\mathbf{T}_u}$ and we will handle each of the terms separately.

First we consider $T_{\mathbf{T}_u}$. Take $P, P' \in \mathbf{T}_u$ with $P \neq P'$. We claim that $u_P \cap u_{P'} = \emptyset$. The only nontrivial case of the claim is when P and P' are comparable, say $P \prec P' \prec R$. But then P, P', and R have a common nonempty intersection necessarily contained in $l_P \cap l_{P'}$ by the definition of $\mathbf{T}_u$. It follows that $\omega_{P'} \subset \omega_P$ and the inclusion is strict since $P \neq P'$. Thus $\omega_{P'} \subset \omega_{l_P}$ so u_P and $u_{P'}$ are disjoint as claimed. It follows that $T_P f$ and $T_{P'}$ are supported on disjoint sets. Recall that for any tile p there is exactly one generalized Walsh wavelet packet $\mathcal{W}_p$ with time-frequency support equal to p. Hence,

$$\Pi_p f(x) = \chi_{Q_p}(x) \sum_{Q:|Q||\omega_p|=1} \langle f, \mathcal{W}_{Q\times\omega_p}\rangle \mathcal{W}_{Q\times\omega_p} = \langle f, \mathcal{W}_{Q_p\times\omega_p}\rangle \mathcal{W}_{Q_p\times\omega_p}, \tag{A.1}$$

from which we get

$$|T_P f(x)| \leq |\Pi_{l_p} f(x)| = |\langle f, \mathcal{W}_{l_p}\rangle \mathcal{W}_{l_p}(x)| \leq \frac{1}{|Q_{l_p}|}\int_{Q_{l_p}} |f(y)|\,dy \leq CMf(x), \tag{A.2}$$

where we have used that A is almost isotropic which implies that the sets Q_{l_p} have bounded eccentricity so there exists an d-ball B centered at x such that $Q_{l_p} \subset B$ and $|Q_{l_p}| \geq c|B|$ with c independent of p. We conclude that $\sum_{P\in\mathbf{P}_u} T_P f(x)$ can be bounded pointwise by a constant times $Mf(x)$.

Next, we turn to $T_{\mathbf{T}_l}$. Pick a frequency $N \in l_R$, and let $T_N = \Pi_{\{(x,n)|n<N\}}$. Notice that $\|T_N\|_q \leq C_q$ by Lemma 3.6. Suppose we can find two tiles p and p' such that

$$T_{\mathbf{T}_l} f(x) = (\Pi_p T_N f)(x) - (\Pi_{p'} T_N f)(x). \tag{A.3}$$

Then using the same argument as above we can bound $T_{\mathbf{T}_l} f(x)$ by $2CMT_N f(x)$ which will prove the Lemma.

Suppose $T_{\mathbf{T}_l} f(x) \neq 0$, and define $E_x = \{n|(x,n) \in \mathbf{P}_l\}$. We let P be a minimal bitile in $\mathbf{P}_l$ such that $(x, b(x)) \in u_P$ and let P' be a maximal bitile with the same property, and define $p = \Omega_p \times \omega_P$ where Ω_p is defined such that p is a tile and $x \in \Omega_p$, and we let $p' = u_{P'}$. The decomposition (A.3) will follow at once if we can prove that $\tilde{E}_x = \{n|\, n < N, n \in \omega_p, \text{ and } n \notin \omega_{p'}\}$ equals E_x. Given $(x,n) \in E_U$ with $U \in \mathbf{P}_l$. Then $(x, b(x)) \in u_U$ and $(x,n) \in l_U$. Moreover, $U \prec R$ so $\omega_R \subset \omega_U$ which implies that $(x, N) \in u_U$ (note that $(x,N) \notin l_U$ since $U \in \mathbf{P}_l$). Hence $n < N$ and $\omega_U \subset \omega_p$ since $\omega_p = \Omega_P$ and $P \prec U$ so $n \in \omega_p$. Also, $(x, b(x)) \in u_U \cap u_{P'} \neq \emptyset$ so $\omega_{u_{P'}} \subset \omega_{u_U}$ since $U \prec P'$. But $n \in \omega_{l_U}$ so $n \notin \omega_{p'} \subset \omega_{u_U}$. Hence $E_x \subset \tilde{E}_x$. Conversely, given $n \in \tilde{E}_x$. Then $n < N$ and $\{(x,N),(x,b(x))\} \subset u_{P'}$ but $(x,n) \notin u_{P'}$. Thus, $n < b(x)$ and we can find a bitile V such that $(x,n) \in E_V$ satisfying $P \prec V \prec P'$ so $V \in \mathcal{T}_{k,i,R}$ by convexity. It also follows that $V \in \mathbf{P}_l$ which implies $\tilde{E}_x \subset E_x$ and we are done. ∎

