

On reconstructions from measurements with binary functions

Robert Calderbank* Anders Hansen† Bogdan Roman‡
Laura Thesing†

Abstract

We consider the problem of reconstructions from linear measurements with binary functions. That is, the samples of the function are given by inner products with functions taking only the values 0 and 1. We consider three particular methods for this problem, the parametrised-background data-weak (PBDW)-method, Generalised sampling and infinite-dimensional compressed sensing (CS). The first two methods are dependent on knowing the stable sampling rate when considering samples by Walsh function and wavelet reconstruction. We establish linearity of the stable sampling rate, which is sharp, allowing for optimal use of these methods. In addition we provide recovery guaranties for infinite-dimensional compressed sensing with Walsh functions and wavelets.

1 Introduction

A classical problem in sampling theory is to reconstruct a function f , that is typically in $L^2([0, 1]^d)$, from linear measurements in the form of inner products. One of the most famous problems of this type is to reconstruct f from its Fourier coefficients, where one can view the Fourier coefficients as values obtained from inner products of f with the basis of complex exponentials. In a more general abstract form the problem is as follows.

Problem 1.1. *An element $f \in \mathcal{H}$, where \mathcal{H} is a separable Hilbert space, is to be reconstructed from measurements with linear functionals $(l_i)_{i \in \mathbb{N}} : \mathcal{H} \rightarrow \mathbb{C}$ that can be represented by elements $\zeta_i \in \mathcal{H}$ as $l_i(f) = \langle f, \zeta_i \rangle$. The key issue is that the l_i cannot be chosen freely, but are dictated by the modality of the sampling device.*

Classical Fourier sampling problems in applications include Magnetic Resonance Imaging (MRI), radio interferometry etc., which is a natural consequence of the frequent appearance of the Fourier transform in the sciences. However, there is another important form of measurements, namely sampling with binary functions. By sampling with binary functions we mean sampling with inner products of functions $\{\zeta_i\}_{i \in \mathbb{N}} \subset L^2([0, 1]^d)$ that only take the values 0 and 1. Just

*Department of Electrical and Computer Engineering, Duke University, USA

†Department of Applied Mathematics and Theoretical Physics, University of Cambridge, UK

‡Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, UK

as Fourier sampling occurs naturally in many sampling devices, sampling with binary functions is a phenomenon that occurs as a consequence of a sampling apparatus being "on" or "off", which occurs in digital signal processing, such as $\Sigma\Delta$ quantization, or newer forms of compressed measurements in microscopy or imaging.

There is a standard trick to convert sampling with binary functions to measurements with functions that take the values $\{-1, 1\}$ rather than $\{0, 1\}$, by multiplying every measurement by 2 and subtracting the sample done with the constant function. Thus, one may assume, if one is willing to accept a potential change in the statistical noise model, that the measurements are done with $\{-1, 1\}$ valued functions. One motivation for converting from $\{0, 1\}$ valued sampling to $\{-1, 1\}$ is that the latter allows for the use of Walsh functions. These functions have very similar qualities to the complex exponentials in Fourier, and the Walsh transform is a close cousin of the Fourier transform. Moreover, the classical discrete Fourier transform obeys a fast implementation. This is also existent for the Walsh case via the Hadamard transform.

Given the extensive theory of Walsh functions and the benefits listed above we will from now on assume that the sampling functions $\{\zeta_i\}_{i \in \mathbb{N}} \subset L^2([0, 1]^d)$ are the Walsh functions. A bonus property of the Walsh functions is that when combined with wavelets φ_j , $j \in \mathbb{N}$, spanning $L^2([0, 1]^d)$, in a change of basis matrix

$$U = \{\{\varphi_j, \zeta_i\}\}_{i, j \in \mathbb{N}}, \quad (1)$$

one obtains a very structured infinite matrix. This infinite matrix shares many structural similarities with the change of basis matrix obtained by combining complex exponentials and wavelets. In particular, both types of infinite matrices become almost block diagonal, a feature that will be highly useful as we will see below.

In this paper we will address three methods for Problem 1.1, two linear and one non-linear method. We choose the two linear ones because of their optimality with respect to the reconstruction error. The first is optimal in the class of linear methods that are consistent with the measurement and the second is optimal for linear algorithms that map into the reconstruction space. The non-linear method is the algorithm which takes the most structure into account and hence allows very good reconstruction guarantees. In all cases we assume sampling with Walsh functions. The methods are as follows.

- (i) The parametrised-background data-weak (PBDW)- method (linear)
- (ii) Generalised sampling (linear)
- (iii) Infinite-dimensional compressed sensing (non-linear)

The PBDW-method originated with the work by Maday, Patera, Penn and Yano in (51), and was further developed and analysed by by Binev, Cohen, Dahmen, DeVore, Petrova, and Wojtaszczyk (10, 22, 13). Generalised sampling has been studied by Adcock, Hansen, Hrycak, Gröchenig, Kutyniok, Ma, Poon, Shadrin and others (1, 4, 2, 43, 41, 6, 50), and the predecessor; consistent sampling,

has been analysed by Aldroubi, Eldar, Unser and others (8, 64, 29, 30, 31, 32). Infinite-dimensional compressed sensing has been developed and studied by Adcock, Hansen, Kutyniok, Lim, Poon and Roman (7, 47, 5, 55).

The successful use of the two first methods when reconstructing in a wavelet basis is completely dependent on the stable sampling rate which is defined below in terms of subspace angles between sampling and reconstruction spaces. It dictates the size of the sampling space as a function of the dimension of the reconstruction space in order to ensure accurate reconstructions. The main question we want to answer is:

What is the stable sampling rate when given Walsh samples and a wavelet reconstruction basis?

The key issue is that the error bounds for these methods depend (sharply) on the subspace angle, and fortunately we can provide sharp results on the stable sampling rate. In the case of infinite-dimensional compressed sensing one cannot directly use the stable sampling rate, however, we provide estimates on the size of the sampling space as well as recovery guarantees from sub-sampled data.

To define the stable sampling rate we need to introduce some notation. The goal is to reconstruct f from the finite number of samples $\{l_i(f)\}_{i=1}^M$ for some $M \in \mathbb{N}$. The space of the functions ζ_i is called the *sampling space* and is denoted by $\mathcal{S} = \overline{\text{span}}\{\zeta_i : i \in \mathbb{N}\}$, meaning the closure of the span. In practice, one can only acquire a finite number of samples. Therefore, we denote by $\mathcal{S}_M = \text{span}\{\zeta_i : i = 1, \dots, M\}$ 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 $(\varphi_i)_{i \in \mathbb{N}}$, i.e. $\mathcal{R} = \overline{\text{span}}\{\varphi_i : i \in \mathbb{N}\}$. As in the case 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 $\mathcal{R}_N = \text{span}\{\varphi_i : i = 1, \dots, N\}$.

One of the main questions in reconstruction theory is how many samples are needed to guarantee a stable and accurate recovery? In the first part of this paper we want to analyse this question for linear methods. The *stable sampling rate* captures the number of samples necessary to obtain a stable and accurate reconstruction of a certain number of coefficients in the reconstruction space. More precisely, we are interested in the dimension of the sampling space \mathcal{S}_M in relation to the reconstruction space \mathcal{R}_N . Section 4 talks about the linear reconstruction method, and there we see that the accuracy and stability of the methods depend on the subspace angle between \mathcal{R}_N and \mathcal{S}_M . In particular,

$$\cos(\omega(\mathcal{R}_N, \mathcal{S}_M)) := \inf_{r \in \mathcal{R}_N, \|r\|=1} \|P_{\mathcal{S}_M} r\|, \quad (2)$$

where $P_{\mathcal{S}_M}$ is the orthogonal projection onto the sampling space. Mainly, one is interested in the reciprocal value

$$\sigma(\mathcal{R}_N, \mathcal{S}_M) = 1/\cos(\omega(\mathcal{R}_N, \mathcal{S}_M)) \in [1, \infty], \quad (3)$$

which, as we will see later, plays a key role in all the error estimates of the two linear algorithms discussed here. Due to the definition of cosine, σ takes values in

$[1, \infty]$. The stable sampling rate is then given by

$$\Theta(N, \theta) = \min \{M \in \mathbb{N} : \sigma(\mathcal{R}_N, \mathcal{S}_M) \leq \theta\}. \quad (4)$$

This function was analysed for different reconstruction methods in the Fourier case. We know that the stable sampling rate is linear for Fourier-wavelet (6) and Fourier-shearlet (50) reconstructions. This is the best one can wish for as it allows to reconstruct nearly as well from Fourier samples as sampling directly in the wavelet or shearlet system. However, this is not always the case. In the Fourier-polynomial reconstruction we get that the stable sampling rate is polynomial which leads to a large number of necessary samples and makes this only feasible for very sparse signals. For the Walsh case it was shown in (40) that the stable sampling rate is also linear in the Walsh-wavelet case and it is even possible to determine the slope for Walsh-Haar wavelets (60).

The analysis for the non-linear reconstruction is a bit more involved and needs a detailed analysis of the change of basis matrix as well as the reconstruction space. It is common to use the sparsity of the coefficients in the wavelet space of natural images. However, it is known that this can be described more detailed with structured sparsity. This reduces the size of the class and hence allows better reconstruction guarantees. Similarly we have that the change of basis matrix is not incoherent but asymptotically incoherent. This leads to a new version of CS with highly improved reconstruction quality from fewer samples. We will see that the impact of elements outside the diagonal boxes decays exponentially. This allows us to use more sub-sampling than previously described by the classical CS literature. We are here as well mainly interested in the reconstruction from Walsh samples with wavelets.

