View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Apollo

ON THE STABLE SAMPLING RATE FOR BINARY MEASUREMENTS AND WAVELET RECONSTRUCTION

A. C. HANSEN AND L. TERHAAR

ABSTRACT. This paper is concerned with the problem of reconstructing an infinite-dimensional signal from a limited number of linear measurements. In particular, we show that for binary measurements (modelled with Walsh functions and Hadamard matrices) and wavelet reconstruction the stable sampling rate is linear. This implies that binary measurements are as efficient as Fourier samples when using wavelets as the reconstruction space. Powerful techniques for reconstructions include generalized sampling and its compressed versions, as well as recent methods based on data assimilation. Common to these methods is that the reconstruction quality depends highly on the subspace angle between the sampling and the reconstruction space, which is dictated by the stable sampling rate. As a result of the theory provided in this paper, these methods can now easily use binary measurements and wavelet reconstruction bases.

1. INTRODUCTION

Reconstructing infinite-dimensional signals from a limited number of linear measurements is a key problem in sampling and approximation theory, and has received substantial attention over the last decades due to its many applications. The list of fields is comprehensive and includes Magnetic Resonance Imaging (MRI) [30, 41], electron tomography [38, 39], lensless cameras, fluorescence microscopy [49, 51], X-ray computed tomography [17, 48], surface scattering [35] among others. Efficient methods for such problems date back to Shannon's sampling theorem [34, 50, 52], however, over the years, more modern approaches have been developed. Indeed, new methods include generalized sampling, which has been studied by Adcock, Hansen, Hrycak, Gröchenig, Kutyniok, Ma, Poon, Shadrin and others [1, 2, 4, 6, 32, 33, 42], its compressed versions investigated by Adcock, Hansen, Kutyniok, Lim, Poon and Roman [5, 7, 37, 47] as well as the predecessor; consistent sampling, analysed by Aldroubi, Eldar, Unser and others [8, 23–26, 53]. Note that consistent sampling is very much related to the finite section method [13, 29, 31, 40]. More recently, new methods based on data assimilation have been successfully developed and analysed. A first approach for the same number of measurements M and reconstructed coefficients N was introduced under the name *generalized empirical interpolation method* by Maday, Patera, Penn and Yano in [44]. This was then further extended to *Parametrized-Background Data-Weak (PBDW) approach* with $M \geq N$ in [43, 45]. The PBDW approach was additionally analysed and shown to be optimal by Binev, Cohen, Dahmen, DeVore, Petrova, and Wojtaszczyk [10, 11, 21].

The problem is given as follows. An element $f \in \mathcal{H}$, where $\mathcal{H}$ is a separable Hilbert space, is to be reconstructed from measurements with linear functionals $(m_i)_{i\in\mathbb{N}} : \mathcal{H} \to \mathbb{C}$ that can be represented by elements $s_i \in \mathcal{H}$ as $m_i(f) = \langle f, s_i\rangle$. The key issue is that the m_i cannot be chosen freely, but are dictated by the modality of the sampling device, for example a Magnetic Resonance Imaging (MRI) scanner providing Fourier samples or a fluorescence microscope giving binary measurements. The goal is to reconstruct f from the finite number of samples $\{m_i(f)\}_{i=1}^M$ for some $M \in \mathbb{N}$. The space of the functions s_i is called the *sampling space* and is denoted by $\mathcal{S} = \overline{\text{span}}\{s_i : i \in \mathbb{N}\}$, meaning the closure of the span. In practice, one can only acquire a finite number of samples. Therefore, we denote by $\mathcal{S}_M = \text{span}\{s_i : i = 1, \ldots, M\}$ the sampling space of the first M elements. The reconstruction is typically done via a reconstruction space denoted by $\mathcal{R}$ and spanned by reconstruction functions $(r_i)_{i\in\mathbb{N}}$, i.e. $\mathcal{R} = \overline{\text{span}}\{r_i : i \in \mathbb{N}\}$. As in the case

Key words and phrases. Sampling theory, Generalized sampling, Wavelets, Walsh functions, Stable Sampling rate, Data assimilation, Hilbert spaces.

of the sampling space, it is impossible to acquire and save an infinite number of reconstruction coefficients. Hence, one has to restrict to a finite reconstruction space, which is denoted by $\mathcal{R}_N = \operatorname{span}\{r_i : i = 1, \dots, N\}$. The key is that the r_i can be tailored to the desired signal type to be reconstructed. For example, spaces spanned by X-lets (wavelets, curvelets, contourlets, shearlets) [14–16, 19, 20, 22, 36, 46] may be preferable as reconstruction spaces in imaging applications, whereas polynomials may be useful when recovering very smooth functions.

The methods mentioned above can be described as follows: for $f \in \mathcal{H}$ and $N, M \in \mathbb{N}$, we define the reconstruction method of *generalized sampling* $G_{N,M} : \mathcal{H} \to \mathcal{R}_N$ by

$$\langle P_{\mathcal{S}_M} G_{N,M}(f), r_j \rangle = \langle P_{\mathcal{S}_M} f, r_j \rangle, \quad r_j \in \mathcal{R}_N,$$

where $P_{\mathcal{S}_M}$ denotes the orthogonal projection on the subspace $\mathcal{S}_M$. Note that the stability and accuracy of this method depends on the subspace angle between the sampling and the reconstruction space, i.e.

$$\|f - G_{N,M}(f)\| \leq \mu(\mathcal{R}_N, \mathcal{S}_M) \|f - P_{\mathcal{R}_N} f\|,$$

where we define the subspace angle between closed subspaces $U, V \in \mathcal{H}$

$$\cos(\omega(U,V)) := \frac{1}{\mu(U,V)} := \inf_{u \in U, \|u\|=1} \|P_V u\|,$$

$\omega(U,V) \in [0, \pi/2]$. Moreover, the condition number κ of $G_{N,M}$ is also given by $\kappa(G_{N,M}) = \mu(\mathcal{R}_N, \mathcal{S}_M)$. For the PBDW method one calculates

$$F_{N,M}(f) = \operatorname*{argmin}_{u \in P_{\mathcal{S}_M} f + \mathcal{S}_M^\perp} \|u - P_{\mathcal{R}_N} u\|,$$

and it can be shown that the accuracy then depends on subspace angle as follows

$$\|f - F_{N,M}(f)\| \leq \mu(\mathcal{R}_N, \mathcal{S}_M) \operatorname{dist}(f, \mathcal{R}_N \oplus (\mathcal{S}_M \cap \mathcal{R}_N)^\perp).$$

Moreover, this is sharp because the constant $\mu(\mathcal{R}_N, \mathcal{S}_M)$ cannot be improved. It is clear that, in both approaches, the key to success lies in the ability to make sure that

$$\mu(\mathcal{R}_N, \mathcal{S}_M) \leq \theta, \qquad \theta \in (1, \infty).$$

Thus, we need to balance the number of samples M with the number of reconstruction vectors N, and this leads to the so-called *stable sampling rate*:

$$\Theta(N, \theta) = \min\{M \in \mathbb{N} : \mu(\mathcal{R}_N, \mathcal{S}_M) \leq \theta\}.$$

The methods above can only be used efficiently when the stable sampling rate is known and reasonable. In particular, numerical calculations of the stable sampling rate are very time consuming. Moreover, if the stable sampling rate is worse than linear, the approximation quality of the reconstruction space must allow for rapid approximation to compensate for the "slow" sampling rate. Fortunately, it is possible to obtain sharp results on describing $\Theta(N, \theta)$ for popular sampling and reconstruction spaces, and often, especially for the reconstruction with X-lets, one can establish linearity. In this paper we do so for sampling with Walsh functions and reconstructing with wavelets. Remark that this is not always the case. For example in the case of the reconstruction with Legendre polynomials, the stable sampling rate is known to be quadratic for Fourier measurements [33]. Additionally, for the PBDW-method there are approaches where the reconstruction space is fixed by the PDE and the sampling space is chosen adaptively. This is a very different setting from the one discussed here. Details and convergence rates can be found in [12].

1.1. **Connection to previous work and novelty of the paper.** The stable sampling rate is well understood when the samples $m_i(f) = \langle f, s_i \rangle$ are Fourier measurements. In other words, the s_i are complex exponentials and $m_i(f)$ are the Fourier coefficients. In this case the stable sampling rate is linear for many X-lets including wavelets and shearlets. Fourier samples and X-lets are a natural starting point given the vast applications that are based on Fourier measurements (MRI, tomography problems with parallel beam, surface scattering, radio interferometry etc.). However, the next question regards binary measurements. By binary measurements we mean that the sampling functions s_i can only take two values either $\{0, 1\}$ or $\{-1, 1\}$. Without loss of generality we can assume that the model uses $\{-1, 1\}$, as one can, by adding one extra measurement with the constant function, convert from the $\{0, 1\}$ setup to the $\{-1, 1\}$ model.

Binary measurements are a mainstay in signal and image processing due to the "on-off" nature of many physical sampling devices. Microscopy is an obvious application as well as the newly emerging techniques of lensless cameras. In the discrete setting binary measurements are often modelled with Hadamard matrices, and this is one of the reasons why Hadamard matrices are so important in signal processing. To model binary measurements we change the model from Fourier samples $\langle f, s_i \rangle$, where the s_i are complex exponentials to letting the s_i be Walsh functions. The Walsh functions are the binary counterpart to Fourier samples and complex exponentials. Thus, the key question is as follows: what is the stable sampling rate when sampling with Walsh functions and reconstructing with wavelets? The answer is that it is linear regardless of the dimension when we consider separable boundary wavelets. This means that sampling with binary measurements is as efficient (up to potentially a different constant) as sampling with Fourier samples when reconstructing with wavelets. We expect the techniques used in this paper to extend to other X-lets as well; however, the extension, as in the Fourier case, is non-trivial.

1.2. **Main Theorem.** We consider the sampling space $\mathcal{S}$ of Walsh functions, which will be described in more detail in Chapter 2, and let the reconstruction space $\mathcal{R}$ be the space of boundary-corrected Daubechies wavelets (see Chapter 3 for details). The main theorem states that the stable sampling rate is indeed linear in N.

Theorem 1.1. *Let $\mathcal{S}$ and $\mathcal{R}$ be the sampling and reconstruction space spanned by the d-dimensional Walsh functions and separable boundary wavelets respectively. Moreover, let $N = 2^{dR}$ with $R \in \mathbb{N}$. Then for all $\theta \in (1, \infty)$ there exists S_θ such that for all $M \geq 2^{dR} S_\theta$ we have $\mu(\mathcal{R}_N, \mathcal{S}_M) \leq \theta$. In particular, one gets $\Theta \leq S_\theta N$. Hence, the relation $\Theta(N; \theta) = \mathcal{O}(N)$ holds for all $\theta \in (1, \infty)$.*

2. Walsh functions - Defining the Sampling Space $\mathcal{S}_M$

Due to the fact that we are dealing with the d-dimensional case, we introduce multi-indices to make the notation more readable. Let $j = \{j_1, \ldots, j_d\} \in \mathbb{N}^d$, $d \in \mathbb{N}$ be a multi-index. A natural number n is in the context of a multi-index interpreted as a multi-index with the same entry, i.e. $n = \{n, \ldots, n\}$. Then we define the addition of two multi-indices for $j, r \in \mathbb{N}^d$ by the pointwise addition, i.e. $j + r = \{j_1 + r_1, \ldots, j_d + r_d\}$ and the sum

$$\sum_{j=k}^{r} x_j := \sum_{j_1=k_1}^{r_1} \ldots \sum_{j_d=k_d}^{r_d} x_{j_1,\ldots,j_d},$$

where $k, r \in \mathbb{N}^d$. The multiplication of an multi-index with a real number is understood pointwise, as well as the division by a multi-index. The absolute value of a multi-index j is given by $|j| = j_1 + \ldots + j_d$. The d dimensional functions that we use in this paper are constructed by the tensor product. For a function $f : \mathbb{R} \to \mathbb{R}$ and an input parameter $\{x_i\}_{i=1,\ldots,d} = x \in \mathbb{R}^d$ with $x_i \in \mathbb{R}$ we use the following notation to present the d-times tensor product of f, i.e.

$$f(x) = f(x_1) \otimes \ldots \otimes f(x_d) \text{ (d-times)}.$$

It should be clear from the input parameter, whether f represents the function on $\mathbb{R}$ or $\mathbb{R}^d$.