Lemma A.3. *For $q \in (1,\infty)$ there is a constant C_q such that for every tree $\mathcal{T}_{k,i,R}$,*

$$\|T_{\mathcal{T}_{k,i,R}} f\|_q \leq C2^k|\Omega_R|^{1/q},$$

where C does not depend on the fixed function f.

Proof. The area E of the tree $\mathcal{T}_{k,i,R}$ is a convex union of bitiles so it follows from Lemma A.1.3 that E can be written as a disjoint union of tiles. $E\backslash l_P$ is also a disjoint union of tiles, so using (A.1) we obtain that for $P \in \mathcal{T}_{k,i,R}$ the projections $\Pi_{E\backslash l_P}$ and Π_{l_P} are orthogonal. Hence,

$$\Pi_{l_P}\Pi_E = \Pi_{l_P}(\Pi_{l_P} + \Pi_{E\backslash l_P}) = \Pi_{l_P}\Pi_{l_P} = \Pi_{l_P},$$

and we deduce that $T_P f(x) = T_P \Pi_E f(x)$. Consequently $T_{\mathcal{T}_{k,i,R}} f = T_{\mathcal{T}_{k,i,R}} \Pi_E f$ and $\|T_{\mathcal{T}_{k,i,R}} f\|_q \leq \|T_{\mathcal{T}_{k,i,R}}\|_q \|\Pi_E f\|_q$. The support of $\Pi_E f$ is contained in Ω_R. Fix $x \in \Omega_R$ and let P be the minimal bitile in the tree containing x. Then ω_P is exactly the frequencies n such that $(x, n) \in E$. To see this we suppose $(x, n) \in E$. Then there is a bitile P' containing (x, n). Since P' and P are smaller than R, their frequency intervals both contain a point $\tilde{n} \in \omega_R$. Hence P and P' are comparable and $P \prec P'$ by the definition of P. Thus $(x, n) \in P$. The opposite inclusion is trivial. Hence, $\Pi_E f(x) = \Pi_P f(x)$ so from the densities of the bitiles in $\mathcal{T}_{k,i,R}$ we get $\|\Pi_E f\|_\infty \leq 2^{k+1}$. Using that the support of Π_E is contained in Ω_R we get the estimate $\|\Pi_E f\|_q \leq 2^{k+1}|\Omega_R|^{1/q}$. Combined with the previous lemma this gives us $\|T_{\mathcal{T}_{k,i,R}} f\|_q \leq C 2^k |\Omega_R|^{1/q}$. ■

Completion of the proof. The area of two distinct trees $\mathcal{T}_{k,i,R_1}$ and $\mathcal{T}_{k,i,R_2}$ from the same forest are clearly disjoint so we have, for $q > 0$,

$$|T_{\mathcal{T}_{k,i}} f(x)|^q = \sum_{R \in \mathcal{T}_{k,i}^{\max}} |T_{\mathcal{T}_{k,i,R}} f(x)|^q,$$

which combined with Lemma A.3 implies

$$\|T_{\mathcal{T}_{k,i}} f\|_q \leq C 2^k \left(\sum_{R \in \mathcal{T}_{k,i}^{\max}} |\Omega_R| \right)^{1/q} \tag{A.4}$$