This paper is structured as follows. First we discuss in chapter 2 the Walsh functions which are the building block of the sampling space. Then we revise the basics about boundary corrected wavelets in chapter 3. With this information at hand we are able to present the linear reconstruction methods and their analysis in chapter 4. We then continue with the non-linear method. We analyse the change of basis matrix and discuss the classical theory, different problem types and sum up with a detailed comparison between the new theory and the older versions. This is underlined by some numerical examples.

2 Walsh functions

In this section we introduce Walsh functions, which span the sampling space \mathcal{S}_M . First, we want to discuss the use of multi-indices. This is important because we want to deal with one- and d -dimensional functions. A multi-index j is commonly defined by $j = (j_1, \dots, j_d) \in \mathbb{N}^d$, $d \in \mathbb{N}$. The basic operations such as addition and multiplication are understood point-wise, i.e. $j \star r = (j_1 \star r_1, \dots, j_d \star r_d)$. This can also be done with a natural number n which is then interpreted as a multi-index with the same entry $n = (n, \dots, n)$. Finally, the sum over a vector indexed by a

multi-index is given by

$$\sum_{j=k}^r x_j := \sum_{j_1=k_1}^{r_1} \cdots \sum_{j_d=k_d}^{r_d} x_{j_1, \dots, j_d}, \quad (5)$$

where $k, r \in \mathbb{N}^d$.

The multi-indices can be used to define functions in higher dimensions by the tensor product of the one-dimensional functions. The tensor product of $f : \mathbb{R} \rightarrow \mathbb{R}$ is given by

$$f(x) = f(x_1) \cdot \dots \cdot f(x_d), \quad (6)$$

where $\{x_i\}_{i=1, \dots, d} = x \in \mathbb{R}^d$ with $x_i \in \mathbb{R}$. Hence, the input parameter defines the dimensions. This simplifies the transition between the one- and d -dimensional case.

2.1 Definition

It is important to notice that Walsh functions behave very similarly to the complex exponential functions when the setting is changed from the decimal to the dyadic analysis. Dyadic analysis is a framework where functions are analysed for the situation where decimal addition is replaced by dyadic addition. Therefore, we start with a short review of the dyadic representation and addition. Let $x \in \mathbb{R}_+$, the dyadic representation is given by

$$x = \sum_{i \in \mathbb{Z}} x_i 2^i, \quad (7)$$

where $x_i \in \{0, 1\}$ for all $i \in \mathbb{Z}$. To make this representation unique we use the one that ends in 0 instead of 1 if there is a choice. The dyadic addition of two numbers $x, y \in \mathbb{R}_+$ is given by

$$x \oplus y = \sum_{i \in \mathbb{Z}} (x_i \oplus_2 y_i) 2^i, \quad (8)$$

where $x_i \oplus_2 y_i$ is addition modulo two, i.e. $0 \oplus_2 0 = 0, 0 \oplus_2 1 = 1, 1 \oplus_2 0 = 1, 1 \oplus_2 1 = 0$. This definition can also be extended to all \mathbb{R} and works as in the decimal case. We use the convention to denote negative numbers with a minus sign in front of the dyadic representation of the absolute value.

With this information at hand, we can now define the Walsh functions, which span the sampling space \mathcal{S}_M .

Definition 2.1 ((36)). *Let $t \in \mathbb{N}$ and $x \in [0, 1)$ with the dyadic representation (t_0, \dots) and (\dots, x_{-1}) . Then there exists a unique $n = n(t) \in \mathbb{N}$ such that $t = \sum_{i=0}^{n-1} t_i 2^i$, in particular $t_n \neq 0$ and $t_k = 0$ for all $k > n$. Let $t^n = (t_0, \dots, t_n)$ and for $x = \sum_{i=-\infty}^{-1} x_i 2^i$ define $x^n = (x_{-n}, \dots, x_{-1})$, and $C_W : \mathbb{R}^n \mapsto \mathbb{R}^n$ by*

$$C_W = \begin{pmatrix} 0 & \cdots & 0 & 1 & 1 \\ \vdots & \ddots & \ddots & 1 & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ 1 & 1 & \ddots & & \vdots \\ 1 & 0 & \cdots & \cdots & 0 \end{pmatrix}. \quad (9)$$

The Walsh functions are then given by

$$\text{wal}(t; x) = (-1)^{t^n \cdot C_W x^n}. \quad (10)$$

This definition is a bit longer to write, however, it gives an interesting insight of the ordering of the Walsh functions. There are a lot of different orderings of the Walsh functions available. The first choice for the matrix C_W might be the identity. The functions are then called Walsh-Kronecker functions. The problem with this ordering is that the functions change completely if the maximal element $n(t)$ is changed. Therefore, they are seldom used in practice. One attempt to overcome this problem is the Walsh-Paley ordering which is given by the reversal matrix:

$$C_{WP} = \begin{pmatrix} 0 & \cdots & 0 & 0 & 1 \\ \vdots & \ddots & \ddots & 1 & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & 1 & \ddots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 0 \end{pmatrix}. \quad (11)$$

In this scenario the functions stay the same with changing $n(t)$. Hence they overcome one drawback of the Walsh-Kronecker ordering. However, they are not ordered with increasing number of zero crossing. This would be a desirable property because, first it makes them similar to exponential functions and second it relates well to the level ordering of the wavelets. Therefore, we use the presented definition as it obeys none of the discussed drawbacks.

It is also possible to extend the classical Walsh functions to inputs in $\mathbb{R}_+ \times \mathbb{R}_+$, i.e.

$$\text{Wal}(t, x) = (-1)^{t_0 x_0} \text{wal}([t]; x) \text{wal}([x]; t), \quad (12)$$

where t and x have the dyadic representation $(t_i)_{i \in \mathbb{Z}}$ and $(x_i)_{i \in \mathbb{Z}}$, t_0, x_0 are the corresponding elements of the sequence and $[\cdot]$ denotes the rounding down operation. We get the same functions if C_W is defined over \mathbb{N} instead of $\{1, \dots, n(t)\}$. For negative inputs we take the same definition as in (36)

$$\text{Wal}(-t, x) := -\text{Wal}(t, x) \quad (13)$$

$$\text{Wal}(t, -x) := -\text{Wal}(t, x). \quad (14)$$

With the presentation of the multi-indices and the generalised Walsh functions it is now easy to define them in higher dimensions with the tensor product by

$$\text{Wal}(t, x) = \bigotimes_{k=1}^d \text{Wal}(t_k, x_k), \quad (15)$$

where $t = (t_k)_{k=1, \dots, d}$, $x = (x_k)_{k=1, \dots, d} \in \mathbb{R}^d$. These function then span the sampling space, i.e.

$$\mathcal{S} = \overline{\text{span}} \{ \text{Wal}(t, \cdot), t \in \mathbb{N}^d \} \subset L^2([0, 1]^d) \quad (16)$$

and for $M = m^d$ for some $m \in \mathbb{N}$ we have

$$\mathcal{S}_M = \text{span} \{ \text{Wal}(t, \cdot), t_i \leq m, i = 1, \dots, d \} \subset L^2([0, 1]^d) \quad (17)$$

For discrete signals in \mathbb{C}^N the orthogonal projection onto the sampling space is often denoted by Ψ , which we shall discuss later on.

Finally, we define a continuous and a discrete transform. For the definition of the continuous transform we have to ensure that the integral exists. Therefore, let $f \in L^2([0, 1]^d)$ then the *generalised Walsh transform* is given almost everywhere by

$$\widehat{f}^w(t) = \mathcal{W}\{f(\cdot)\}(t) = \langle f(\cdot), \text{Wal}(t, \cdot) \rangle = \int_{[0, 1]^d} f(x) \text{Wal}(t, x) dx, \quad t \in \mathbb{R}^d. \quad (18)$$

The restrictions to functions which are supported in $[0, 1]^d$ leads to the use of boundary corrected wavelets which are presented in §3. For the discrete transform let $N = 2^n$, $n \in \mathbb{N}$ and $x = (x_0, \dots, x_{N-1}) \in \mathbb{R}^N$ then the one dimensional *discrete Walsh transform* of x is given by $X = (X_0, \dots, X_{N-1})$ with

$$X_j = \frac{1}{N} \sum_{k=0}^{N-1} x_k \text{Wal}\left(j, \frac{k}{N}\right). \quad (19)$$

As discussed before, the Walsh functions are desirable because of the fast transform. It can be seen here, that this indeed corresponds to the Hadamard transform and therefore the Walsh functions are its kernel.