2.1. **Defining Walsh functions.** The key property that makes Walsh functions attractive in many applications is that they take only the values 1 and -1. However, as Walsh functions are defined in the dyadic analysis, some properties only hold for dyadic addition. Recalling the basics of dyadic addition, we represent elements $x \in \mathbb{R}_+$ with their dyadic representation as follows

$$x = \sum_{i \in \mathbb{Z}} x_i 2^i,$$

where $x_i \in \{0, 1\}$ for all $i \in \mathbb{Z}$. The natural extension always ends in 0 for dyadic rational numbers and is infinite for dyadic irrational numbers. The representation is therefore unique. Elements of $\mathbb{R}_-$ are represented as in the decimal analysis with an additional $-$ in front of the representation. In the dyadic analysis the addition $\oplus : \mathbb{R}_+ \mapsto \mathbb{R}_+$ is defined by

$$x \oplus y = \sum_{i \in \mathbb{Z}} (x_i \oplus_2 y_i) 2^i,$$

where $x_i \oplus_2 y_i$ is addition modulo two, i.e. $0 \oplus_2 0 = 0, 0 \oplus_2 1 = 1, 1 \oplus_2 0 = 1, 1 \oplus_2 1 = 0$. For negative numbers one has $-x \oplus y = x \oplus -y = -(x \oplus y)$. Note that there is no closed form between the decimal and dyadic addition. In particular, for two numbers $x, h \in \mathbb{R}$ the expression of the decimal sum $x + h$ has a different expression in the dyadic addition for every pair of numbers. This leads to further investigation during the proof of the main theorem. One part on the way to control the subspace angle is not to deal with all timeshifts of the wavelet but instead transfer them to the Walsh function and deal with the Walsh polynomial. Unfortunately, all properties of Walsh functions rely on dyadic rather than decimal addition. Therefore, this difference of the additions needs special care. In chapter 4.1 we will see that under mild assumptions the additions can made be equal. In chapter 4.2 we tweak the wavelets to match these assumptions.

Now, we present the Walsh functions, which are used to represent the sampling space $\mathcal{S}_M$. Therefore, we give a definition of the classical Walsh functions that highlights the difference between the possible orderings.

Definition 2.1 ([27]). *Let $s \in \mathbb{N}$ and $x \in [0, 1)$. Then there exists a unique $n = n(s) \in \mathbb{N}$ such that $s = \sum_{i=0}^{n-1} s_i 2^i$, in particular $s_{n-1} \neq 0$ and $s_k = 0$ for all $k \geq n$. Let $s^n = \{s_0, \ldots, s_{n-1}\}$ and for $x = \sum_{i=-\infty}^{-1} x_i 2^i$ define $x^n = \{x_{-n}, \ldots, x_{-1}\}$, and $\omega_W \colon \mathbb{R}^n \mapsto \mathbb{R}^n$ by*

$$\omega_W = \begin{pmatrix} 0 & \cdots & 0 & 1 & 1 \\ \vdots & \iddots & \iddots & 1 & 0 \\ 0 & \iddots & \iddots & \iddots & \vdots \\ 1 & 1 & \iddots & & \vdots \\ 1 & 0 & \cdots & \cdots & 0 \end{pmatrix}.$$

The Walsh functions *are then given by*

$$\mathrm{wal}(s; x) = (-1)^{s^n \cdot \omega_W x^n}.$$

By changing the matrix ω_W one gets different orderings of the Walsh functions. For example, the identity matrix leads to the Walsh-Kronecker functions, which have the drawback that with a change of $n(s)$ all functions are altered, hence one has to fix the maximal s in advance. The Walsh-Paley ordering is obtained by replacing ω_W with the reversal matrix, i.e.

$$\omega_{WP} = \begin{pmatrix} 0 & \cdots & 0 & 0 & 1 \\ \vdots & \iddots & \iddots & 1 & 0 \\ 0 & \iddots & \iddots & \iddots & 0 \\ 0 & 1 & \iddots & \iddots & \vdots \\ 1 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

They overcome the previous problem, but the functions are not ordered such that the number of zero crossings increases with s. Both drawbacks are overcome with the Walsh-Kaczmarz ordering presented in the previous definition.

The classical Walsh functions can be extended to the *generalized Walsh functions* $\mathrm{Wal} : \mathbb{R}_+^2 \to \{-1, 1\}$ which are defined with the classical Walsh functions and the periodic continuation with period 1 by

$$\mathrm{Wal}(s, x) = (-1)^{s_0 x_0} \,\mathrm{wal}([s]\,; x)\,\mathrm{wal}([x]\,; s),$$

where s and x have the dyadic representation $(s_i)_{i\in\mathbb{Z}}$ and $(x_i)_{i\in\mathbb{Z}}$ and s_0, x_0 are the corresponding elements of the sequence. This extension can also be defined by letting ω_W be infinite, i.e. be defined over $\mathbb{Z}$ instead of $\mathbb{N}$ and hence allow inputs with infinite dyadic representations over $\mathbb{Z}$. Moreover, the Walsh functions can also be extended to negative inputs. Therefore, we define the following equality as in [27]

$$\mathrm{Wal}(-s, x) := -\,\mathrm{Wal}(s, x)$$
$$\mathrm{Wal}(s, -x) := -\,\mathrm{Wal}(s, x).$$

The Walsh functions in higher dimensions are obtained by the tensor product, i.e. for $s = \{s_k\}_{k=1,\dots,d}$, $x = \{x_k\}_{k=1,\dots,d} \in \mathbb{R}^d$

$$\mathrm{Wal}(s, x) = \bigotimes_{k=1}^{d} \mathrm{Wal}(s_k, x_k).$$

The Walsh functions can also be combined to *Walsh polynomials* similar to trigonometric polynomials.

Definition 2.2. *Let $A, B \in \mathbb{Z}^d$ such that $A_i \le B_i, i = 1, \dots, d$ and $\alpha_{j_i} \in \mathbb{R}$. Then for $z \in \mathbb{R}_+^d$ we define the* Walsh polynomial *of order $n = |B|$ by $\Phi(z) = \sum_{j=A}^{B} \alpha_j \,\mathrm{Wal}(j, z)$. The set of all Walsh polynomials up to degree n is given by*

$$WP_n = \left\{ \sum_{j=A}^{B} \alpha_j \,\mathrm{Wal}(j, z), \alpha_{j_i} \in \mathbb{R}, A, B \in \mathbb{Z}^d, |B| \le n \right\}.$$

With the generalized Walsh functions one can define a continuous and discrete transform. To ensure that the following integral exists, let $f \in L^2([0,1]^d)$, the *generalized Walsh transform* is given almost everywhere by

$$\widehat{f}^{W}(s) = \langle f(\cdot), \mathrm{Wal}(s, \cdot)\rangle = \int_{[0,1]^d} f(x)\,\mathrm{Wal}(s, x)dx, \quad s \in \mathbb{R}^d.$$

This is suitable for our setting, because we consider only the Walsh transform of functions that are supported in $[0, 1]^d$. In the discrete setting we have for $N = 2^n$, $n \in \mathbb{N}$ and $x = \{x_0, \dots, x_{N-1}\} \in \mathbb{R}^N$ that the one dimensional *discrete Walsh transform* of x is given by $X = \{X_0, \dots, X_{N-1}\}$ with

$$X_j = \frac{1}{N} \sum_{k=0}^{N-1} x_k \,\mathrm{Wal}(j, \frac{k}{N}).$$

This transform corresponds, as mentioned, to the multiplication with a Hadamard matrix. By the definition of Wal it corresponds to the Hadamard matrix in Walsh-Kaczmarz ordering. The addition here is again the dyadic addition. The discrete d-dimensional Walsh transformed of $x \in \mathbb{R}^{N_1\times\dots\times N_d}$ where $x_{k_i} \in \mathbb{R}$, $k = \{k_i\}_{i=1,\dots,d}$, $k_i = 0, \dots, N_i - 1$ is given by $X = \{X_j\} \in \mathbb{R}^{N_1\times\dots\times N_d}$, where $X_{j_i} \in \mathbb{R}$, $j = \{j_i\}_{i=1,\dots,d}$, $j_i = 0, \dots, N_i - 1$, with

$$X_j = \frac{1}{\prod_{i=1}^{d} N_i} \sum_{k=0}^{N-1} x_k \,\mathrm{Wal}(j, \frac{k}{N}).$$

2.2. **Properties of Walsh functions.** The Walsh functions obey the following properties: They are symmetric,

$$\mathrm{Wal}(s,x) = \mathrm{Wal}(x,s) \text{ for all } s,x \in \mathbb{R},$$

and they obey the *scaling property* as well as the *multiplicative identity*, i.e

$$\mathrm{Wal}(2^k s,x) = \mathrm{Wal}(s,2^k x) \text{ for all } s,x \in \mathbb{R},\ k \in \mathbb{N} \tag{2.1}$$

and

$$\mathrm{Wal}(s,x)\,\mathrm{Wal}(s,t) = \mathrm{Wal}(s,x \oplus t) \text{ for all } s,x \in \mathbb{R}. \tag{2.2}$$

These properties can be directly transferred to the d-dimensional Walsh functions and the continuous Walsh transform, i.e. it holds, that the continuous Walsh transform is linear

$$\mathcal{W}\{af(t) + bg(t)\} = a\mathcal{W}\{f(t)\} + b\mathcal{W}\{g(t)\} \text{ for all } a,b \in \mathbb{R} \text{ and } f,g \in L^2([0,1]^d),$$

and obeys the following *shift* and *scaling property*, i.e.

$$\mathcal{W}\{f(t \oplus x)\}(s) = \mathcal{W}\{f(t)\}(s)\,\mathrm{Wal}(x,s) \text{ for all } x \in \mathbb{R}^d \text{ and } f \in L^2([0,1]^d)$$

and

$$\mathcal{W}\{f(2^m t)\}(s) = \frac{1}{2^m}\mathcal{W}\{f(t)\}\left(\frac{s}{2^m}\right) \text{ for all } m \in \mathbb{N}^d \text{ and } f \in L^2([0,1]^d).$$

3. Wavelets - Defining the Reconstruction Space $\mathcal{R}_N$

3.1. **Boundary Wavelets.**

3.1.1. *Boundary Wavelet space in one dimension.* Daubechies boundary wavelets are deduced from general Daubechies wavelets. They have the advantage that they keep desirable properties, such as smoothness and vanishing moments, from their mother wavelet. In contrast, other approaches to find an orthonormal wavelet basis for $L^2([0,1])$, such as extension with zero, periodising or folding, loose smoothness. For the construction of the Daubechies boundary wavelets, as presented in [18], one starts with the Daubechies scaling functions. First, we deal with the scaling functions on the positive line $[0,\infty)$. Remember that a Daubechies scaling function ϕ of order p has the support $[-p+1,p]$. Then the functions $\phi_n(x) = \phi(x-n)$ have their support completely in $[0,\infty)$ for $n \geq p-1$. However, these functions do not even generate the polynomials on $[0,\infty)$, so they do not represent smooth functions well. Therefore, the following functions are added to circumvent this issue:

$$\widetilde{\phi}_n^{\text{left}}(x) = \sum_{l=0}^{2p-2} \binom{l}{n} \phi(x+l-p+1).$$

It is shown in [18] that these functions together with the inner ones, i.e. the translates of the scaling function, whose support is completely contained in the positive real line, span all polynomials with degree smaller or equal to $p-1$ on $[0,\infty)$. The same construction can be done for the negative line $(-\infty,0]$ and then be shifted by 1 to get to the desired interval. This means in detail that the scaling function on the right hand side can be deduced from those on the left side, i.e. the construction for the right hand side results from a shift of 1 in the functions that do intersect with the right end of the interval and a reflection. We have that

$$\widetilde{\phi}_n^{\text{right}}(x) = \widetilde{\phi}_{-1-n}^{\text{left}}(-x).$$