For $P \in \mathcal{T}_k^{\max}$ consider the bitiles R in $\mathcal{T}_{k,i}^{\max}$ which are smaller than P. The time intervals of these bitiles are contained in Ω_P and must be pairwise disjoint because otherwise the frequency intervals of two such bitiles with nonempty intersection would both contain ω_P and thus make two of the bitiles in $\mathcal{T}_{k,i}^{\max}$ comparable, which is clearly not the case. This observation gives us the following estimate

$$\sum_{R \in \mathcal{T}_{k,i}^{\max} : R \prec P} |\Omega_R| \leq |\Omega_P|.$$

We add this inequality up for all the bitiles $P \in \mathcal{T}_k^{\max}$, using the fact that each $R \in \mathcal{T}_{k,i}^{\max}$ dominates at least 2^i bitiles from $\mathcal{T}_k^{\max}$, to obtain

$$2^i \sum_{R \in \mathcal{T}_{k,i}^{\max}} |\Omega_R| \leq \sum_{P \in \mathcal{T}_k^{\max}} |\Omega_P|. \tag{A.5}$$

Next, we observe that any tile p we have the important property that $\|\Pi_p f\|_2^2 = \|\Pi_p f\|_\infty^2 |\Omega_p|$, which follows from (A.1) Thus for any bitile P,

$$\begin{aligned} 2\|\Pi_P f\|_2^2 &= 2(\|\Pi_{u_P} f\|_2^2 + \|\Pi_{l_P} f\|_2^2) \\ &= 2|\Omega_P|((\|\Pi_{u_P} f\|_\infty^2 + \|\Pi_{l_P} f\|_\infty^2) \\ &\geq \|\Pi_P f\|_\infty^2 |\Omega_P|. \end{aligned}$$

So from the fact that the time intervals of the bitile in $\mathcal{T}_k^{\max}$ are pairwise disjoint we have

$$\|f\|_2^2 \geq \sum_{P \in \mathcal{T}_k^{\max}} \|\Pi_P f\|_2^2 \geq \sum_{P \in \mathcal{T}_k^{\max}} \frac{1}{2} \|\Pi_P f\|_\infty^2 |\Omega_P| \geq \frac{1}{2} 2^{2k} \sum_{P \in \mathcal{T}_k^{\max}} |\Omega_P|, \tag{A.6}$$

where we used the definition of the density of the tiles in $\mathcal{T}_k^{\max}$. We use (A.5) in (A.6) and combine with (A.4) to conclude that

$$\|T_{\mathcal{T}_{k,i}} f\|_q \leq C 2^k \|f\|_2^{2/q} 2^{-(2k+i)/q}.$$

Fix $K \in \mathbb{Z}$, and let $q > 2$. We add all bitiles with density less than or equal to K to get

$$\Big\| \sum_{P:a_P \leq K} T_P f \Big\|_q \leq C \|f\|_2^{2/q} \sum_{k<K, i\geq 0} \frac{2^{k(1-2/q)}}{2^{i/q}} \leq C \|f\|_2^{2/q} 2^{K(1-2/q)}, \tag{A.7}$$

from which we obtain the following weak estimate

$$\big|\big\{x : \big| \sum_{P:a_P \leq K} T_P f(x) \big| > 2^K \big\}\big| \leq C \|f\|_2^2 \frac{2^{K(q-2)}}{2^{Kq}} = C \frac{\|f\|_2^2}{2^{2K}} \tag{A.8}$$

To get the general result we follow R. Hunt and verify that restricted type inequalities holds for the Carleson operator, and then use interpolation of the restricted type inequalities (see e.g. [16, Chap. V]) to get the full result. Let us suppose $f = \chi_\Omega$, $\Omega \subset Q$. Then $\|f\|_2^2 = \|f\|_p^p$ for $1 < p < \infty$. Notice that no bitile can have density larger than 1 so taking taking $K = 1$ in (A.7) immediately gives us the bound $\|T_P f\|_p \leq C \|f\|_p$, which is the required restricted inequality. For $1 < p < 2$ we put $r - p = p(r - s)$ in (A.8) to get

$$\big|\big\{x : \big| \sum_{P:a_P \leq pK} T_P f(x) \big| > 2^{pK} \big\}\big| \leq C \frac{\|f\|_2^2}{2^{pK}} = C \frac{\|f\|_p^p}{2^{pK}}.$$