In higher dimensions we get for $x \in \mathbb{R}^{N_1 \times \dots \times N_d}$ where $x_{k_i} \in \mathbb{R}$, $k = (k_i)_{i=1, \dots, d}$, $k_i = 0, \dots, N_i - 1$ the discrete Walsh transformed $X = (X_j) \in \mathbb{R}^{N_1 \times \dots \times N_d}$, where $X_{j_i} \in \mathbb{R}$, $j = (j_i)_{i=1, \dots, d}$, $j_i = 0, \dots, N_i - 1$, with

$$X_j = \frac{1}{\prod_{i=1}^d N_i} \sum_{k=0}^{N-1} x_k \text{Wal}\left(j, \frac{k}{N}\right). \quad (20)$$

2.2 Properties

In this section we recall the most important and useful properties of the Walsh functions and transfer them to the continuous transform. The Walsh functions are symmetric,

$$\text{Wal}(t, x) = \text{Wal}(x, t) \quad \forall t, x \in \mathbb{R}, \quad (21)$$

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

$$\text{Wal}(2^k t, x) = \text{Wal}(t, 2^k x) \quad \forall t, x \in \mathbb{R}, \quad k \in \mathbb{N} \quad (22)$$

and

$$\text{Wal}(t, x) \text{Wal}(t, y) = \text{Wal}(t, x \oplus y) \quad \forall t, x, y \in \mathbb{R}. \quad (23)$$

Due to the tensor product definition these properties also hold in the d -dimensional case. Moreover, we have for the transform, that it is linear:

$$\mathcal{W}\{af(x) + bg(x)\} = a\mathcal{W}\{f(x)\} + b\mathcal{W}\{g(x)\} \quad \forall a, b \in \mathbb{R}, \quad f, g \in L^2([0, 1]^d), \quad (24)$$

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

$$\mathcal{W}\{f(x \oplus y)\}(t) = \mathcal{W}\{f(y)\}(t) \text{Wal}(x, t) \quad \forall x \in \mathbb{R}^d, \quad f \in L^2([0, 1]^d) \quad (25)$$

and

$$\mathcal{W}\{f(2^m x)\}(t) = \frac{1}{2^m} \mathcal{W}\{f(x)\}\left(\frac{t}{2^m}\right) \quad \forall m \in \mathbb{N}^d, \quad f \in L^2([0, 1]^d). \quad (26)$$

3 Reconstruction space

The reconstruction space should be chosen appropriately to the given data. For image and audio signals wavelets have proven to be very useful as they are able to present the data sparsely. In the following we will deal with Daubechies wavelets. Normally, they are defined on the whole \mathbb{R}^d . However, we need to have that $\mathcal{S}_M^+ \cap \mathcal{R}_N = \{0\}$ because otherwise there are elements in the reconstruction space which cannot be captured with the sampling space and makes it impossible to have a unique solution to the reconstruction problem. Hence, we have to restrict ourselves to wavelets that are only defined on the cube $[0, 1]^d$. For this sake we use boundary corrected wavelets and in higher dimensions separable boundary corrected wavelets which are constructed by tensor products. We follow the construction as in (20).

For a smoother outline of boundary wavelets we start with the one-dimensional case. We denote the mother wavelet with ψ and the corresponding scaling function with ϕ , which is equal to the common literature in this area. The corresponding wavelet and scaling spaces are spanned by the scaled and translated versions

$$\psi_{r,j}(x) := 2^{r/2}\psi(2^r x - j) \text{ and } \phi_{r,j}(x) := 2^{r/2}\phi(2^r x - j), \quad (27)$$

where $r, j \in \mathbb{Z}$. With this we obtain the wavelet space $W_r := \text{span}\{\psi_{r,j} : j \in \mathbb{Z}\}$ at level r and accordingly the scaling space $V_r := \text{span}\{\phi_{r,j} : j \in \mathbb{Z}\}$. As discussed in the beginning of the chapter and in the previous chapter we have to restrict ourselves to functions defined on $[0, 1]^d$. Therefore, we take boundary corrected Daubechies wavelets which are introduced in chapter 4 in (20). They have two major advantages. The first is the maintained smoothness and compactness properties of the original wavelet. Second, they also keep the multi-resolution analysis. This is important for the definition of the higher dimensional wavelets. It allows us to keep the structure also in higher dimensions. We can still represent the reconstruction space with the scaling space in only one level.

We start with the scaling function at the level J_0 such that the functions can only intersect with one boundary at a time. The scaling space is then given by

$$V_{J_0}^b = \text{span}\left\{\phi_{J_0,j} : j = 0, \dots, 2^{J_0} - p - 1, \phi_{J_0,j}^\# : j = 2^{J_0} - p, \dots, 2^{J_0} - 1\right\}, \quad (28)$$

where $\phi^\#$ is the scaling function reflected at 1. The definition for higher levels $r > J_0$ works accordingly. We denote the boundary wavelets by ψ^b and $\psi_{j,m}^b(x) = 2^{j/2}\psi(2^j x - m)$ for $j \geq J_0$. We are only interested in the smoothness properties of the wavelet, which stay the same to the generating wavelet ψ . Therefore, we do not get into the details about the construction of the ψ^b . Interested readers should seek out for (20) for a detailed explanation. The wavelet space is then given by

$$W_r^b = \text{span}\{\psi_{r,j}^b : j = 0, \dots, 2^r\}. \quad (29)$$

For the linear reconstruction methods it suffices to only consider the reconstruction space as a whole. Therefore, we exploit the multi-resolution analysis and we can

represent the wavelet spaces up to level $R - 1$ by the scaling space at level R , i.e.

$$\bigcup_{r < R} W_r^b = V_R^b. \quad (30)$$

This allows us for a number of coefficients related to the levels, i.e. $N = 2^R$ to represent the reconstruction space as

$$\mathcal{R}_N := V_R^b. \quad (31)$$

This has the positive byproduct that we do not have to deal with the internal ordering.

For the non-linear methods this internal ordering becomes more important. We then let

$$\mathcal{R}_N := V_{J_0}^b \oplus W_{J_0}^b \dots \oplus W_{R-1}^b. \quad (32)$$

We now get to the definition in higher dimensions. The scaling space is defined by the tensor product, i.e.

$$\mathcal{R}_N = V_R^{b,d} := V_R^b \otimes \dots \otimes V_R^b \quad (\text{d-times}) \quad (33)$$

for $N = 2^{dR}$. It is important to note that the wavelet space in higher dimensions is not simply the tensor product of the one-dimensional wavelets, but the combination of wavelets and scaling functions, i.e.

$$V_j^{b,d} = V_j^b \otimes \dots \otimes V_j^b = (V_{j-1}^b \oplus W_{j-1}^b) \otimes \dots \otimes (V_{j-1}^b \oplus W_{j-1}^b) = V_{j-1}^{b,d} \oplus W_{j-1}^{b,d}. \quad (34)$$

And hence,

$$W_{j-1}^{b,d} := (V_{j-1}^b \oplus W_{j-1}^b) \otimes \dots \otimes (V_{j-1}^b \oplus W_{j-1}^b) \ominus V_{j-1}^{b,d}. \quad (35)$$

Let $\phi_{J_0,m}^{b,d} = \otimes_{i=1}^d \phi_{J_0,m_i}^b$ and $\psi_{j,m}^{b,d} = \otimes_{i=1}^d \psi_{j,m_i}^b$, where ϕ^b can stand for ϕ or $\phi^\#$ depending on m . For the reconstruction space this results in

$$\mathcal{R} = \left\{ \phi_{J_0,m}^{b,d}, m = (m_1, \dots, m_d), m_i = 0, \dots, 2^{J_0} - 1 \right. \quad (36)$$

$$\left. \phi_{j,m}^{b,d-1} \otimes \psi_{j,m}, \dots, \phi_{j,m}^b \otimes \psi_{j,m}^{b,d-1}, \psi_{j,m}^{b,d} \right. \quad (37)$$

$$\left. j \geq J_0, m = (m_1, \dots, m_d), m_i = 0, \dots, 2^j - 1 \right\}. \quad (38)$$

Note the abuse of notation in $\phi_{j,m}^{b,d-r} \oplus \psi_{j,m}^{b,r}$. Only the parts of the multi-index $m = (m_1, \dots, m_d)$ related to the position of the function in the tensor product are used for the shift of the function. Moreover, we have 2^d different possibilities to combine the scaling function and the wavelets by the tensor product. Hence, there are $2^{dj}(2^d - 1)$ elements at every level j . In case of doubt of the dimension we will use an upper index d to make the distinction clear, i.e. for $u_{i,j}^d$. For the discrete setting the orthogonal projection onto the reconstruction space is often denoted by Φ , which will be discussed in more details with the numerical experiments.

4 Linear Reconstruction Methods

In this section we are concerned with two different linear reconstruction methods the *PBDW-method* and *generalised sampling*. They both have in common that they are linear and share the same condition number and that the accuracy is highly dependent on the stable sampling rate which is analysed in the last subsection.

4.1 PBDW-method

The PBDW-method as introduced in (51) and analysed in (10, 13, 22) is based on the following idea. Given the measurements $l := P_{\mathcal{S}_M} f$, where $P_{\mathcal{S}_M}$ denotes the orthogonal projection onto the subspace \mathcal{S}_M , one tries to find an approximation that is consistent with the measurements and close to the reconstruction space \mathcal{R}_N which is measured with the distance $\text{dist}(f, \mathcal{R}_N) = \min \{\|f - \varphi\|_2 : \varphi \in \mathcal{R}_N\}$ and bounded by a sequence $\{\epsilon_N\}_{N \in \mathbb{N}}$. Mathematically, one tries to approximate f by $f^* \in \mathcal{K}_l$ where we define

$$\mathcal{K} = \{f \in \mathcal{H} : \text{dist}(f, \mathcal{R}_N) \leq \epsilon_N\} \text{ and } \mathcal{H}_l = \{f \in \mathcal{H} : P_{\mathcal{S}_M} f = l\}, \quad (39)$$

and the space of possible approximation is then the intersection $\mathcal{K}_l := \mathcal{K} \cap \mathcal{H}_l$. Obviously, we try to find the closest element $f^* \in \mathcal{K}_l$ to the true solution f , hence we solve the minimization problem

$$g^* = \text{argmin}_{g \in \mathcal{R}_N} \|l - P_{\mathcal{S}_M} g\|^2. \quad (40)$$

The outcome g^* is then adjusted to be consistent with the measurements by

$$f^* = l + P_{\mathcal{S}_M^\perp} g^*. \quad (41)$$

Then f^* is the solution to the PBDW-method and was analysed in (13) and shown to be optimal with respect to the distance to the true function for all functions that are consistent with the measurements. We have the error estimate

$$\|f - f^*\| \leq \sigma(\mathcal{R}_N, \mathcal{S}_M) \text{dist}(f, \mathcal{R}_N). \quad (42)$$

This error estimate was then improved in (51) to

$$\|f - f^*\| \leq \sigma(\mathcal{R}_N, \mathcal{S}_M) \text{dist}(f, \mathcal{R}_N \oplus (\mathcal{S}_M \cap \mathcal{R}_N^\perp)). \quad (43)$$

However, it was shown in (13) that the factor of the subspace angle $\sigma(\mathcal{R}_N, \mathcal{S}_M)$ cannot be removed or improved.