In the next step we bring both systems together on $[0,1]$. To make sure that each shift of the scaling function is either an inner a left or a right scaling function, we consider scaling functions at a level $j \geq J_0$, where $2^{J_0} \geq 2p-1$. This way the support size of the scaling function at that scale is smaller than 1. Therefore, the scaling function can intersect only with 0 or 1 and hence the correction is well defined. The functions are now all corrected on the boundaries and they span the desired space $L^2([0,1])$. To form an orthonormal basis, we simply apply a Gram-Schmidt procedure. The new functions, after the orthonormalisation, are denoted

by $\phi_n^{\text{left}}, \phi_n^{\text{right}}$. The functions have staggered support, i.e. $\operatorname{supp} \phi_n^{\text{left}} = [0, p+n]$. Therefore, all ϕ have support length at most $2p-1$. Hence, the change to the boundary wavelet preserves the favourable property of a small support size. The dilated boundary scaling functions can be deduced from this construction, like the scaling functions for the real line. With this construction we obtain 2^j+2 scaling functions at every scale j, but in many applications one prefers to have 2^j scaling functions. Therefore, we remove the two outermost interior scaling functions, i.e. those with the support closest to 0 and 1 but not intersecting with them. This results in the subspaces

$$V_j^b = \operatorname{span}\left\{\phi_{j,n}^b : n = 0, \dots 2^j - 1\right\},$$

where

$$\phi_{j,n}^b(x) = \begin{cases} 2^{j/2}\phi_n^{\text{left}}(2^j x) & n = 0, \dots p-1 \\ 2^{j/2}\phi_n(2^j x) & n = p, \dots 2^j - p - 1 \\ 2^{j/2}\phi_{2^j-n-1}^{\text{right}}(2^j(x-1)) & n = 2^j - p, \dots 2^j - 1. \end{cases} \tag{3.1}$$

In [18] it is proven that we can define the wavelet space at every scale j similar to the case on the real line by

$$W_j^b = V_{j+1}^b \cap (V_j^b)^\perp.$$

The original wavelet functions $\psi_{j,k}$ from the real line are in W_j^b for $k = p, \dots, 2^j - p - 1$. Because of the dimension of the scaling space one knows that $\dim W_j^b = 2^j$. Hence, one has to add $2p$ wavelets. As they do not play an important role for the investigation of the main theorem we point the interested reader to [18] for detailed information.

Now that the necessary information about wavelets is introduced, we discuss the reconstruction space. The data is usually sparsely represented in the wavelet scheme, i.e. they only have large coefficients up to a certain scale. Therefore, the reconstruction space contains only the wavelets up to some scale R. Moreover, the low frequency part can be represented by the scaling space at some level $J \geq J_0$. In theory the choice of the lowest level J is free. Nevertheless, it is common to use $J = J_0$. This results in the following reconstruction space. For $R \in \mathbb{N}$, the space of wavelets up to a scaling of R is given by

$$\mathcal{R}_N = V_J^b \oplus W_J^b \oplus \dots \oplus W_{R-1}^b = V_R^b. \tag{3.2}$$

and has $N = 2^R$ elements. Due to the construction the "left" scaling functions are translates of the mother scaling functions and the "right" scaling functions are reflected translated scaling functions, denoted by $\phi^\#$. Therefore,

$$V_R^b = \operatorname{span}\left\{\phi_{R,n} : n = 0, \dots, 2^R - p - 1, \phi_{R,n}^\# : n = 2^R - p, \dots, 2^R - 1\right\} \tag{3.3}$$

and every $\varphi \in \mathcal{R}_N$ with $\|\varphi\| = 1$ has the representation

$$\varphi = \sum_{n=0}^{2^R-p-1} \alpha_k \phi_{R,n} + \sum_{n=2^R-p}^{2^R-1} \beta_k \phi_{R,n}^\# \text{ with } \sum_{n=0}^{2^R-p-1} |\alpha_n|^2 + \sum_{n=2^R-p}^{2^R-1} |\beta_n|^2 = 1. \tag{3.4}$$

3.1.2. *Boundary wavelets in higher dimensions.* In this paper, we also consider the d-dimensional case. For the reconstruction in d-dimensions we focus on separable boundary wavelets. Therefore, the d-dimensional wavelets can be derived from the one dimensional case by tensoring the scaling space and then studying the according wavelet space.

From (3.1) we have the one dimensional boundary scaling function. With the tensor product we get the d dimensional one, i.e. $\phi_{j,n}^d = \phi_{j,n_1}^b \otimes \dots \otimes \phi_{j,n_d}^b$ for $n = \{n_1, \dots, n_d\} \in \mathbb{N}^d$, $j \geq J_0$. To make this easier to read we set $\phi_{j,n} := \phi_{j,n}^d$, as the dimension is defined by the context. Then the d-dimensional scaling space is given by

$$V_J^{b,d} := V_J^b \otimes \dots \otimes V_J^b \ (d \text{ times}).$$

For the purpose of constructing higher dimensional boundary wavelets we exploit the MRA structure. We have that

$$V_j^b = V_{j-1}^b \oplus W_{j-1}^b.$$

Therefore, we can divide in higher dimensions the scaling space at one level in the scaling space and the wavelet space in the lower level

$$V_j^{b,d} = V_j^b \otimes \ldots \otimes V_j^b = (V_{j-1}^b \oplus W_{j-1}^b) \otimes \ldots \otimes (V_{j-1}^b \oplus W_{j-1}^b) = V_{j-1}^{b,d} \oplus W_{j-1}^{b,d}.$$

This way we have defined the d dimensional boundary-corrected wavelet space $W_{j-1}^{b,d}$ by

$$W_{j-1}^{b,d} := (V_{j-1}^b \oplus W_{j-1}^b) \otimes \ldots \otimes (V_{j-1}^b \oplus W_{j-1}^b) \ominus V_{j-1}^{b,d}.$$

Due to (3.2) we only have to focus on the scaling space, as the sum over the wavelet spaces can be represented by the scaling space at highest scale. Therefore, we do not explain details about the wavelets here. We have with (3.3)

$$V_j^{b,d} = V_j^b \otimes \ldots \otimes V_j^b = \operatorname{span}\left\{\phi_{j,n}^d := \phi_{j,n_1}^b \otimes \ldots \otimes \phi_{j,n_d}^b : n = \{n_1, \ldots, n_d\}, n_i = 0, \ldots 2^j - 1\right\}.$$

According to the size of the one dimensional scaling space, we know that the d-dimensional scaling space has size 2^{dj}. The reconstruction space for $N = 2^{dR}$ is then

$$\mathcal{R}_N = V_R^{b,d}.$$

In order to get the scaling space for the boundary wavelets in one dimension we had to reflect the scaling function for translates $k = 2^j - p, \ldots, 2^j - 1$. This means that $V_j^{b,d}$ is spanned by the translates of 2^d functions, which are constructed by tensoring of the original scaling function $\phi^0 := \phi$ and the translated version $\phi^1 := \phi^{\#}$. Define $K_0 = \left\{0, \ldots, 2^j - p - 1\right\}$ and $K_1 = \left\{2^j - p, \ldots, 2^j - 1\right\}$, then the mapping $m : \left\{0, \ldots, 2^j - 1\right\} \mapsto \{0, 1\}$ is given by

$$m(n) = \begin{cases} 0 & n \in K_0 \\ 1 & n \in K_1. \end{cases}$$

This allows us to represent $\varphi \in V_j^{b,d}$ with $||\varphi|| = 1$ by

$$\varphi = \sum_{n=0}^{2^j-1} \alpha_n \phi_{j,n_1}^{m(n_1)} \otimes \ldots \otimes \phi_{j,n_d}^{m(n_d)} = \sum_{s \in \{0,1\}^d} \sum_{n \in K_s} \alpha_n \bigotimes \phi_{j,n}^s \text{ with } \sum_{s \in \{0,1\}^d} \sum_{n \in K_s} |\alpha_n|^2 = 1, \tag{3.5}$$

where

$$\bigotimes \phi_{j,n}^s = \phi_{j,n_1}^{s_1} \otimes \ldots \otimes \phi_{j,n_d}^{s_d}.$$

So at this point we have the wavelets and the scaling functions which span the space $L^2([0,1]^d)$ and therefore the reconstruction from Walsh functions in the wavelet space is guaranteed.

4. The main theorems and its proof

Taken together, we can now prove the main result.

4.1. Useful lemmas about Walsh functions.

For the proof of theorem 1.1 we have to combine the properties of Walsh functions in the dyadic analysis with the properties of the wavelets in the decimal analysis. Therefore, we consider under which conditions the decimal and dyadic additions are equal. This is important to combine the multiplicative identity of the Walsh functions with the translates of the wavelets.

Lemma 4.1. *Let $x \in [0,1)$ and $m \in \mathbb{N}$ then $x + m = x \oplus m$.*

Proof. The dyadic representation of x is $\{\ldots, 0, x_{-1}, x_{-2}, \ldots\}$ and the dyadic representation of

$$m \text{ is } \{\ldots, m_2, m_1, m_0, 0, 0, \ldots\}.$$

Because the representations do not have non-zero elements at the same position, one achieves the following

$$x \oplus m = \sum_{i=-\infty}^{\infty} (x_i \oplus_2 m_i) 2^i = \sum_{i=-\infty}^{\infty} (x_i + m_i) 2^i = x + m.$$

□

Next, we look at the inverse element for the dyadic addition. This is also discussed in [28] and will be used in Corollary 4.3.

Lemma 4.2. *The dyadic sum of two numbers $x, y \in \mathbb{R}_+$ is 0 if and only if $x = y$.*

Proof. Let $x, y \in \mathbb{R}_+$ with the dyadic representation $\{x_i\}_{i \in \mathbb{Z}}$ and $\{y_i\}_{i \in \mathbb{Z}}$. Then

$$x \oplus y = \sum_{i=-\infty}^{\infty} (x_i \oplus_2 y_i) 2^i = 0$$

if and only if $x_i \oplus_2 y_i = 0$ for all $i \in \mathbb{Z}$. This is the case if and only if $x_i = y_i$ for all $i \in \mathbb{Z}$, i.e. $x = y$. □

With this the relation between the decimal addition and the multiplicative identity of the Walsh functions can be found.

Corollary 4.3. *Let $t \in \mathbb{N}$ and $x \in [0, 1)$, then the following holds:*

$$\mathcal{W}\{f(x+t)\}(s) = \mathcal{W}\{f(x)\}(s) \operatorname{Wal}(t, s).$$

Proof. With Lemma 4.1 we have that $x \oplus t = x + t$. This allows

$$\mathcal{W}\{f(x+t)\}(s) = \mathcal{W}\{f(x \oplus t)\}(s) = \mathcal{W}\{f(x)\}(s) \operatorname{Wal}(t, s).$$

□

Next, we analyse the sum of Walsh functions with equally distributed inputs. This will be used in Lemma 4.5.

Lemma 4.4. *Let $N = 2^n$, $n \in \mathbb{N}$, then for all $s \in \mathbb{N}_0$ the following property holds:*

$$\sum_{i=0}^{N-1} \operatorname{wal}(s, \frac{i}{N}) = \begin{cases} N & \text{if } s = 0 \\ 0 & \text{else.} \end{cases} \tag{4.1}$$

Proof. The first case for $s = 0$ follows directly by the definition of the Walsh function as $\operatorname{wal}(0; x) \equiv 1$ for all $x \in [0, 1)$. For the second part we use the equal distribution of Walsh functions in intervals where it takes the values -1 and 1, i.e. for $s \leq 2^m$, $m \in \mathbb{N}_0$ the Walsh function $\operatorname{wal}(s; x)$ takes the value 1 on 2^{m-1} intervals of length $\frac{1}{2^m}$ and -1 on the same number of intervals of that length [27]. As the sequence $\{i/N\}_{i=0,\ldots N-1}$ is equally distributed on this interval, the sum equals 0. □

With this information in hand we can now prove the following lemma, which shows a relation between the values of the discrete Walsh transform and the signal itself. This will then be used in Lemma 4.6.