Next, consider $g = \sum_{P:a_P > pK} T_P f$. If x is in the support of g then x is contained in the time interval of some bitile with density larger than pK, and it follows from (A.2) that $Mf(x) > C 2^{pK}$. Hence

$$|\{x : |g(x)| > 2^{pK}\}| \leq |\{x : Mf(x) > C 2^{pK}\}| \leq C \frac{\|f\|_1}{2^{pK}} = \frac{\|f\|_p^p}{2^{pK}}.$$

The strong estimate now follows by interpolation. ∎

References

[1] E. Belogay and Y. Wang. Arbitrarly smooth orthogonal nonseparable wavelets in $\mathbb{R}^2$. *SIAM J. Math. Anal.*, 30(3):678–697, 1999.

[2] P. Billard. Sur la convergence presque partout des séries de Fourier-Walsh des fontions de l'espace $L^2(0,1)$. *Studia Math.*, 28:363–388, 1967.

[3] R. R. Coifman, Y. Meyer, S. R. Quake, and M. V. Wickerhauser. Signal processing and compression with wavelet packets. In Y. Meyer and S. Roques, editors, *Progress in Wavelet Analysis and Applications*, 1992.

[4] I. Daubechies. *Ten Lectures on Wavelets*. SIAM, 1992.

[5] B. Golubov, A. Efimov, and V. Skvortsov. *Walsh Series and Transforms*. Kluwer Academic Publishers, 1991.

[6] J. C. Lagarias and Y. Wang. Corrigendum and addendum to: Haar bases for $L^2(\mathbb{R}^n)$ and algebraic number theory. *J. Number Theory*, 57:181–197, 1996.

[7] J. C. Lagarias and Y. Wang. Haar bases for $L^2(\mathbb{R}^n)$ and algebraic number theory. *J. Number Theory*, 57:181–197, 1996.

[8] F. G. Meyer and R. R. Coifman. Brushlets: a tool for directional image analysis and image compression. *Appl. Comput. Harmon. Anal.*, 4(2):147–187, 1997.

[9] Y. Meyer. *Ondelettes et Opérateurs II: Opérateurs de Calderón-Zygmund*. Hermann, 1990.

[10] Y. Meyer. *Wavelets and Operators*. Cambridge University Press, 1992.

[11] M. Nielsen. *Size Properties of Wavelet Packets*. PhD thesis, Washington University, St. Louis, 1999.

[12] M. Nielsen. Walsh type wavelet packet expansions. *Appl. and Comp. Harmonic Anal.*, to appear.

[13] F. Schipp, W. R. Wade, and P. Simon with assistance from J. Pál. *Walsh Series*. Adam Hilger, 1990.

[14] P. Sjölin. An inequality of Paley and convergence a.e. of Walsh-Fourier series. *Arkiv för Matematik*, 7:551–570, 1969.

[15] E. M. Stein. *Singular Integrals and Differentiabillity Properties of Functions.* Princeton University Press, 1970.

[16] E. M. Stein and G. Weiss. *Introduction to Fourier Analysis on Euclidean Spaces.* Princeton Univ. Press. 1971.

[17] T. Tao. On the almost everywhere convergence of wavelet summation methods. *Appl. and Comp. Harmonic Anal.*, 3:384–387, 1996.

[18] C. Thiele. The quartile operator and pointwise convergence of Walsh series. *Trans. Amer. Math. Soc.*, to appear.

[19] P. Wojtaszczyk. *A Mathematical Introduction to Wavelets.* Cambridge University Press, 1997.

Department of Mathematics, University of South Carolina, South Carolina 29208
E-mail address: nielsen@math.sc.edu