This underlines again that it is important to make sure that $\mathcal{R}_N \cap \mathcal{S}_M^\perp = \{0\}$. Moreover, it underlines the importance of the stable sampling rate and estimates of the relation between the number of samples M and the number of reconstructed coefficients N to get a stable and accurate reconstruction. In the next section we discuss the concept of generalised sampling and see that the condition number of the PBDW-method also equals the subspace angle, which underlines its importance.

4.2 Generalised Sampling

We now study a different linear reconstruction technique: *generalised sampling*. Unlike PBDW, it forces the solution to stay in the reconstruction space. In particular, for very sparse data in the reconstruction space, it improves the reconstruction quality.

The method is an extension of the finite section methods (14, 38, 39, 48). In important cases like Fourier-wavelet or Walsh-wavelet the finite section method is very unstable. The advantage of generalised sampling is that it allows a different number of samples than reconstructed coefficients, which makes the method stable and accurate. The question of how to choose the number of measurements with respect to the number of coefficients is answered by the stable sampling rate. We give this method now, and then explain how it can be cast into a least squares problem.

Definition 4.1 ((1)). *For $f \in \mathcal{H}$ and $N, M \in \mathbb{N}$ with $M \geq N$ we define the reconstruction method of generalised sampling $G_{N,M} : \mathcal{H} \rightarrow \mathcal{R}_N$ by*

$$\langle P_{S_M} G_{N,M}(f), \varphi_i \rangle = \langle P_{S_M} f, \varphi_i \rangle, \quad i = 1, \dots, N, \quad (44)$$

where $\varphi_i, i = 1, \dots, N$ span \mathcal{R}_N . We refer to $G_{N,M}(f)$ as the generalised sampling reconstruction of f .

Equation [Eq. 44] can be rewritten as the following least squares problem: We search for a solution $\alpha^{[N]} \in \mathbb{R}^N$ of

$$U^{[N,M]} \alpha^{[N]} = l(f)^{[M]}, \quad (45)$$

where

$$U^{[N,M]} = \begin{pmatrix} u_{11} & \dots & u_{1N} \\ \vdots & \ddots & \vdots \\ u_{M1} & \dots & u_{MN} \end{pmatrix} \quad (46)$$

and $u_{ij} = \langle \varphi_j, \zeta_i \rangle$, $l(f)^{[M]} = (l_1(f), \dots, l_M(f)) \in \mathbb{R}^M$. The solution of the method is then given by $G_{N,M}(f) = \sum_{i=1}^N \alpha_i \varphi_i$. The change of basis matrix U can be seen in Figure 1 for Walsh measurements and different wavelets.

Generalised sampling was widely studied and it was shown that [Eq. 44] yields a solution if the number of samples is large enough.

Theorem 4.2 ((2)). *Let $N \in \mathbb{N}$. Then, there exists $M_0 \in \mathbb{N}$ such that for every $f \in \mathcal{H}$ Equation [Eq. 44] has a unique solution $G_{N,M}(f)$ for all $M \geq M_0$. Moreover, the smallest M_0 is the least number such that $\cos(\omega(\mathcal{R}_N, \mathcal{S}_{M_0})) > 0$.*

Additionally, the condition number and optimality was analysed. Here it is interesting to notice, that the generalised sampling as well as the PBDW-method are optimal in their setting and that in both cases the performance depends on the subspace angle between the sampling and reconstruction space $\sigma(\mathcal{R}_N, \mathcal{S}_M)$.

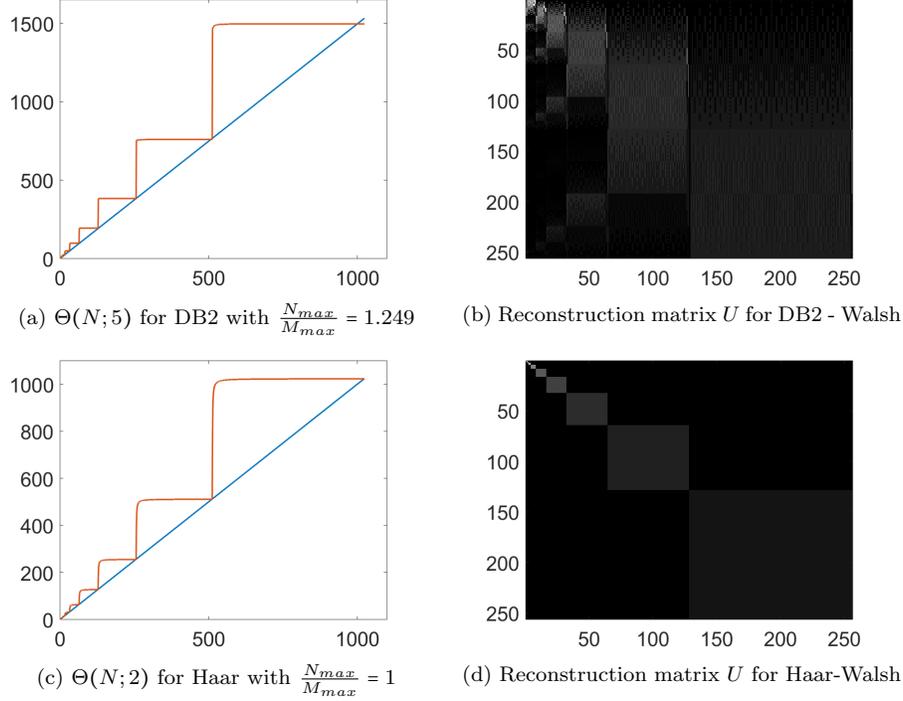


Figure 1: Stable sampling rate and change of basis matrix for different wavelets with Walsh functions.

Theorem 4.3 ((2)). *Retaining the definitions and notations from this section, for all $f \in \mathcal{H}$ we have*

$$\|G_{N,M}(f)\| \leq \sigma(\mathcal{R}_N, \mathcal{S}_M) \|f\|, \quad (47)$$

and

$$\|f - P_{\mathcal{R}_N} f\| \leq \|f - G_{N,M}(f)\| \leq \sigma(\mathcal{R}_N, \mathcal{S}_M) \|f - P_{\mathcal{R}_N} f\|. \quad (48)$$

In particular, these bounds are sharp.

Remark that the same least square problem is solved for the PBDW-method and generalised sampling. Therefore, the analysis of the condition number $\kappa(\mathcal{R}_N, \mathcal{S}_M)$ of the generalised sampling approach in (4) translates directly to the PBDW-method. We get

$$\kappa(\mathcal{R}_N, \mathcal{S}_M) = \sigma(\mathcal{R}_N, \mathcal{S}_M). \quad (49)$$

Hence, the stable sampling rate is important to analyse the accuracy and also the stability.

4.3 The stable sampling rate for the Walsh-wavelet case

In this section we recall the results from (40) about the stable sampling rate for the Walsh-wavelet case.

Theorem 4.4 ((40)). *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 $\sigma(\mathcal{R}_N, \mathcal{S}_M) \leq \theta$. In particular one gets $\Theta \leq S_\theta N$. Hence, the relation $\Theta(N; \theta) = \mathcal{O}(N)$ holds for all $\theta \in (1, \infty)$.*

In Figure 1 the stable sampling rate is displayed for different Daubechies wavelets. One can see that the slope S_θ is smaller for wavelets with a more block diagonal change of basis matrix. A direct relation between the number of vanishing moments and the value of S_θ is not known due to the very different behaviour of Walsh functions and wavelets.

One should note that for the case of Haar wavelets the slope $S_\theta = 1$ for all $\theta \in (1, \infty)$. This relation was analysed in more detail in (60).

Theorem 4.5 ((60)). *Let the sampling space \mathcal{S} be spanned by the Walsh functions and the reconstruction space \mathcal{R} by the Haar wavelets in $L^2([0, 1]^d)$. If $N = 2^{dR}$ for some $R \in \mathbb{N}$, then for every $\theta \in (1, \infty)$ we have that the stable sampling rate is the identity, i.e. $\Theta(N, \theta) = N$.*

These results show that sampling with Walsh functions is nearly as good as sampling directly with wavelets. Hence, the presented algorithms allow to improve the recovery quality. This can be seen in the comparison with direct inversion where one gets a lot of block artefacts from Walsh functions or the Gibbs phenomena with Fourier samples. These are mostly removed after the reconstruction method. We can analyse mathematically the approximation rate in the different bases.

4.3.1 Approximation qualities

Approximation theory provides a useful tool for comparing the representation qualities of different bases. Let $\{\varphi_i\}_{i \in \mathbb{N}}$ be an orthonormal basis for $L^2([0, 1])$, hence every $f \in L^2([0, 1])$ can be represented by

$$f = \sum_{i \in \mathbb{N}} \langle f, \varphi_i \rangle \varphi_i. \quad (50)$$

In practice this is not a feasible representation approach. This is due to the fact that we can only access and store a finite number of coefficients. Hence, instead of the true f we can only have an approximation $f_N = \sum_{i=1}^N \langle f, \varphi_i \rangle \varphi_i$. The resulting approximation error is given by

$$\epsilon(N, f) = \|f - f_N\|_2^2 = \int |f - f_N|^2 dx = \sum_{i > N} |\langle f, \varphi_i \rangle|^2. \quad (51)$$

In approximation theory one compares bases and representation systems in terms of the decay of $\epsilon(N, f)$ with respect to N for functions f in some specified function class. A very fast decay with N is desirable, because this allows a good representation from only a few coefficients, which then results in less measurements.