Lemma 4.5. *Let $N = \{N_i\}_{i=1,\ldots,d}$, where $N_i = 2^{n_i}, n_i \in \mathbb{N}$ and $i = 1, \ldots, d$. Let $x \in \mathbb{R}^{N_1 \times \ldots \times N_d}$, where $x = \{x_k\}$ and $x_{k_i} \in \mathbb{R}$, $k = \{k_i\}_{i=1,\ldots,d}, k_i = 0, \ldots, N_i - 1$ be a discrete d-dimensional signal. Given the discrete d-dimensional Walsh transformed by $X = \{X_j\} \in \mathbb{R}^{N_1 \times \ldots \times N_d}$, where $X_{j_i} \in \mathbb{R}$, $j = \{j_i\}_{i=1,\ldots,d}, j_i = 0, \ldots, N_i - 1$, with*

$$X_j = \frac{1}{\prod_{i=1}^{d} N_i} \sum_{k=0}^{N-1} x_k \operatorname{Wal}(j, \frac{k}{N}),$$

it follows that

$$\sum_{j=0}^{N-1} |X_j|^2 = \frac{1}{\prod_{i=1}^d N_i} \sum_{k=0}^{N-1} |x_k|^2.$$

Proof. First, one observes that by definition and the fact that $N_i = 2^{n_i}, i = 1, \dots, d$ the following holds:

$$\begin{aligned} |X_j|^2 &= \frac{1}{\prod_{i=1}^d N_i^2} \left(\sum_{k=0}^{N-1} x_k \operatorname{Wal}(j, \frac{k}{N}) \right) \left(\sum_{k=0}^{N-1} x_k \operatorname{Wal}(j, \frac{k}{N}) \right) \\ &= \frac{1}{\prod_{i=1}^d N_i^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} x_k x_l \operatorname{Wal}(j_1, \frac{k_1 \oplus l_1}{N_1}) \cdot \ldots \cdot \operatorname{Wal}(j_d, \frac{k_d \oplus l_d}{N_d}). \end{aligned}$$

Next, recalling (4.1) and Lemma 4.2 we directly get the desired property by

$$\begin{aligned} \sum_{j=0}^{N-1} |X_j|^2 &= \frac{1}{\prod_{i=1}^d N_i^2} \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} x_k x_l \operatorname{Wal}(j_1, \frac{k_1 \oplus l_1}{N_1}) \cdot \ldots \cdot \operatorname{Wal}(j_d, \frac{k_d \oplus l_d}{N_d}) \\ &= \frac{1}{\prod_{i=1}^d N_i^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} x_k x_l \sum_{j_1=0}^{N_1-1} \operatorname{Wal}(j_1, \frac{k_1 \oplus l_1}{N_1}) \cdot \ldots \cdot \sum_{j_d=0}^{N_d-1} \operatorname{Wal}(j_d, \frac{k_d \oplus l_d}{N_d}) \\ &= \frac{1}{\prod_{i=1}^d N_i} \sum_{k=0}^{N-1} |x_k|^2. \end{aligned}$$

□

With this we can achieve a useful relation between the Walsh polynomial, i.e. $\Phi(z) = \sum_{j=A}^B \alpha_j \operatorname{Wal}(j, z)$ with $A, B \in \mathbb{Z}^d$ and $\alpha_{j_i} \in \mathbb{R}$ for all $j_i = A_i, \dots, B_i$, $i = 1, \dots, d$, and its coefficients α similar to the trigonometric polynomial in [3].

Lemma 4.6. *Let $A, B \in \mathbb{Z}^d$ such that $A_i \le B_i, i = 1, \dots, d$ and consider the Walsh polynomial $\Phi(z) = \sum_{j=A}^B \alpha_j \operatorname{Wal}(j, z)$ for $z \in \mathbb{R}_+^d$. If $L = \{L_1, \dots, L_d\}$ with $L_i = 2^{n_i}$, $n_i \in \mathbb{N}, i = 1, \dots, d$ such that $2L_i \ge B_i - A_i + 1$, then*

$$\sum_{j=0}^{2L-1} \frac{1}{\prod_{i=1}^d 2L_i} \left| \Phi(\frac{j}{2L}) \right|^2 = \sum_{j=A}^B |\alpha_j|^2.$$

Proof. For the proof Lemma 4.5 is used. Therefore, let $x = \{x_k\}$, where $x_{k_i} \in \mathbb{R}$, $k = \{k_i\}_{i=1,\dots,d}, k_i = 0, \dots, 2L_i - 1$ and $x \in \mathbb{R}^{2L_1 \times \ldots \times 2L_d}$ with the discrete Walsh transformed $X = \{X_j\} \in \mathbb{R}^{2L_1 \times \ldots \times 2L_d}$, where $X_{j_i} \in \mathbb{R}$, $j = \{j_i\}_{i=1,\dots,d}, j_i = 0, \dots, 2L_i - 1$ and $X \in \mathbb{R}^{2L_1 \times \ldots \times 2L_d}$. Consider the sums $k_i + A_i$, $i = 1, \dots, d$, there exists a number $\tilde{A}_i(k)$, such that $k_i + A_i = k_i \oplus \tilde{A}_i(k_i)$ for all $i = 1, \dots, d$. As before, we denote by $\tilde{A}(k)$ the multi-index containing all $\tilde{A}_i(k_i)$. Define the coefficients

$$\tilde{\alpha}_{k_i} = \frac{\alpha_{k_i}}{\operatorname{Wal}(\frac{j_i}{2L}, \tilde{A}_i(k_i))}$$

and the sequence x as follows:

$$x_{k+L} = \begin{cases} \tilde{\alpha}_{k+A+L} & -L_i \le k_i \le -L_i + B_i - A_i \\ 0 & \text{otherwise.} \end{cases}$$

Then one gets with the scaling property (2.1) and the multiplicative identity (2.2)

$$\begin{aligned} X_j &= \frac{1}{\prod_{i=1}^d 2L_i} \sum_{k=0}^{2L-1} x_k \operatorname{Wal}(j, \frac{k}{2L}) = \frac{1}{\prod_{i=1}^d 2L_i} \sum_{k=-L}^{L-1} x_{k+L} \operatorname{Wal}(j, \frac{k+L}{2L}) \\ &= \frac{1}{\prod_{i=1}^d 2L_i} \sum_{k=A}^B \tilde{\alpha}_k \operatorname{Wal}(\frac{j}{2L}, k + A) = \frac{1}{\prod_{i=1}^d 2L_i} \sum_{k=A}^B \tilde{\alpha}_k \operatorname{Wal}(\frac{j}{2L}, k \oplus \tilde{A}(k)) \\ &= \frac{1}{\prod_{i=1}^d 2L_i} \sum_{k=A}^B \tilde{\alpha}_k \operatorname{Wal}(\frac{j}{2L}, \tilde{A}(k)) \operatorname{Wal}(\frac{j}{2L}, k) = \frac{1}{\prod_{i=1}^d 2L_i} \sum_{k=A}^B \alpha_k \operatorname{Wal}(k, \frac{j}{2L}) = \frac{1}{\prod_{i=1}^d 2L_i} \Phi(\frac{j}{2L}). \end{aligned}$$

With that one can conclude

$$\sum_{j=0}^{2L-1} \frac{1}{\prod_{i=1}^{d} 2L_i} \left|\Phi(\frac{j}{2L})\right|^2 = \sum_{j=0}^{2L-1} \prod_{i=1}^{d} 2L_i |X_j|^2 = \sum_{j=0}^{2L-1} |x_j|^2 = \sum_{j=A}^{B} |\tilde{\alpha}_j|^2 = \sum_{j=A}^{B} |\alpha_j|^2.$$

□

4.2. **Changes of Wavelets.** As mentioned in chapter 2 the decimal and dyadic addition do not correspond directly to each other. In particular, the representation of the decimal addition with a number h to a number x depends on both parts of the sum. However, in Corollary 4.3 we have seen that for $x \in [0,1)$ and $n \in \mathbb{N}$ the dyadic and decimal addition coincide. In the proof of the main theorem we want to transfer the time shifts of the wavelet to the Walsh function, i.e.

$$\begin{aligned}\int_{2^{-R}(n-p+1)}^{2^{-R}(n+p)} 2^{R/2}\phi(2^R x - n)\,\mathrm{Wal}(k,x)dx &= 2^{-R/2}\int_{-p+1}^{p} \phi(x)\,\mathrm{Wal}(k, 2^{-R}(x+n))dx \\ &\neq 2^{-R/2}\int_{-p+1}^{p} \phi(x)\,\mathrm{Wal}(k, 2^{-R}(x \oplus n))dx.\end{aligned}$$

Therefore, to enable us to use Corollary 4.3 and make the last equation an equality, the domain of the wavelets needs to be restricted to $[0,1]^d$. This is not a contradiction to the construction of the previous chapter, because the functions $\phi_{R,n}$ are indeed supported in $[0,1]^d$. However, the scaling function at level 0 is not only supported there and that is the function that we are dealing with after the change of variables in the integral. To solve this problem we represent the scaling function as a sum of functions that are supported in $[0,1]$, i.e.

$$\phi(x) = \sum_{i=-p+2}^{p} \phi_i(x-i+1) \text{ with } \phi_i(x) = \phi(x+i-1)\mathcal{X}_{[0,1]}(x) \tag{4.2}$$

and

$$\phi_{R,n} = 2^{R/2} \sum_{i=-p+2}^{p} \phi_i(2^R x - i + 1 - n).$$

This can also be done accordingly for the reflected function $\phi^{\#}$. In the higher dimensional case we have

$$\phi(x) = \phi_1(x_1) \otimes \ldots \otimes \phi_d(x_d) = \sum_{i=-p+2}^{p} \phi_{i_1}(x_1 - i_1 + 1) \cdot \ldots \cdot \phi_{i_d}(x_d - i_2 + 1)$$

and ϕ_{i_k} defined as above. This way the multiplicative identity holds also for the decimal time shift of the wavelets.

4.3. **Useful lemma about wavelets.** For the proof of the main theorem we have to bound the decay rate of the Wavelets under the Walsh transform. We analyse the Haar wavelets and the boundary-corrected Daubechies wavelet separately.

For the Haar wavelet we have that the scaling function $\phi = \mathcal{X}_{[0,1]}$. Hence,

$$\begin{aligned}\widehat{\phi}^W(\frac{j}{2^R} + m) &= \int_0^1 \mathrm{Wal}(\frac{j}{2^R} + m, x)dx = \int_0^1 \mathrm{Wal}(j + 2^R m, \frac{x}{2^R})dx \\ &= \int_0^{2^{-R}} \mathrm{Wal}(j + 2^R m, x)2^R dx = \begin{cases} 1 & m = 0, j < 2^R \\ 0 & \text{else.}\end{cases}\end{aligned}$$

Therefore we have

$$\widehat{\phi}^W(\frac{j}{2^R} + m) \leq \frac{1}{m} \quad \text{for } j = 0, \ldots, 2^R - 1, m \geq 1. \tag{4.3}$$

Next, we use the result from [9] regarding the decay rate of Hölder continuous functions. We have that the ϕ_i and $\phi_i^{\#}$ have the same smoothness properties as the original function ϕ. Hence, we have for the Daubechies

wavelet of order 2 that it is Hölder continuous with coefficient $\alpha = 0.55$. The higher order wavelets are all Hölder continuous with coefficient $\alpha = 1$. The results in [9] hold only for inputs in $\mathbb{N}$. We will use the same proof strategy and notation for inputs in the form of $\frac{k}{2^R}$ for some $k, R \in \mathbb{N}$.

Lemma 4.7. *Let ϕ be a Daubechies wavelet of order $p > 1$. And let $m \geq 1, R \in \mathbb{N}$, $L = 2^R$ and $j = 0, \ldots, L-1$. Then we have that*

$$|\widehat{\phi_i}^W(\frac{j}{L} + m)| \leq \frac{C}{m^\alpha}$$

and

$$|\widehat{\phi_i^\#}^W(\frac{j}{L} + m)| \leq \frac{C^\#}{m^\alpha}$$

for some constants C and $C^\#$.

Proof. First, observe that

$$\int_0^1 \phi_i(x)\,\mathrm{Wal}(\frac{j}{L} + m, x)dx = \int_0^1 \phi_i(x)\,\mathrm{Wal}(j + Lm, \frac{x}{L})dx.$$

We want to use the same technique as in [9] to divide the integral into parts where the Walsh function takes the values $+1$ and -1 on intervals of the same length. Let $q \in \mathbb{N}$ such that $2^q \leq m < 2^{q+1}$, then we have that $2^{q+R} \leq j + Lm < 2^{q+R+1}$. Now, define the intervals $\Delta_k^q = [2^{-q}k, 2^{-q}(k+1))$. It was shown that the function $\mathrm{Wal}(j + Lm, s)$ is constant on the interval Δ_{2l}^{q+R+1} and Δ_{2l+1}^{q+R+1} and takes the values $+1$ on one of them and -1 on the other. Hence, $\mathrm{Wal}(j + Lm, \frac{x}{L})$ takes these values on the intervals Δ_{2k}^{q+1} and Δ_{2k+1}^{q+1}. Due to the Hölder continuity we have that there exists a constant C such that $\phi_i(x) \leq \phi_i(s) + C|x - s|^\alpha$ for all $s \in [0, 1]$. With this one gets

$$\sup_{x \in \Delta_k^q} \phi_i(x) \leq \phi_i(2^{-q}k + 2^{-q-1}) + C2^{-(q+1)\alpha}$$
$$\sup_{x \in \Delta_k^q} -\phi_i(x) \leq -\phi_i(2^{-q}k + 2^{-q-1}) + C2^{-(q+1)\alpha}.$$

Hence,

$$\begin{aligned}
&|\int_{\Delta_k^q} \phi_i(x)\,\mathrm{Wal}(j + Lm, \frac{x}{L})dx| \\
&\qquad \leq 2^{-q}|(\phi_i(2^{-q}k + 2^{-p-1}) + C2^{-(p+1)\alpha}) \ + (-\phi_i(2^{-q}k + 2^{-p-1}) + C2^{-(p+1)\alpha})| \\
&\qquad \leq 2^{-q}C2^{-q\alpha}.
\end{aligned}$$

Thus, we get for the complete integral on $[0, 1]$

$$\begin{aligned}
\int_0^1 \phi_i(x)\,\mathrm{Wal}(\frac{j}{L} + m, x)dx &= \sum_{k=0}^{2^q+1} \int_{\Delta_k^q} \phi(x)\,\mathrm{Wal}(j + Lm, \frac{x}{L})dx \\
&\leq \sum_{k=0}^{2^q+1} 2^{-q}|2^{-q}|^\alpha \leq C2^{-q\alpha} \leq \frac{2C}{m^\alpha}.
\end{aligned}$$

The proof holds analogously for $\phi_i^\#$. □

4.4. **Proof of the main theorem.** With the tools established above, we can now prove the main result. To make the exposition easier to read we first prove the theorem in one dimension and then make the generalisation to several dimensions in a separate proof. Given the setup with the multi-indices framework, this can be done reasonably smoothly.

Proof of Theorem 1.1 in one dimension. The aim of this proof is to find for every $\theta \in (1, \infty)$ an integer $S_\theta \in \mathbb{N}$, such that for all $M \geq S_\theta N$ the subspace angle is bounded, i.e. $\mu(\mathcal{R}_N, \mathcal{S}_M) \leq \theta$. Let $R \in \mathbb{N}$ be the number of reconstructed levels, i.e. $N = 2^R$. We start with a suitable representation of $\cos(\omega(\mathcal{R}_N, \mathcal{S}_M))$.

There exists $\varphi \in \mathcal{R}_N$ with $\|\varphi\| = 1$ such that $\inf_{f\in\mathcal{R}_N,\|f\|=1} \|P_{\mathcal{S}_M} f\| = \|P_{\mathcal{S}_M}\varphi\|$, because the closed unit ball in $\mathcal{R}_N$ is compact and $P_{\mathcal{S}_M}$ is continuous. By (3.4) we can represent φ as

$$\varphi = \sum_{l=0}^{2^R-p-1} \alpha_l \phi_{R,n} + \sum_{l=2^R-p}^{2^R-1} \beta_l \phi^{\#}_{R,n} \text{ with } \sum_{l=0}^{2^R-p-1} |\alpha_n|^2 + \sum_{l=2^R-p}^{2^R-1} |\beta_n|^2 = 1 \tag{4.4}$$

and

$$\begin{aligned} \cos(\omega(\mathcal{R}_N, \mathcal{S}_M)) &= \inf_{f\in\mathcal{R}_N,\|f\|=1} \|P_{\mathcal{S}_M} f\| = \|P_{\mathcal{S}_M}\varphi\| \\ &= \|\varphi - P^{\perp}_{\mathcal{S}_M}\varphi\| \geq \|\varphi\| - \|P^{\perp}_{\mathcal{S}_M}\varphi\| = 1 - \|P^{\perp}_{\mathcal{S}_M}\varphi\|. \end{aligned} \tag{4.5}$$

The first equation (4.4) allows us to deal only with the scaling function instead of both the wavelets at different scales and the scaling function. The second one (4.5) enables us to bound $P^{\perp}_{\mathcal{S}_M}\varphi$ from above in lieu of $P_{\mathcal{S}_M}\varphi$ from below.

Instead of dealing with all different shifts of the scaling function, we aim for a closed form that only depends on the functions ϕ and $\phi^{\#}$. An essential part in the construction of this is the use of the scaling property in Corollary 4.3. Therefore, it is necessary, that $m \in \mathbb{N}$ and $x \in [0,1)$. For this sake, the functions ϕ_i were defined in (4.2) and we define

$$p_R : \mathbb{Z} \to \mathbb{N} \tag{4.6}$$

with $z \mapsto p_R(z)$ and $p_R(z)$ being the smallest integer such that $p_R(z)2^R + z > 0$. This yields

$$\begin{aligned} \langle \phi_{R,n}, \mathrm{Wal}(k,\cdot)\rangle &= \sum_{i=-p+2}^{p} \langle \phi_{i,R,n}, \mathrm{Wal}(k,\cdot)\rangle \\ &= 2^{R/2} \sum_{i=-p+2}^{p} \int_{2^{-R}(n+i-1)}^{2^{-R}(n+i)} \phi_i(2^R x - n - i + 1)\,\mathrm{Wal}(k,x)dx \\ &= 2^{-R/2} \sum_{i=-p+2}^{p} \int_0^1 \phi_i(x)\,\mathrm{Wal}(k, 2^{-R}(x+n+i-1))dx \\ &= 2^{-R/2} \sum_{i=-p+2}^{p} \int_0^1 \phi_i(x)\,\mathrm{Wal}(k, 2^{-R}(x+n+i-1+2^R p_R(i-1)))dx, \end{aligned} \tag{4.7}$$

where we used in the last line the fact that the Walsh functions are 1-periodic, if the other input data is an integer. Then, we have that $x \in [0,1]$ and $n+i-1+2^R p_R(n+i-1) \in \mathbb{N}$. Hence, Corollary 4.3 can be used in the third line to get

$$\begin{aligned} &\langle \phi_{R,n}, \mathrm{Wal}(k,\cdot)\rangle \\ &= 2^{-R/2} \sum_{i=-p+2}^{p} \int_0^1 \phi_i(x)\,\mathrm{Wal}(k, 2^{-R}(x+(n+i-1+2^R p_R(i-1))))dx \\ &= 2^{-R/2} \sum_{i=-p+2}^{p} \mathrm{Wal}(k, 2^{-R}(n+i-1+2^R p_R(i-1))) \int_0^1 \phi_i(x)\,\mathrm{Wal}(k, 2^{-R}x)dx \\ &= 2^{-R/2} \sum_{i=-p+2}^{p} \mathrm{Wal}(n+i-1+2^R p_R(i-1), \frac{k}{2^R})\widehat{\phi_i}^{w}(\frac{k}{2^R}). \end{aligned} \tag{4.8}$$

With

$$\Phi_i(z) = \sum_{n=0}^{2^R-p-1} \alpha_n \,\mathrm{Wal}(n+i-1+2^R p_R(i-1), z) \tag{4.9}$$

it results in

$$\sum_{n=0}^{2^R-p-1} \alpha_n \langle \phi_{i,R,n}, \mathrm{Wal}(k,\cdot)\rangle = 2^{-R/2} \sum_{n=0}^{2^R-p-1} \alpha_n \,\mathrm{Wal}(n+i-1+2^R p_R(i-1), \frac{k}{2^R}) \widehat{\phi_i}^W(\frac{k}{2^R}) \tag{4.10}$$
$$= 2^{-R/2} \widehat{\phi_i}^W(\frac{k}{2^R}) \Phi_i(\frac{k}{2^R}).$$

Analogously this can be done for the reflected function $\phi^{\#}$. Thus, by using

$$\Phi_i^{\#}(z) = \sum_{n=2^R-p}^{2^R-1} \beta_n \,\mathrm{Wal}(n+i-1+2^R p_R(i-1), z)$$

similar to the above, we get

$$\sum_{n=2^R-p}^{2^R-1} \beta_n \langle \phi_{i,R,n}^{\#}, \mathrm{Wal}(k,x)\rangle = 2^{-R/2} \sum_{n=2^R-p}^{2^R-1} \beta_n \,\mathrm{Wal}(n+i-1+2^R p_R(i-1), \frac{k}{2^R}) \widehat{\phi_i^{\#}}^W(\frac{k}{2^R}) \tag{4.11}$$
$$= 2^{-R/2} \widehat{\phi_i^{\#}}^W(\frac{k}{2^R}) \Phi_i^{\#}(\frac{k}{2^R}).$$

This representation is very useful. Indeed, one only has to pay attention to the decay rate of the Walsh transform of the pieces of the mother scaling function, its reflection and the Walsh polynomial. Moreover, the pieces fulfil $f(t) = 0$ for $t < 0$, such that the Walsh transform with one kernel can be used for the analysis. Consider

$$\|P_{\mathcal{S}_M}^{\perp}\varphi\| = \|P_{\mathcal{S}_M}^{\perp}(\sum_{n=0}^{2^R-p-1} \alpha_n \phi_{R,n} + \sum_{n=2^R-p}^{2^R-1} \beta_n \phi_{R,n}^{\#})\| \tag{4.12}$$
$$= \|P_{\mathcal{S}_M}^{\perp}(\sum_{n=0}^{2^R-p-1} \alpha_n \sum_{i=-p+2}^{p} \phi_{i,R,n} + \sum_{n=2^R-p}^{2^R-1} \beta_n \sum_{i=-p+2}^{p} \phi_{i,R,n}^{\#})\|.$$

Next, one uses the linearity of the orthogonal projection to change the order of the summands, such that the sum over the scaling function pieces can be dealt with in the end. First, we take out the sum over the parts of the wavelet $\phi_{i,R,n}$ to handle every cut-out of the wavelet separately. In particular, by (4.12)

$$\|P_{\mathcal{S}_M}^{\perp}\varphi\| = \|\sum_{i=-p+2}^{p} P_{\mathcal{S}_M}^{\perp}(\sum_{n=0}^{2^R-p-1} \alpha_n \phi_{i,R,n} + \sum_{n=2^R-p}^{2^R-1} \beta_n \phi_{i,R,n}^{\#})\| \tag{4.13}$$
$$\leq \sum_{i=-p+2}^{p} \|P_{\mathcal{S}_M}^{\perp}(\sum_{n=0}^{2^R-p-1} \alpha_n \phi_{i,R,n} + \sum_{n=2^R-p}^{2^R-1} \beta_n \phi_{i,R,n}^{\#})\|$$
$$= \sum_{i=-p+2}^{p} \sqrt{\sum_{k>M} \left| \sum_{n=0}^{2^R-p-1} \alpha_n \langle \phi_{i,R,n}, \mathrm{Wal}(k,\cdot)\rangle + \sum_{n=2^R-p}^{2^R-1} \beta_n \langle \phi_{i,R,n}^{\#}, \mathrm{Wal}(k,\cdot)\rangle \right|^2}.$$

Second, it follows, by using (4.10), (4.11) and the Cauchy-Schwarz inequality, that

$$\|P_{\mathcal{S}_M}^{\perp}\varphi\| \leq \sum_{i=-p+2}^{p} \sqrt{\sum_{k>M} 2^{-R} \left| \widehat{\phi_i}^W(\frac{k}{2^R})\Phi_i(\frac{k}{2^R}) + \widehat{\phi_i^{\#}}^W(\frac{k}{2^R})\Phi_i^{\#}(\frac{k}{2^R}) \right|^2} \tag{4.14}$$
$$\leq \sum_{i=-p+2}^{p} \left(\sum_{k\geq M} 2^{-R} \left| \widehat{\phi_i}^W(\frac{k}{2^R})\Phi_i(\frac{k}{2^R}) \right|^2 + \sum_{k\geq M} 2^{-R} \left| \widehat{\phi_i^{\#}}^W(\frac{k}{2^R})\Phi_i^{\#}(\frac{k}{2^R}) \right|^2 \right.$$
$$\left. +2 \left(\sum_{k\geq M} 2^{-R} \left| \widehat{\phi_i}^W(\frac{k}{2^R})\Phi_i(\frac{k}{2^R}) \right|^2 \right)^{1/2} \left(\sum_{k\geq M} 2^{-R} \left| \widehat{\phi_i^{\#}}^W(\frac{k}{2^R})\Phi_i^{\#}(\frac{k}{2^R}) \right|^2 \right)^{1/2} \right)^{1/2}.$$