The decay rate of the Walsh transform of Lipschitz continuous functions is analysed in (9). It was shown that

$$\langle f, \text{Wal}(n, \cdot) \rangle = \widehat{W}f(n) \leq 2^{-p}, \quad (52)$$

where $2^p \leq n < 2^{p+1}$. With this we get for the approximation error

$$\epsilon(N, f) \leq \sum_{i>N} \frac{1}{2i^2} \leq \frac{1}{2N}$$

which then lies in $\mathcal{O}(N^{-1})$. In contrast to the Fourier transform this does not improve if the function gets smoother or periodic. The resulting artefacts can be seen in Figure 2. Therefore, the reconstruction techniques as the PBDW-method or generalised sampling are very useful because they allow to change the basis in which we represent our data. We then use a basis such as wavelets with a much faster decay rate. The decay rate is analysed in (52). Daubechies wavelets of order p represent functions f in the Sobolev space $W^\gamma([0, 1])$ for some $\gamma < p$ with an approximation rate of

$$\epsilon(N, f) = \mathcal{O}(N^{-2\gamma}). \quad (53)$$

This underlines the advantage of representing smooth functions with Daubechies wavelets instead of Walsh functions. Due to the findings in Theorem 4.4 it is possible to highly improve the reconstruction quality from measurements with binary functions. In particular, because of the linearity of the stable sampling rate and a reasonable slow slope, for example $S_2 = 2$ for Daubechies wavelets of order 8, we get an improved representation from $\mathcal{O}(N^{-1})$ to $\mathcal{O}(N^{-2\gamma})$.

5 Non-linear reconstruction methods

In the previous section we have seen linear reconstruction methods and discussed their convergence properties in view of subspace angles and the stable sampling rate. Even though they offer good results and fast computations, they are rather restrictive in terms of adaptivity to the problem at hand, in particular, they do not allow for sub-sampling. Hence, we want to extend our analysis to the non-linear methods where we focus on compressed sensing and the structure, provided by measurements with binary functions, that allows for substantial under sampling. The main issue with this extension is that classical compressed sensing considers finite-dimensional signals. However, we are dealing with infinite-dimensional ones. Hence, the classic compressed needs to be extended to infinite-dimensional compressed sensing introduced in (5, 7). Nevertheless, we start this chapter with a quick review of the standard finite-dimensional compressed sensing.

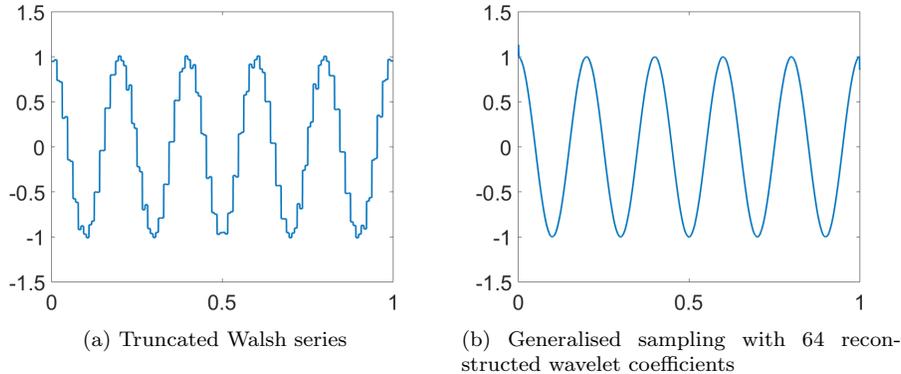


Figure 2: Reconstruction from 77 measurements with binary functions, where both examples use exactly the same samples. The figure illustrates the change in approximation rate when converting the Walsh samples to wavelet coefficients via generalised sampling.

5.1 Classical compressed sensing

Compressed sensing was introduced by Candès, Romberg & Tao (19) and Donoho (24) and is formulated in the finite-dimensional setting, stating that under appropriate conditions one can overcome the Nyquist sampling barrier and recover signals using far fewer samples than dictated by the classical Shannon theory.

A traditional CS set-up is as follows. The aim is to recover a signal f from an incomplete (sub-sampled) set of measurements y . Here, f is represented as a vector in \mathbb{C}^N and is assumed to be s -sparse in some orthonormal basis $\Phi \in \mathbb{C}^{N \times N}$ (e.g. wavelets) called the *reconstruction* or *sparsity* basis. This means that its vector of coefficients $x = \Phi f$ has at most s nonzero entries. Let $\Psi \in \mathbb{C}^{N \times N}$ be an orthonormal basis, called *sensing* or *sampling* basis, and write $U = \Psi \Phi^* = (u_{ij})$, which is an isometry and the discrete version of U in [Eq. 46]. The coherence of U is

$$\mu(U) = \max_{i,j} |u_{ij}|^2 \in [1/N, 1]. \quad (54)$$

and U is said to be perfectly incoherent if $\mu(U) = 1/N$.

Let the *sub-sampling pattern* be the set $\Omega \subseteq \{1, \dots, N\}$ of cardinality m with its elements chosen uniformly at random. This is one of the main differences to the previous discussed linear reconstruction methods. For the linear methods we restrict ourselves to the first N measurements instead of choosing the most beneficial ones. Owing to a result by Candès & Plan (16) and Adcock & Hansen (5), if we have access to the subset of noisy measurements $y = P_\Omega \Psi f + e$ then f can be recovered from y exactly (up to the noise level) with probability at least $1 - \epsilon$ if

$$m \gtrsim \mu(U) \cdot N \cdot s \cdot (1 + \log(1/\epsilon)) \cdot \log(N), \quad (55)$$

where $P_\Omega \in \{0, 1\}^{N \times N}$ is the diagonal projection matrix with the j^{th} entry 1 if

$j \in \Omega$ and 0 otherwise, and the notation $a \gtrsim b$ means that $a \geq Cb$ where $C > 0$ is some constant independent of a and b . Then, f is recovered by solving

$$\min_{z \in \mathbb{C}^N} \|z\|_1 \quad \text{subject to} \quad \|y - P_\Omega U z\| \leq \eta. \quad (56)$$

where η is chosen according to the noise level, i.e. $\|e\| \leq \eta$. The key estimate [Eq. 55] shows that the number of measurements m required is, up to a log factor, on the order of the sparsity s , provided the coherence $\mu(U) = \mathcal{O}(1/N)$. This is the case, for example, when U is the DFT, which was studied in some of the first CS papers (19).

The main reason why we want to consider infinite dimensional CS is threefold. First, our signal is defined in a continuous setting in $L^2([0, 1]^d)$ instead of \mathbb{R}^n , hence it is sensible to adapt the reconstruction problem accordingly. Second, the discrete setting leads to the measurement mismatch and wavelet crime. The measurement mismatch comes from the fact that the discrete Hadamard transform leads to an approximation of the signal by step function, which always results in an additional approximation error for our method. The wavelet crime describes the error which results from assuming that the discrete inverse of the wavelet coefficients leads to the point evaluations of the signal. This is an approximation which uses the fact that for high order scaling function the support is nearly point-wise. However, as this is only an approximation we also add up this error. Finally, we want to analyse the change of basis matrix. For the analysis it is easier to consider the inner products of the wavelets and the Walsh function than to work with the discrete matrices. This allows us to develop a rich analysis.

5.2 Types of compressed sensing problems

CS problems can be roughly divided into two types. **Type I** are problems where the physical device imposes the sampling operator, but allows some limited freedom to design the sampling strategy. This category is vast, with examples including Magnetic Resonance Imaging (MRI), Electron Microscopy (EM), Computerized Tomography, Seismic Tomography and Radio Interferometry. **Type II** are problems where the sensing mechanism offers freedom to design both the sampling operator and the strategy. Examples include Fluorescence Microscopy (FM) and Compressive Imaging (CI) (e.g. single pixel and lensless cameras). In these two examples, many practical set-ups still impose some restrictions regarding the sampling operator, e.g. measurements must typically be binary.

In a simplified view, traditional CS assumes three main principles: *sparsity* (there are s important coefficients in the vector to be recovered, however, the location is arbitrary), *incoherence* (the values in the measurements matrix should be uniformly spread out) and *sampling is performed with some degree of randomness*.

In many Type I practical problems, some of the above principles as introduced in the traditional CS model are lacking. For example, many Type I problems are coherent due to the physics of the underlying sensing mechanism. However, CS was used successfully in such problems, though with very different sampling

techniques than uniform random sub-sampling. For Type II problems the traditional CS framework is applicable, e.g. in compressive imaging or fluorescence microscopy one can use random Bernoulli matrices. However, as we shall see, the use of complete randomness does not allow one to exploit the structure of the signal during the sampling procedure; it can still be taken into account after sampling (during recovery) but not as efficiently.

5.3 Taking structure and infinite dimensionality into account

The problem considered in this paper is a Type II problem, and there are several ways one can choose the sampling. However, the finite-dimensional set-up in §5.1 must be extended in order to address Problem 1.1. The first question one may ask oneself is: how should one carry out the sub-sampling? Indeed, would choosing some $M \in \mathbb{N}$ and then, as suggested in §5.1, choosing uniformly at random m indices from $\{1, \dots, M\}$ be a reasonable idea?