We will only deal with the first summand and the other follow analogously. In the following step the 1-periodicity of the Walsh function is used. Let $S \in \mathbb{R}_+$ such that $M = S2^R$, then by replacing $k \geq M$ with

$k = mL + j$ with $L = 2^R$, $m \geq S$ and $j = 0, \ldots, L-1$ we have

$$\sum_{k\geq M} 2^{-R} \left|\widehat{\phi_i}^W(\frac{k}{2^R})\Phi_i(\frac{k}{2^R})\right|^2 \leq \sum_{j=0}^{L-1} \frac{1}{L}\left|\Phi_i(\frac{j}{L})\right|^2 \sum_{m\geq S}\left|\widehat{\phi_i}^W\left(\frac{j}{L}+m\right)\right|^2.$$

The last sum can be estimated via Lemma 4.7 and (4.3) by

$$\sum_{m\geq S}\left|\widehat{\phi_i}^W\left(\frac{j}{L}+m\right)\right|^2 \leq \sum_{m\geq S}\frac{A^2}{m^{2\alpha}} \leq \frac{A^2}{S^{2\alpha-1}} \tag{4.15}$$

with $\alpha = 1$ for the Haar wavelet and α being the Hölder coefficient for the other Daubechies wavelets of order $p > 1$. For the first sum $\sum_{j=0}^{L-1}\frac{1}{L}\left|\Phi_i(\frac{j}{L})\right|^2$ we have with $\alpha_n = 0$ for $n = -p+1, \ldots, -1$ that

$$\begin{aligned}\Phi_i(z) &= \sum_{n=0}^{2^R-p}\alpha_n \operatorname{Wal}(n+i-1+2^R p_R(i-1), z)\\ &= \sum_{n=-p+1}^{2^R-p}\alpha_l \operatorname{Wal}(n+i-1+2^R p_R(i-1), z)\\ &= \sum_{n=-p+1+i-1+2^R p_R(i-1)}^{2^R-p+i-1+2^R p_R(i-1)}\alpha_{n-i+1-2^R p_R(i-1)} \operatorname{Wal}(n, z)\end{aligned} \tag{4.16}$$

and $L = 2^R$, such that we can use Lemma 4.6 as $2^R - p + i - 1 + 2^R p_R(i-1) - (-p+1+i-1+2^R p_R(i-1)) + 1 = 2^R - p - (-p+1) + 1 = 2^R \geq L$ and obtain

$$\sum_{j=0}^{L-1}\frac{1}{L}\left|\Phi_i(\frac{j}{L})\right|^2 = \sum_{l=-p+1+i-1+2^R p_R(i-1)}^{2^R-p+i-1+2^R p_R(i-1)}|\alpha_{l-i+1-2^R p_R(i-1)}|^2 = \sum_{n=-p+1}^{2^R-p}|\alpha_n|^2 \leq 1. \tag{4.17}$$

Altogether this gives with Lemma 4.7

$$\sum_{k\geq M} 2^{-R}\left|\widehat{\phi_i}^W(\frac{k}{2^R})\Phi_i(\frac{k}{2^R})\right|^2 \leq \frac{A^2}{S^{2\alpha-1}}, \tag{4.18}$$

and similarly

$$\sum_{k\geq M} 2^{-R}\left|\widehat{\phi_i^\#}^W(\frac{k}{2^R})\Phi_i^\#(\frac{k}{2^R})\right|^2 \leq \frac{{A^\#}^2}{S^{2\alpha-1}}. \tag{4.19}$$

Using (4.17), (4.18) and (4.19) yields the following estimation

$$\begin{aligned}\|P^\perp_{\mathcal{S}_M}\varphi\| &\leq \sum_{i=-p+2}^{p}\left(\frac{A^2}{S^{2\alpha-1}} + \frac{{A^\#}^2}{S^{2\alpha-1}} + 2\frac{AA^\#}{S^{2\alpha-1}}\right)^{1/2}\\ &\leq (2p-2)\left(\frac{4\max\left\{A^2, {A^\#}^2\right\}}{S^{2\alpha-1}}\right)^{1/2} = (2p-2)\left(\frac{C^2}{S^{2\alpha-1}}\right)^{1/2}.\end{aligned}$$

Thus, $\|P^\perp_{\mathcal{S}_M}\varphi\| \leq \gamma$ whenever

$$S \geq \left(\frac{C(2p-2)}{\gamma}\right)^{2/2\alpha-1}, \tag{4.20}$$

where $C = 4\max\left\{A^2, {A^\#}^2\right\}$. It follows from (4.5) that $\cos(\omega(\mathcal{R}_N, \mathcal{S}_M)) \geq 1-\gamma \geq \frac{1}{\theta}$,i.e. $\mu(\mathcal{R}_N, \mathcal{S}_M) \leq \theta$, whenever the constant S, which is dependent on θ and therefore denoted by S_θ, fulfils (4.20) with $\gamma = 1 - \frac{1}{\theta}$, i.e

$$S_\theta \geq \left(\frac{C(2p-2)\theta}{\theta-1}\right)^{2/2\alpha-1}$$

and $M = S_\theta L = S_\theta N$. □

Proof of Theorem 1.1 in d-dimensions. In higher dimensions we represent $\varphi \in \mathcal{R}_N$ in terms of the sum over 2^d different tensor products. Then, we need to investigate the inner products of these summands with the Walsh functions as in the one dimensional case. At this point the results from the one dimensional case come into play. Next, one investigates the parts of the set $I_M^\perp$, where $I_M = \{l = \{l_1, \dots, l_d\}, l_k = 0, \dots, M_k - 1\}$ and $M = \{M_1, \dots, M_d\} \in \mathbb{N}^d$ and $I_M^\perp = \mathbb{N}^d \setminus I_M$, which correspond to the largest estimates of the inner products of the summands of the wavelet and the Walsh function. Finally, these can be bounded with estimates from the one dimensional case and additional care for the finite sums.

Now, we present the described steps in more detail. Let $\varphi \in \mathcal{R}_N$ with $\|\varphi\| = 1$. Then we can represent φ as in (3.5) in the following sum

$$\varphi = \sum_{s \in \{0,1\}^d} \sum_{n \in K_s} \alpha_n \bigotimes \phi_{R,n}^s.$$

In the one dimensional case we derived the representation of $\sum_{n=0}^{2^R-p-1} \alpha_n \langle \phi_{i,R,n}, \mathrm{Wal}(k,\cdot)\rangle$ in terms of the Walsh transform of the wavelet and the Walsh polynomial, i.e. $2^{-R/2}\widehat{\phi_i}^W(\frac{k}{2^R})\Phi_i(\frac{k}{2^R})$. This equality from (4.10) is used to represent the inner product in higher dimensions. For this we need to define p_R from (4.6) for higher dimensions. In particular, let $p_R : \mathbb{Z}^d \to \mathbb{N}^d$ with $\{z_i\}_{i=1,\dots,d} = z \mapsto p_R(z) = \{p_R(z)_i\}_{i=1,\dots,d}$ and $p_R(z)_i$ being the smallest integer such that $p_R(z)_i 2^R - z_i > 0$ for all $i = 1, \dots, d$. Further, let $l = (l_1, \dots, l_d) \in \mathbb{Z}^d$. This yields

$$\langle \phi_{i,R,n}^s, \mathrm{Wal}(l,\cdot)\rangle = \prod_{k=1}^d \langle \phi_{i,R,n_k}^{s_k}, \mathrm{Wal}(l_k,\cdot)\rangle.$$

Here, the problem is reduced to the one dimensional case and we can apply (4.7) and (4.8) to get

$$\begin{aligned}\langle \phi_{i,R,n}^s, \mathrm{Wal}(l,\cdot)\rangle &= \prod_{k=1}^d 2^{-dR/2}\,\mathrm{Wal}(n_k + i_k - 1 + 2^R p_R(i_k - 1), \frac{l_k}{2^R})\widehat{\phi_{i_k}^{s_k}}^W(\frac{l_k}{2^R}) \\ &= \mathrm{Wal}(n + i - 1 + 2^R p_R(i-1), \frac{l}{2^R})\widehat{\phi_i^s}^W(\frac{l}{2^R}).\end{aligned}$$

This is defined now as in (4.9)

$$\Phi_i^s(z) = \sum_{n \in K_s} \alpha_n \,\mathrm{Wal}(n + i - 1 + 2^R p_R(i-1), \frac{z}{2^R}).$$

Note that the different definitions from the one dimensional case for Φ and $\Phi^\#$ are combined in the notation with the K_s. We get with this the presentation of the inner products as desired:

$$\sum_{n \in K_s} \alpha_n \langle \phi_{i,R,n}^s, \mathrm{Wal}(l,\cdot)\rangle = 2^{-dR/2}\Phi_i^s(\frac{l}{2^R})\widehat{\phi_i^s}(\frac{l}{2^R}).$$

For the representation of indices that correspond to the sampling functions, let $I_M = \{l = \{l_1, \dots, l_d\}, l_k = 0, \dots, M_k - 1\}$, where $M = \{M_1, \dots, M_k\} \in \mathbb{N}^d$ is the number of samples. Then $l \notin I_M$ corresponds to $l > m$ in the one dimensional case. We now want to analyse the orthogonal projection on the orthogonal complement of the sampling space

$$\|P_{\mathcal{S}_M}^\perp \varphi\| = \|P_{\mathcal{S}_M}^\perp(\sum_{s \in \{0,1\}^d} \sum_{n \in K_s} \alpha_n \phi_{R,n}^s)\| = \|P_{\mathcal{S}_M}^\perp(\sum_{s \in \{0,1\}^d} \sum_{n \in K_s} \alpha_n \sum_{i=-p+2}^p \phi_{i,R,n}^s)\|.$$

This way we ensured the use of Corollary 4.3. Next we change the order again to deal with the different cut-out functions separately. This was seen already in (4.13). We get

$$\|P_{\mathcal{S}_M}^\perp \varphi\| = \|\sum_{i=-p+2}^p P_{\mathcal{S}_M}^\perp(\sum_{s \in \{0,1\}^d} \sum_{n \in K_s} \alpha_n \sum_{i=-p+2}^p \phi_{i,R,n}^s)\| \le \sum_{i=-p+2}^p \|P_{\mathcal{S}_M}^\perp(\sum_{s \in \{0,1\}^d} \sum_{n \in K_s} \alpha_n \sum_{i=-p+2}^p \phi_{i,R,n}^s)\|.$$

With the Cauchy Schwarz inequality and a careful reordering we get as in (4.14)

$$\|P^{\perp}_{\mathcal{S}_M}\varphi\| = \sum_{i=-p+2}^{p}\sqrt{\sum_{l\notin I_{\mathcal{M}}}\left|\sum_{s\in\{0,1\}^d}\sum_{n\in K_s}\alpha_n\langle\phi^s_{i,R,n},\mathrm{Wal}(l,\cdot)\rangle\right|^2} = \sum_{i=-p+2}^{p}\sqrt{\sum_{l\notin I_{\mathcal{M}}}2^{-dR}\left|\sum_{s\in\{0,1\}^d}\Phi^s_i(\frac{l}{2^R})\widehat{\phi^s_i}^{W}(\frac{l}{2^R})\right|^2}$$

$$\leq \sum_{i=-p+2}^{p}\sum_{s\in\{0,1\}^d}\left(\sum_{l\notin I_{\mathcal{M}}}2^{-dR}|\Phi^s_i(\frac{l}{2^R})\widehat{\phi^s_i}^{W}(\frac{l}{2^R})|^2\right)^{1/2}.$$