In order to answer this question it may be of interest to investigate the relationship between the sampling space of Walsh functions and the reconstruction space of, for example, wavelets. Consider the infinite change of basis matrix

$$U = \begin{pmatrix} \langle \varphi_1, \zeta_1 \rangle & \langle \varphi_2, \zeta_1 \rangle & \cdots \\ \langle \varphi_1, \zeta_2 \rangle & \langle \varphi_2, \zeta_2 \rangle & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}, \quad (57)$$

where the φ_j s are the Haar wavelets and the ζ_j s are the Walsh functions. In Figure 3 the absolute values of the matrix elements (the 1024×1024 finite section of U) are displayed in a grey scale for the two dimensional Walsh and Haar functions. It is evident from the figure that there is a very clear block-diagonal structure. Moreover, in Figure 1 we can see the stable sampling rate and the absolute values of U for Daubechies 2 and Haar wavelets given sampling with Walsh functions in the one dimensional case. It is clear that the matrix U obeys a lot of structure. Indeed, for the Walsh-Haar case we observe perfect block diagonality, which leads to a slope of the stable sampling rate of 1. These observations can be made rigorous in the following results.

Proposition 5.1. *Let $\psi = \chi_{[0,1/2]} - \chi_{(1/2,1]}$ be the Haar wavelet, where χ denotes the characteristic function. Then, we have that*

$$|\langle \psi_{R,j}, \text{Wal}(n, \cdot) \rangle| = \begin{cases} 2^{-R/2} & 2^R \leq n < 2^{R+1}, 0 \leq j \leq 2^R - 1 \\ 0 & \text{otherwise,} \end{cases} \quad (58)$$

where we recall the wavelet notation with subscripts from [Eq. 27].

Note that the Haar wavelet is only defined on $[0, 1]$ and hence does not need to be boundary corrected in contrast to higher order Daubechies wavelets.

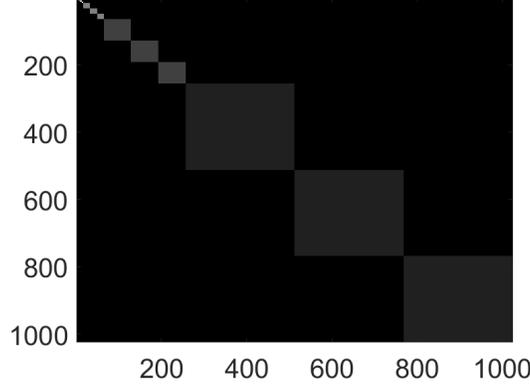


Figure 3: Colour plot of the absolute values of the first 1024×1024 matrix elements of the infinite matrix U in [Eq. 57] for two dimensional Haar wavelets and Walsh functions. White colour denotes 1 whereas black illustrates 0.

Proposition 5.2 ((60)). *Let $\phi = \chi_{[0,1]}$ be the Haar scaling function. Then, we have that the Walsh transform obeys the following block and decay structure*

$$|\langle \phi_{R,j}, \text{Wal}(n, \cdot) \rangle| = \begin{cases} 2^{-R/2} & n < 2^R, 0 \leq j \leq 2^R - 1 \\ 0 & \text{otherwise,} \end{cases} \quad (59)$$

where we recall the wavelet notation with subscripts from [Eq. 27].

These results can be combined into a theorem describing the situation in two dimensions. Indeed, recall the standard two-dimensional set-up for the Haar wavelet:

$$\psi_{R,j_1,j_2,l}(x_1, x_2) = \begin{cases} \phi_{R,j_1}(x_1)\psi_{R,j_2}(x_2) & l = 1 \\ \psi_{R,j_1}(x_1)\phi_{R,j_2}(x_2) & l = 2 \\ \psi_{R,j_1}(x_1)\psi_{R,j_2}(x_2) & l = 3. \end{cases} \quad (60)$$

We then get the theoretical justification for the observed structure in the change of basis matrix in Figure 3.

Theorem 5.3 ((60)). *Let $\psi_{R,j_1,j_2,l}$ be the Haar wavelet defined as in [Eq. 60]. Then, the Walsh transform has the following property. For $0 \leq j_1, j_2 \leq 2^R - 1$,*

$$|\langle \psi_{R,j_1,j_2,1}, \text{Wal}(n_1, n_2, \cdot, \cdot) \rangle| = \begin{cases} 2^{-R} & n_1 \leq 2^R, 2^R \leq n_2 < 2^{R+1} \\ 0 & \text{otherwise,} \end{cases} \quad (61)$$

$$|\langle \psi_{R,j_1,j_2,2}, \text{Wal}(n_1, n_2, \cdot, \cdot) \rangle| = \begin{cases} 2^{-R} & 2^R \leq n_1 < 2^{R+1}, n_2 \leq 2^R \\ 0 & \text{otherwise} \end{cases} \quad (62)$$

and for the third version

$$|\langle \psi_{R,j_1,j_2,3}, \text{Wal}(n_1, n_2, \cdot, \cdot) \rangle| = \begin{cases} 2^{-R} & 2^R \leq n_1 < 2^{R+1}, 2^R \leq n_2 < 2^{R+1} \\ 0 & \text{otherwise.} \end{cases} \quad (63)$$

Theorem 5.3 describes the block-diagonal structure visualised in Figure 3. These findings suggest that also for the compressed sensing approach it is sensible to take additional structure that can be observed for wavelets and Walsh functions into account. This motivated the introduction of an extended framework for CS (3) by generalising the traditional CS principles of incoherence and sparsity into *asymptotic incoherence* and *asymptotic sparsity*, proposing a matched sampling procedure called *multilevel sampling*. In Section 5.4 we shall also discuss structured sampling in contrast with structured recovery, and implications regarding sampling operators in the context of measurements with binary functions.

5.3.1 Multilevel sampling

High coherence in the first few rows of U means that important information about the signal to be recovered is likely to be contained in the corresponding measurements, and thus we should fully sample these rows. Once outside this region, as coherence starts decreasing, we can sub-sample gradually. This is also the wisdom behind the various variable density sampling strategies, which were first introduced in (49).

Definition 5.4 (Multilevel sampling). *Let $r \in \mathbb{N}$, $\mathbf{M} = (M_0, \dots, M_r) \in \mathbb{N}^{r+1}$ with $1 \leq M_1 < \dots < M_r$, $\mathbf{m} = (m_1, \dots, m_r) \in \mathbb{N}^r$, with $m_k \leq M_k - M_{k-1}$, $k = 1, \dots, r$, and that $\Omega_k \subseteq \{M_{k-1} + 1, \dots, M_k\}$, $|\Omega_k| = m_k$, are chosen uniformly at random, where $M_0 = 0$. We refer to the set $\Omega = \Omega_{\mathbf{M}, \mathbf{m}} = \bigcup_{k=1}^r \Omega_k$ as an (\mathbf{M}, \mathbf{m}) -multilevel sampling scheme (using r levels).*

Briefly, for a vector x , the sampling amount m_k needed in each sampling band Ω_k is determined by the sparsity of x in the corresponding sparsity band Δ_k and the asymptotic coherence $\mu(P_{M_k}^\perp U)$.

5.3.2 Asymptotic sparsity

Let us consider a wavelet basis indexed by one variable in the canonical way according to the different scales $\{\varphi_n\}_{n \in \mathbb{N}}$. There is a natural decomposition of \mathbb{N} into finite subsets according to the wavelet scales, $\mathbb{N} = \bigcup_{k \in \mathbb{N}} \{N_{k-1} + 1, \dots, N_k\}$, where $0 = N_0 < N_1 < N_2 < \dots$ and $\{N_{k-1} + 1, \dots, N_k\}$ is the set of indices corresponding to the k^{th} scale. For the boundary wavelets we have $N_i = 2^{d(J_0 + i)}$. Let $x \in l^2(\mathbb{N})$ be the coefficients of a function f in this basis, $\epsilon \in (0, 1]$ and define the global sparsity, s , and the sparsity at the k^{th} level, s_k as follows:

$$s = s(\epsilon) = \min \left\{ n : \left\| \sum_{i \in \mathcal{N}_n} x_i \varphi_i \right\| \geq \epsilon \left\| \sum_{j=1}^{\infty} x_j \varphi_j \right\| \right\}, \quad (64)$$

$$s_k = s_k(\epsilon) = |\mathcal{N}_{s(\epsilon)} \cap \{N_{k-1} + 1, \dots, N_k\}|,$$

where \mathcal{N}_n is the set of indices of the largest n coefficients in absolute value and $|\cdot|$ is the set cardinality. Figure 4 shows that besides being sparse, images have more structure, namely *asymptotic sparsity*, i.e. the relative per-level sparsity

$$s_k/(N_k - N_{k-1}) \longrightarrow 0 \quad (65)$$

rapidly as $k \rightarrow \infty$ for any fixed $\epsilon \in (0, 1]$. In particular, images are far sparser at fine scales (large k) than at coarse scales (small k). This also holds for other function systems, e.g. curvelets (15), contourlets (23) or shearlets (21). Note that asymptotic sparsity is a rather different, and much more general structure than the connected tree structure of wavelet coefficients (53). [Eq. 64] and [Eq. 65] do not assume such a tree structure, but only different local sparsities s_k at different levels.

Given the structure of function systems such as wavelets and their generalisations, we instead consider the notion of sparsity in levels:

Definition 5.5 (Sparsity in levels). *Let x be an element of either \mathbb{C}^N or $l^2(\mathbb{N})$. For $r \in \mathbb{N}$ let $\mathbf{N} = (N_0, \dots, N_r) \in \mathbb{N}^r$ and $\mathbf{s} = (s_1, \dots, s_r) \in \mathbb{N}^r$, with $s_k \leq N_k - N_{k-1}$, $k = 1, \dots, r$, where $N_0 = 0$. We say that x is (\mathbf{s}, \mathbf{N}) -sparse if, for each $k = 1, \dots, r$ we have $|\Delta_k| \leq s_k$, where*

$$\Delta_k := \text{supp}(x) \cap \{N_{k-1} + 1, \dots, N_k\}.$$

We write $\Sigma_{\mathbf{s}, \mathbf{N}}$ for the set of (\mathbf{s}, \mathbf{N}) -sparse vectors.

5.3.3 Asymptotic incoherence.

In contrast with random matrices, such as Gaussian or Bernoulli, many sampling and sparsifying operators typically found in practice yield fully coherent problems, such as the Hadamard with wavelets case discussed earlier. Indeed, Figure 1 shows the absolute values of the entries of the matrix U with Haar and Daubechies 2 wavelets. Although there are large values of U in both cases (since U is coherent as per [Eq. 54]), these are isolated to a leading submatrix. Values get asymptotically smaller once we move away from this region. This motivates the following definition.

Definition 5.6 (Asymptotic incoherence). *Let $\{U_N\}$ be a sequence of isometries with $U_N \in \mathbb{C}^{N \times N}$. Then $\{U_N\}$ is asymptotically incoherent if both $\mu(P_K^\perp U_N)$, $\mu(U_N P_K^\perp) \rightarrow 0$ when $K \rightarrow \infty$ with $N/K = c$, for all $c \geq 1$. Conversely, if $U \in \mathcal{B}(l^2(\mathbb{N}))$, (i.e. U belongs to the space of bounded operators on $l^2(\mathbb{N})$) then we say that U is asymptotically incoherent if $\mu(P_K^\perp U)$, $\mu(U P_K^\perp) \rightarrow 0$ when $K \rightarrow \infty$.*

In brief, U is asymptotically incoherent if the coherences of the matrices formed by removing either the first K rows or columns of U are small. As Figure 1 shows, the change of basis matrix U in [Eq. 57] when considering Walsh functions and Haar wavelets is clearly asymptotically incoherent. However, asymptotic incoherence may not be refined enough to capture the finesses of the fine structures in the change of basis matrix U . In particular, we need the concept of local coherence, which is much more of a scalpel that allows for precise recovery guarantees.

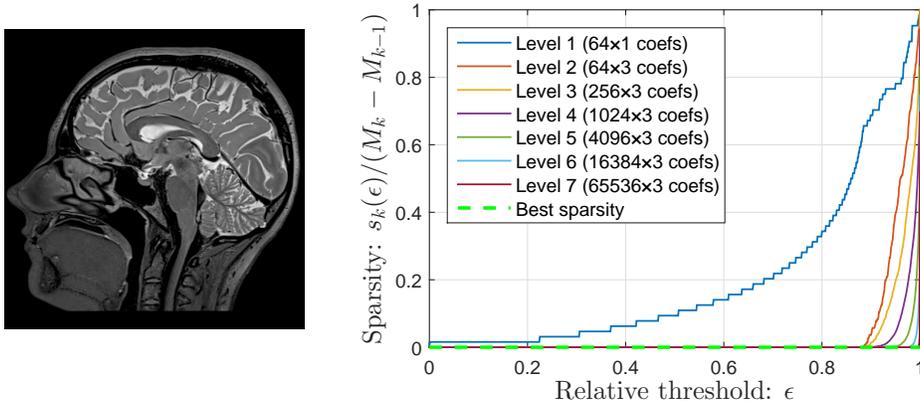


Figure 4: Sparsity of Daubechies-4 coefficients of an MRI image, courtesy of Siemens AG. Levels correspond to wavelet scales and $s_k(\epsilon)$ is given by [Eq. 64]. Each curve shows the relative sparsity at level k as a function of ϵ . Their decreasing nature for increasing k confirms asymptotic sparsity [Eq. 65].

Definition 5.7 (Local coherence). *Let U be an isometry of either \mathbb{C}^N or $l^2(\mathbb{N})$. If $\mathbf{M} = (M_0, \dots, M_r) \in \mathbb{N}^{r+1}$ and $\mathbf{N} = (N_0, \dots, N_r) \in \mathbb{N}^{r+1}$ with $1 \leq M_1 < \dots < M_r$ and $1 \leq N_1 < \dots < N_r$ the (k, l) th local coherence of U with respect to \mathbf{M} and \mathbf{N} is given by*

$$\mu_{\mathbf{M}, \mathbf{N}}(k, l) = \sqrt{\mu(P_{M_k}^{M_{k-1}} U P_{N_l}^{N_{l-1}}) \cdot \mu(P_{M_k}^{M_{k-1}} U)}, \quad k, l = 1, \dots, r,$$

where $N_0 = M_0 = 0$ and P_b^a denotes the projection matrix corresponding to indices $\{a+1, \dots, b\}$. In the case where $U \in \mathcal{B}(l^2(\mathbb{N}))$, we also define

$$\mu_{\mathbf{M}, \mathbf{N}}(k, \infty) = \sqrt{\mu(P_{M_k}^{M_{k-1}} U P_{N_{r-1}}^1) \cdot \mu(P_{M_k}^{M_{k-1}} U)}, \quad k = 1, \dots, r.$$

By estimating the local coherence of U in [Eq. 57] for arbitrary wavelets we can obtain recovery guarantees for infinite-dimensional compressed sensing. These are presented in the next section.

5.3.4 Recovery Guarantees

We are stating the recovery guarantees for Walsh functions and wavelets. For this we consider the ordering of the levels of the sampling and the reconstruction space. We get

$$\mathbf{N} = (N_0^d, N_1^d, \dots, N_r^d) = (0, 2^{d(J_0+1)}, 2^{d(J_0+2)}, \dots, 2^{d(J_0+r)}) \quad (66)$$

and

$$\mathbf{M} = (M_0^d, M_1^d, \dots, M_{r-1}^d, M_r^d) = (0, 2^{d(J_0+1)}, 2^{d(J_0+2)}, \dots, 2^{d(J_0+r-1)}, 2^{d(J_0+r+q)}). \quad (67)$$

Theorem 5.8 ((59)). *Let the notation be as before, i.e. let the sampling space be given by Walsh functions and the reconstruction space spanned by boundary corrected Daubechies wavelets. Additionally, let $\epsilon > 0$ and $\Omega = \Omega_{M,m}$ be a multilevel sampling scheme such that the following holds:*

1. Let $M = M_r$, $K = \max_{k=1,\dots,r} \left\{ \frac{M_k - M_{k-1}}{m_k} \right\}$, $N = N_r$, $s = s_1 + \dots + s_r$ such that

$$M \geq CN^2 \cdot \log_2(4NK\sqrt{s}). \quad (68)$$

2. For each $k = 1, \dots, r$,

$$m_k \geq C \log(\epsilon^{-1}) \log(K^2 s M) \cdot \frac{M_k - M_{k-1}}{M_{k-1}} \cdot \left(\sum_{l=1}^r 2^{-d|k-l|} s_l \right). \quad (69)$$

Then with probability exceeding $1 - \epsilon\epsilon$, any minimizer $\xi \in \ell^1(\mathbb{N})$ of [Eq. 56] satisfies

$$\|\xi - x\| \leq C \cdot \left(\delta\sqrt{K}(1 + L\sqrt{s}) + \sigma_{s,N}(f) \right), \quad (70)$$

for some constant C , where $L = C \cdot \left(1 + \frac{\sqrt{\log_2(6\epsilon^{-1})}}{\log_2(4KN\sqrt{s})} \right)$. If $m_k = M_k - M_{k-1}$ for $1 \leq k \leq r$ then this holds with probability 1.

We see that the impact of the off block parts is exponentially decreasing. This allows us to exploit the asymptotic sparsity and reduce heavily the number of samples.

Remark. *In equation [Eq. 68] we see that the relation between the number of samples and coefficients is squared with an additional log factor. This quadratic term is likely to be an artefact of the proof and not sharp. In (7) it was shown that for the Fourier wavelet case this can be reduced to a linear relation, if the wavelet decays fast under the Fourier transform, i.e. if it is smooth. Unfortunately, there is no direct relation between the smoothness of the wavelet and the decay under the Walsh transform. Therefore, this results are not directly transferable and hence still open research.*

5.3.5 Relation to previous work

It has long been known that wavelet coefficients possess additional structure beyond sparsity. In the CS context, this is the basis for structured recovery algorithms, such as model-based CS (11), Bayesian CS (42) and TurboAMP (57). We discuss these later on. These algorithms exploit the connected tree structure of wavelet coefficients based on the ‘‘persistence across scales’’ phenomenon (53). Asymptotic sparsity assumes only asymptotic decrease of the local sparsities in the individual levels to zero. Asymptotic sparsity is more general, as the levels chosen need not correspond to the $*$ -let levels, where $*$ -lets denote the representation systems that are inspired by wavelets, like shearlets and curvelets. Additionally, it

makes no assumption about dependencies between coefficients such as a connected tree.

A number of different characterizations of non-flat coherence patterns have been introduced in CS previously (25, 33, 61, 62). What differentiates asymptotic incoherence is that it allows one to capture (near) block diagonal structure inherent to ℓ^1 -let bases, by defining a vector of local coherence values for blocks of the coherence matrix, and specifically incorporates the asymptotic decrease of these values and the boundaries of each block. As we shall show, this is key to the practical recovery performance.