Now, let $S\in\mathbb{N}$ be given, such that the number of samples $M=\{M_1,\dots,M_k\}\in\mathbb{N}^d$ is $M_k=S2^R$. Then if $l=\{l_1,\dots,l_d\}\notin I_{\mathcal{M}}$ at least one $l_k>M_k$. The sum is the largest if only one l_k fulfils this estimate. Hence, without loss of generality let $l_1>M_1$ and $l_k\leq M_k$ for $k=2,\dots,d$. Now let $l_k=j_k+u_k2^R$ and $\alpha=0.55$ or $\alpha=1$ depending on the chosen wavelet. Then we get similar to (4.15):

$$\begin{aligned}\sum_{l_1>M_1}\sum_{l_2\leq M_2}\cdots\sum_{l_d\leq M_d}2^{-dJ}\left|\Phi^s_i(\frac{l}{2^R})\widehat{\phi^s_i}^{W}(\frac{l}{2^R})\right|^2 &= \sum_{j=0}^{2^R-1}\frac{1}{2^{dR}}\left|\Phi^s_i(\frac{j}{2^R})\right|^2\sum_{u_1>S}\sum_{u_2\leq S}\cdots\sum_{u_d\leq S}\left|\phi^s_i(\frac{j}{2^R}+u)\right|^2\\ &\leq \sum_{j=0}^{2^R-1}\frac{1}{2^{dR}}\left|\Phi^s_i(\frac{j}{2^R})\right|^2\sum_{u_1>S}\sum_{u_2\leq S}\cdots\sum_{u_d\leq S}\frac{A_1^2}{(1+u_1)^{2\alpha}}\cdots\frac{A_d^2}{(1+u_d)^{2\alpha}}\\ &\leq \frac{C^{d-1}}{S^{2\alpha-1}}\sum_{j=0}^{2^R-1}\frac{1}{2^{dR}}\left|\Phi^s_i(\frac{j}{2^R})\right|^2.\end{aligned}$$

The last sum can be estimated with the help of Lemma 4.6. In (4.16) and (4.17) this was derived in the one dimensional case that can be directly used here, such that

$$\sum_{j=0}^{2^R-1}\frac{1}{2^{dR}}\left|\Phi^s_i(\frac{j}{2^R})\right|^2=\sum_{n\in K_s}|\alpha_n|^2.$$

From the fact that the $\phi^s_{R,n}$ form an orthonormal bases and $\|\varphi\|=1$ we have

$$\sum_{n\in K_s}|\alpha_n|^2\leq 1.$$

This together with the fact that $|\{0,1\}^d|=2^d$ gives

$$\sum_{s\in\{0,1\}^d}\left(\sum_{l_1>M_1}\sum_{l_2\leq M_2}\cdots\sum_{l_d\leq M_d}2^{-dR}\left|\Phi_i(\frac{l}{2^R})\widehat{\phi_i}(\frac{l}{2^R})\right|^2\right)^{1/2}\leq 2^d\left(\frac{C^{d-1}}{S^{2\alpha-1}}\right)^{1/2}.$$

By replacing C^{d-1} with C we have

$$\|P^{\perp}_{\mathcal{S}_M}\varphi\|\leq\sum_{i=-p+2}^{p}2^d\left(\frac{C}{S}\right)^{1/2}=(2p-2)2^d\left(\frac{C}{S^{2\alpha-1}}\right)^{1/2}.$$

Thus, $\|P^{\perp}_{\mathcal{S}_M}\varphi\|\leq\gamma$ whenever

$$S\geq C\left(\frac{(2p-2)2^d}{\gamma}\right)^{2/2\alpha-1}. \tag{4.21}$$

It follows from (4.5) that

$$\cos(\omega(\mathcal{R}_N,\mathcal{S}_M))\geq 1-\gamma\geq\frac{1}{\theta},$$

i.e. $\mu(\mathcal{R}_N,\mathcal{S}_M)\leq\theta$, whenever S_θ fulfils (4.21) with $\gamma=1-\frac{1}{\theta}$, i.e.

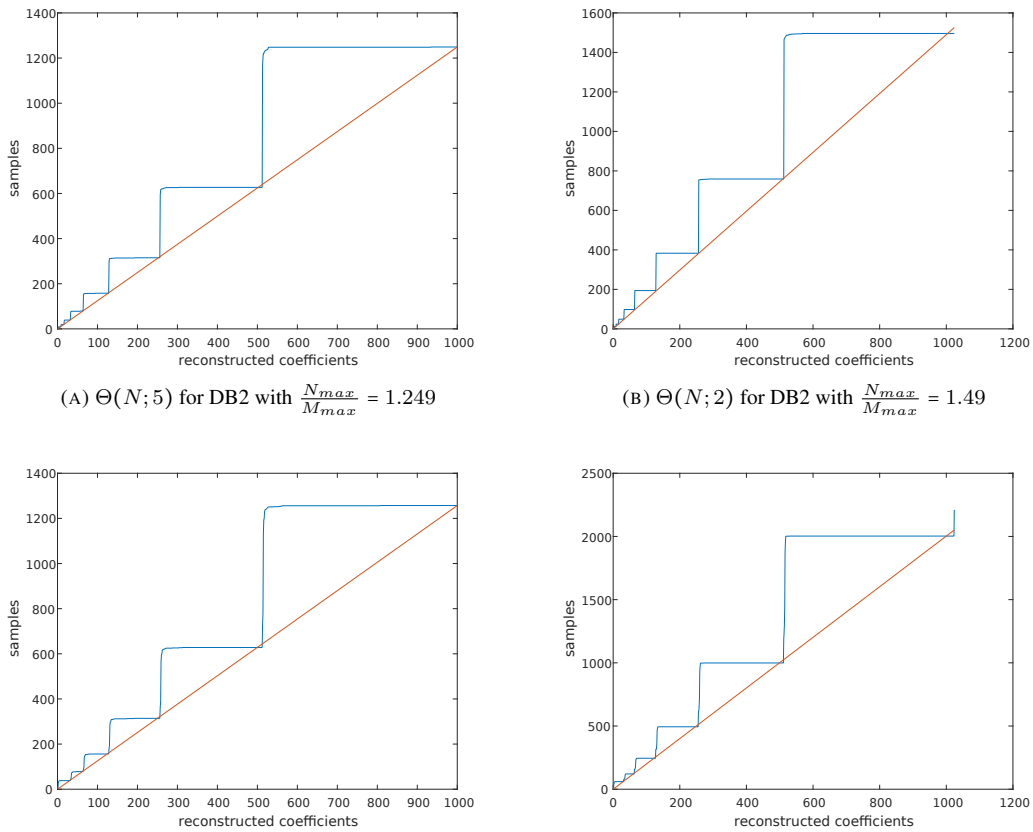
$$S_\theta\geq C\left(\frac{(2p-2)2^d\theta}{\theta-1}\right)^{2/2\alpha-1}$$

and $|M| = M_1 \cdot \ldots \cdot M_d = S_\theta^d 2^{dR} = S_\theta^d N$. □

5. Numerical Experiments

In this chapter we underline the theoretical results with numerical experiments. We first calculate the stable sampling rate for different stabilities θ and Daubechies wavelets. Then, we see that the reconstruction with generalized sampling leads to much better results then the direct inversion with the Walsh transform. Moreover, we point out that it is important to consider the stable sampling rate, as otherwise the reconstruction gets very unstable with meaningless results.

First, we see in Figure 1 the stable sampling rate for the different wavelets and stabilities in the one-dimensional case. One can see that it is indeed linear with jumps according to the levels of the wavelets. Moreover, it is easy to detect that the constant S_θ is considerably low such that the number of samples needed is only marginally larger than the number of coefficients that we reconstruct. It is not surprising that the stable sampling rate gets larger for smaller θ. In the theory of the reconstruction from Fourier measurements we have a direct relation between the smoothness of the wavelets and the size of the stable sampling rate. Similar relations are not known for the Walsh wavelet case.



(A) $\Theta(N;5)$ for DB2 with $\frac{N_{max}}{M_{max}} = 1.249$

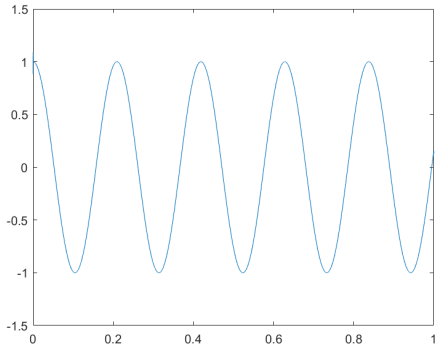
(B) $\Theta(N;2)$ for DB2 with $\frac{N_{max}}{M_{max}} = 1.49$

(C) $\Theta(N;5)$ for DB8 with $\frac{N_{max}}{M_{max}} = 1.257$

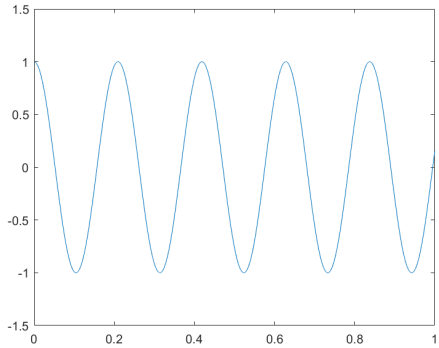
(D) $\Theta(N;2)$ for DB8 with $\frac{N_{max}}{M_{max}} = 2.003$

FIGURE 1. Plots of the stable sampling rate (blue) for Daubechies Wavelet of order 2 and 8 for a threshold $\theta = 2$ and 5 and the linear line with N_{max}/M_{max} (orange).

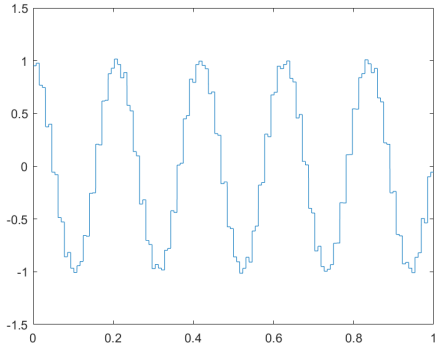
In Figure 2, we demonstrate the reconstruction with generalized sampling. For this sake we consider two different functions on $[0,1]$. First, we look at the cosine function in Figure 2a, taking 77 Walsh samples. In Figure 2c the direct inversion is shown. It is clear that the reconstruction has a lot of block artefacts, whereas the reconstruction with generalized sampling of 64 Daubechies 8 wavelets has nearly no visible artefacts.



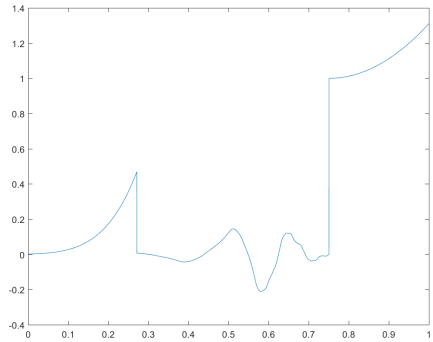
(A) Original function 1



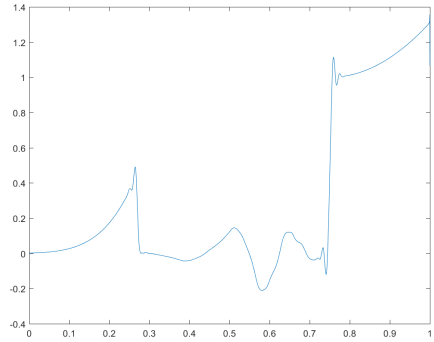
(B) Reconstruction with Generalized Sampling of function 1 with 64 Wavelet coefficients from 77 measurements



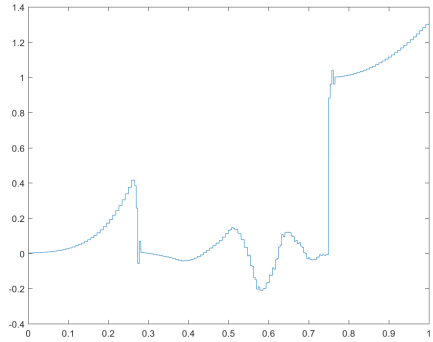
(C) Truncated Walsh series of function 1 from 77 measurements



(D) Original function 2



(E) Reconstruction with Generalized Sampling of function 2 with 128 Wavelet coefficients and 192 measurements



(F) Truncated Walsh series from 192 measurements

FIGURE 2. Reconstruction with Generalized Sampling and Daubechies 8 Wavelets and the inverse Walsh.

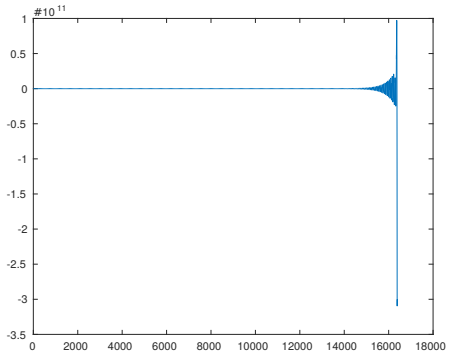


FIGURE 3. Reconstruction with Generalized Sampling below the Stable Sampling Rate with 512 Walsh samples and Daubechies 8 Wavelet coefficients

The same artefacts can be seen in 2f. In this case 192 Walsh samples were taken and 128 wavelet coefficients were reconstructed. The artefacts with the direct Walsh inverse are much stronger than the common Gibbs phenomena for the Fourier case. Because of this, reconstructions with Walsh functions are not feasible in practice. They are also the reason why one does not use Haar wavelets as they obey the same block artefacts. This underlines the need of a reconstruction technique that refers the data from the sampling space to a much more appropriate reconstruction space, where the data is represented sparsely. In this case we get fewer artefacts. For completely continuous functions as in 2b the reconstruction with generalized sampling has nearly no artefacts. In case of discontinuities as in the second function 2d one gets some artefacts at the discontinuities as can be seen in 2e. Even so, the overall reconstruction quality is much better and the reconstruction still obeys the regularity properties of each part of the function.

Nevertheless, it is important to take the stable sampling rate in mind. If one tries to reconstruct with fewer samples then needed, the reconstruction gets very unstable and one gets meaningless results. This can be seen in Figure 3, where a function on $[0, 1]$ is reconstructed from 512, which is much more than the 77 and 192 for the other functions.

6. CONCLUSION

We were able to investigate a very important part of the error estimate for different reconstruction methods. Moreover, we showed that binary measurements modelled by Walsh functions are well suited to reconstruct images with wavelets. This, together with the results in [1, 3], gives a broad knowledge about the accuracy and stability for two major applications of sampling theory, i.e. systems with Fourier samples and those with binary measurements.

ACKNOWLEDGEMENT

LT acknowledges support from the UK EPSRC grant and AH acknowledges support from a Royal Society University Research Fellowship.

REFERENCES

[1] B. Adcock, A. Hansen, G. Kutyniok, and J. Ma. Linear stable sampling rate: Optimality of 2d wavelet reconstructions from fourier measurements. *SIAM J. Math. Anal.*, 47(2):1196–1233, 2015.

[2] B. Adcock, A. Hansen, and C. Poon. Beyond consistent reconstructions: optimality and sharp bounds for generalized sampling, and application to the uniform resampling problem. *SIAM J. Math. Anal.*, 45(5):3132–3167, 2013.

[3] B. Adcock, A. Hansen, and C. Poon. On optimal wavelet reconstructions from fourier samples: linearity and universality. *Appl. Comput. Harmon. Anal.*, 36(3):387–415, 2014.

[4] B. Adcock and A. C. Hansen. A generalized sampling theorem for stable reconstructions in arbitrary bases. *J. Fourier Anal. Appl.*, 18(4):685–716, 2010.

[5] B. Adcock and A. C. Hansen. Generalized sampling and infinite-dimensional compressed sensing. *Foundations of Computational Mathematics*, 16(5):1263–1323, 2016.

[6] B. Adcock, A. C. Hansen, and C. Poon. On optimal wavelet reconstructions from Fourier samples: linearity and universality of the stable sampling rate. *Appl. Comput. Harmon. Anal.*, 36(3):387–415, 2014.

[7] B. Adcock, A. C. Hansen, C. Poon, and B. Roman. Breaking the coherence barrier: A new theory for compressed sensing. In *Forum of Mathematics, Sigma*, volume 5. Cambridge University Press, 2017.

[8] A. Aldroubi and M. Unser. A general sampling theory for nonideal acquisition devices. *IEEE Trans. Signal Process.*, 42(11):2915–2925, 1994.

[9] V. Antun. Coherence estimates between hadamard matrices and daubechies wavelets. Master's thesis, University of Oslo, 2016.

[10] M. Bachmayr, A. Cohen, R. DeVore, and G. Migliorati. Sparse polynomial approximation of parametric elliptic pdes. part ii: lognormal coefficients. *ESAIM: Mathematical Modelling and Numerical Analysis*, 51(1):341–363, 2017.

[11] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk. Data assimilation in reduced modeling. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):1–29, 2017.

[12] P. Binev, A. Cohen, O. Mula, and J. Nichols. Greedy algorithms for optimal measurements selection in state estimation using reduced models. preprint: hal-0163817, Nov. 2017.

[13] A. Böttcher. Infinite matrices and projection methods: in lectures on operator theory and its applications, fields inst. monogr. *Amer. Math. Soc.*, (3):1–72, 1996.

[14] E. J. Candès and L. Demanet. Curvelets and fourier integral operators. *C. R. Acad. Sci.*, 336(1):395–398, 2003.

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

[16] E. J. Candès and L. Donoho. Recovering edges in ill-posed inverse problems: optimality of curvelet frames. *Ann. Statist.*, 30(3):784–842, 2002.

[17] K. Choi, S. Boyd, J. Wang, L. Xing, L. Zhu, and T.-S. Suh. Compressed Sensing Based Cone-Beam Computed Tomography Reconstruction with a First-Order Method. *Medical Physics*, 37(9), 2010.

[18] A. Cohen, I. Daubechies, and P. Vial. Wavelets on the interval and fast wavelet transforms. *Comput. Harmon. Anal.*, 1(1):54–81, 1993.

[19] S. Dahlke, G. Kutyniok, P. Maass, C. Sagiv, H.-G. Stark, and G. Teschke. The uncertainty principle associated with the continuous shearlet transform. *Int. J. Wavelets Multiresolut. Inf. Process.*, 2(6):157–181, 2008.

[20] S. Dahlke, G. Kutyniok, G. Steidl, and G. Teschke. Shearlet coorbit spaces and associated banach frames. *Appl. Comput. Harmon. Anal.*, 2(27):195–214, 2009.

[21] R. DeVore, G. Petrova, and P. Wojtaszczyk. Data assimilation and sampling in banach spaces. *Calcolo*, 54(3):963–1007, 2017.

[22] M. N. Do and M. Vetterli. The contourlet transform: An efficient directional multiresolution image representation. *IEEE Transactions on Image Processing*, 12(14):2091–2106, 2005.

[23] T. Dvorkind and Y. C. Eldar. Robust and consistent sampling. *IEEE Signal Process. Letters*, 16(9):739–742, 2009.

[24] Y. C. Eldar. Sampling with arbitrary sampling and reconstruction spaces and oblique dual frame vectors. *J. Fourier Anal. Appl.*, 9(1):77–96, 2003.

[25] Y. C. Eldar. Sampling without input constraints: Consistent reconstruction in arbitrary spaces. *Sampling, Wavelets and Tomography*, 2003.

[26] Y. C. Eldar and T. Werther. General framework for consistent sampling in hilbert spaces. *Int. J. Wavelets Multiresolut. Inf. Process.*, 3(4):497–509, 2005.

[27] E. Gauss. *Walsh Funktionen für Ingenieure und Naturwissenschaftler*. Springer Fachmedien, Wiesbaden, 1994.

[28] B. Golubov, A. Efimov, and V. Skvortsov. *Walsh Series and Transforms-Theory and Applications*. Springer, Netherlands, 1991.

[29] K. Gröchenig, Z. Rzeszotnik, and T. Strohmer. Quantitative estimates for the finite section method and banach algebras of matrices. *Integral Equations and Operator Theory*, 2(67):183–202, 2011.

[30] M. Guerquin-Kern, M. Häberlin, K. Pruessmann, and M. Unser. A fast wavelet-based reconstruction method for magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, 30(9):1649–1660, 2011.

[31] A. C. Hansen. On the approximation of spectra of linear operators on hilbert spaces. *J. Funct. Anal.*, 8(254):2092–2126, 2008.

[32] A. C. Hansen. On the solvability complexity index, the n-pseudospectrum and approximations of spectra of operators. *J. Amer. Math. Soc.*, 24(1):81–124, 2011.

[33] T. Hrycak and K. Gröchenig. Pseudospectral fourier reconstruction with the modified inverse polynomial reconstruction method. *J. Comput. Phys.*, 229(3):933–946, 2010.

[34] A. J. Jerri. The shannon sampling theorem – its various extensions and applications: A tutorial review. *Proc. IEEE*, (65):1565–1596, 1977.

[35] A. Jones, A. Tamtögl, I. Calvo-Almazán, and A. C. Hansen. Continuous compressed sensing for surface dynamical processes with helium atom scattering. *Nature Sci. Rep.*, 6:27776 EP –, 06 2016.

[36] G. Kutyniok and W.-Q. Lim. Compactly supported shearlets are optimally sparse. *J. Approx. Theory*, (163):1564–1589, 2011.

[37] G. Kutyniok and W.-Q. Lim. Optimal compressive imaging of fourier data. *SIAM Journal on Imaging Sciences*, 11(1):507–546, 2018.

[38] A. F. Lawrence, S. Phan, and M. Ellisman. Electron tomography and multiscale biology. In M. Agrawal, S. Cooper, and A. Li, editors, *Theory and Applications of Models of Computation*, volume 7287 of *Lecture Notes in Computer Science*, pages 109–130. Springer Berlin Heidelberg, 2012.

[39] R. Leary, Z. Saghi, P. A. Midgley, and D. J. Holland. Compressed sensing electron tomography. *Ultramicroscopy*, 131(0):70–91, 2013.

[40] M. Lindner. *Infinite matrices and their finite sections: An introduction to the limit operator method*. Birkhäuser Verlag, Basel, 2006.

[41] M. Lustig, D. Donoho, and J. M. Pauly. Sparse mri: the application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007.

[42] J. Ma. Generalized sampling reconstruction from fourier measurements using compactly supported shearlets. *Appl. Comput. Harmon. Anal.*, 2015.

[43] Y. Maday, T. Anthony, J. D. Penn, and M. Yano. Pbdw state estimation: Noisy observations; configuration-adaptive background spaces; physical interpretations. *ESAIM: Proceedings and Surveys*, 50:144–168, 2015.

[44] Y. Maday and O. Mula. A generalized empirical interpolation method: application of reduced basis techniques to data assimilation. In *Analysis and numerics of partial differential equations*, pages 221–235. Springer, 2013.

[45] Y. Maday, A. T. Patera, J. D. Penn, and M. Yano. A parameterized-background data-weak approach to variational data assimilation: formulation, analysis, and application to acoustics. *International Journal for Numerical Methods in Engineering*, 102(5):933–965, 2015.

[46] D. D.-Y. Po and M. N. Do. Directional multiscale modeling of images using the contourlet transform. *Trans. Img. Proc.*, 6(15):1610–1620, 2006.

[47] C. Poon. Structure dependent sampling in compressed sensing: theoretical guarantees for tight frames. *Applied and Computational Harmonic Analysis*, 42(3):402–451, 2017.

[48] E. T. Quinto. An introduction to X-ray tomography and Radon transforms. In *The Radon Transform, Inverse Problems, and Tomography*, volume 63, pages 1–23. American Mathematical Society, 2006.

[49] B. Roman, B. Adcock, and A. C. Hansen. On asymptotic structure in compressed sensing. *arXiv:1406.4178*, 2014.

[50] C. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, (27):379–423, 623–656, 1948.

[51] V. Studer, J. Bobin, M. Chahid, H. S. Mousavi, E. Candes, and M. Dahan. Compressive fluorescence microscopy for biological and hyperspectral imaging. *Proceedings of the National Academy of Sciences*, 109(26):E1679–E1687, 2012.

[52] M. Unser. Sampling - 50 years after shannon. *Proc. IEEE*, 4(88):569–587, 2000.

[53] M. Unser and J. Zerubia. A generalized sampling theory without band-limiting constraints. *IEEE Trans. Circuits Syst. II.*, 45(8):959–969, 1998.

DAMTP, UNIVERSITY OF CAMBRIDGE
E-mail address: ach70@cam.ac.uk

DAMTP, UNIVERSITY OF CAMBRIDGE
E-mail address: lt420@cam.ac.uk