The idea of sampling the low-order coefficients of an image differently goes back to the early days of CS. In particular, Donoho considers a two-level approach for recovering wavelet coefficients in his seminal paper (24), based on acquiring the coarse scale coefficients directly. This was later extended by Tsaig & Donoho to so-called 'multi-scale CS' in (63), where distinct sub-bands were sensed separately. See also the works by Candès & Romberg (18) and Romberg (56). We note that the sampling schemes of (24, 56), and more recently, the "half-half" scheme of (58) proposed for the application of CS to fluorescence microscopy are examples of two-level sampling strategies within our general framework and were analysed in detail in (3). Our multilevel sampling extends these ideas as part of a formal framework for CS.

5.4 Points of discussion regarding structure

Structured sampling and Structured Recovery. In this work we exploit the sparsity structure at the sampling stage, by sampling asymptotically incoherent matrices, and use standard ℓ^1 minimization algorithms. Alternatively, sparsity structure can be exploited by using universal sampling matrices (e.g. random Gaussian/Bernoulli) and modified recovery algorithms which exploit the structure at the recovery stage.

Structure or Universality. The universality property of random sensing matrices (e.g. Gaussian, Bernoulli), explained later on, is a reason for their popularity in traditional CS. But is universality desirable when the signal sparsity is structured? Should one use universal matrices when there is freedom to choose the sampling operator, i.e. in Type II problems? Random matrices are largely inapplicable in Type I problems where the sampling operator yields coherent operators.

Storage and speed. Random matrices, while popular, require either large storage or are otherwise slow to generate (from a pseudo-random generator point of view), which yields slow recovery and limits the maximum signal size, which adversely affects computations. However, there exist ways to perform CS using fast transforms that emulate the usage of random matrices. Nevertheless, is addressing the speed/space problems via fast transforms or non-random matrices sufficient?

5.4.1 Structured sampling and structured recovery

The asymptotic CS framework takes into account the sparsity structure during the sampling stage via multilevel sampling of non-universal sensing matrices. Sparsity



Original



Random Bernoulli to db4 — ℓ^1 Err 16.0%



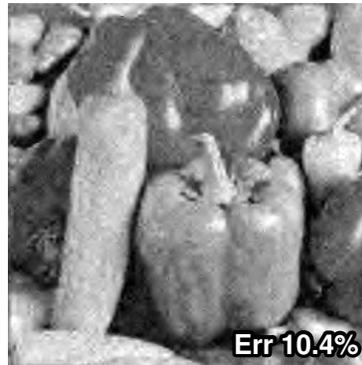
Rnd. Bernoulli to db4 — ModelCS Err 21.2%



Rnd. Bernoulli to db4 — TurboAMP Err 10.7%



Random Bernoulli to db4 — BCS Err 10.6%



Rnd. Bernoulli to db4 — Weighted ℓ^1 Err 10.4%



Multilevel Hadamard to db4 — ℓ^1 Err 7.1%



Multilevel Had. to Curvelets — ℓ^1 Err 6.3%

Figure 5: Compressive Imaging example. 12.5% sub-sampling at 256x256.

structure can also be taken into account in the recovery algorithm. A well-known example of such an approach is model-based CS (11), which assumes the signal is piecewise smooth and exploits the connected tree structure (persistence across scales) of wavelet coefficients (53) to reduce the search space of the matching pursuit algorithm (54). The same tree structure is exploited by the class of message passing and approximate message passing algorithms (12, 27). This can be coupled with hidden Markov trees to model the wavelet structure, such as in the Bayesian CS (42) and TurboAMP (57) algorithms. Another approach is to assign non-uniform weights to the sparsity coefficients (45), to favour the important coefficients during ℓ^1 recovery by assuming some typical decay rate of the coefficients. Another approach assumes the signal (not its representation in a sparsity basis) is sparse and random, and shows promising theoretical results when using spatially coupled matrices (65, 46, 26), yet it is unclear how a practical set-up can be realised where signals are sparse in a transform domain.

The main difference is that the former approach, i.e. multilevel sampling of asymptotically incoherent matrices, incorporates sparsity structure in the sampling strategy and can use standard ℓ^1 minimization algorithms, whereas the latter approaches exploit structure by modifying the recovery algorithm and use universal sampling operators which yield uniform incoherence, e.g. random Gaussian or Bernoulli.

By using universal operators and assuming a sparsity basis, *structured recovery* is typically restricted to Type II problems, where the sensing operator can be designed (see also the remark below), and is limited by the choice of the representation system, whose structure is exploited by the modified algorithm.

Structured sampling is flexibility with regards to the representation system and is applicable in both Type I and Type II problems.

To compare performance, we ran a set of simulations of Compressive Imaging (34, 44), which is a Type II problem, and has utilized universal sensing matrices. Binary measurements y are taken, typically using a $\{-1, 1\}^{N \times N}$ sensing matrix. Any matrix with only two values fits this set-up, such as Hadamard, random Bernoulli, Sum-To-One (37), hence we can directly compare the two approaches. Figure 5 shows a representative example from our set of simulations. One can notice that asymptotic incoherence combined with multilevel sampling of highly non-universal sensing matrices (e.g. Hadamard, Fourier) allows structured sparsity to be better exploited than universal sensing matrices, even when structure is accounted for in the recovery algorithm. The figure also shows the added benefit of being able to use a better sparsifying system, in this case curvelets.

Is it possible to combine the two approaches to leverage further gains? The structured recovery algorithms we have encountered expect the sampling operator to be incoherent with the recovery basis. Replacing those with asymptotically incoherent operators such as Hadamard or Fourier resulted in poorer performance, sometimes failing to produce a result, which isn't totally surprising given that the aforementioned structured recovery algorithms make certain assumptions about the sampling operator. Nevertheless, the successful combination of the two approaches is a promising line of investigation and is the subject of ongoing research.

5.4.2 Structure and universality: Is universality desirable?

Universality is a reason for the popularity in traditional CS of random sensing matrices, e.g. Gaussian or Bernoulli. A random matrix $A \in \mathbb{C}^{m \times N}$ is universal if for every isometry $\Psi \in \mathbb{C}^{N \times N}$, the matrix $A\Psi$ satisfies the restricted isometry property (19) with high probability. For images, a common choice is $\Psi = \Psi_{\text{dwt}}^*$, the inverse wavelet transform. Universality is a key feature when the signal is sparse but possesses no further structure.

But is universality desirable in a sensing matrix when the signal is structured? First, random matrices are applicable mostly in Type II problems, where there is freedom to design the sampling operator. Hence also universal matrices are possible from a practical perspective. But should one use universal matrices there? We argue that universal matrices offer little room to exploit extra structure the signal may have, even in Type II problems.

Typical signals in practice exhibit far more structure than sparsity alone: their sparsity is asymptotic in some basis. Thus, an alternative is to use a non-universal sensing matrix, such as Hadamard, Φ_{Had} . As previously discussed and shown in Figure 1 and 3, $U = \Phi_{\text{Had}} \Psi_{\text{dwt}}^*$ is completely coherent with all wavelets yet asymptotically incoherent, and thus perfectly suitable for a multilevel sampling scheme which can exploit the inherent asymptotic sparsity. This is precisely what we see in Figure 5: a multilevel sampled Hadamard matrix can markedly outperform universal matrices in Type II problems. In Type I problems, many imposed sensing operators are non-universal and asymptotically incoherent with popular sparsity bases, and thus exploitable using multilevel sampling.

The reasons for the superior results are rooted in the incoherence structure. Universal and close to universal sensing matrices typically provide a relatively low and flat coherence pattern. This allows sparsity to be exploited by sampling uniformly at random but, by definition, these matrices cannot exploit the distinct asymptotic sparsity structure when using a typical (ℓ^1 minimization) CS reconstruction.

In contrast, when the sensing matrix provides a coherence pattern that aligns with the signal sparsity pattern, one can fruitfully exploit such structure. A multilevel sampling scheme is likely to give superior results by sampling more in the coherent regions, where the signal is also typically less sparse. The optimum sampling strategy is signal dependent. However, real-world signals, particularly images, share a fairly common structure in the wavelet domain and also in wavelet inspired representation systems. This structure allows to design variable density sampling strategies. An added benefit when this alignment exists, is that the sampling procedure allows for tailoring of the sampling pattern to target application-specific features rather than an all-round approach, e.g. recovering contours better, trading overall quality.

5.4.3 Storage and speed

Random matrices require (large) storage and lack fast transforms. This limits the maximum signal resolution and yields slow recovery. For example, a 1024×1024

recovery with 25% sub-sampling of a random Gaussian matrix would require 2 Terabytes of free memory and $\mathcal{O}(10^{12})$ time complexity, making it impractical. The storage issue could be addressed naively, by storing only the initial seed and generating the matrix on the fly, but that makes the process orders of magnitude slower.

Both the storage and speed issues were in fact addressed to various extents, e.g. pseudo-random column permutations of the columns of orthogonal matrices such as (block) Hadamard or Fourier (17, 35), Kronecker products of random matrix stencils (28), or even fully orthogonal matrices such as the Sum-To-One (STOne) matrix¹ (37) which allows for a fast $\mathcal{O}(N \log N)$ transform. All these solutions in the CS context yield similar statistics to a random matrix: they become universal sampling operators.

Another solution to the storage and speed problem is to instead use structured matrices like Hadamard, DCT or DFT. These have fast transforms but also provide asymptotic incoherence with most sparsity bases, thus a multilevel sub-sampling scheme can be used. This yields significantly better CS recovery when compared to universal matrices, as witnessed previously, and it is also applicable to Type I problems, which impose the sensing operator.

In conclusion, the sensing matrix should contain additional structure besides simply being non-random and/or orthogonal in order to provide asymptotic incoherence. Typically, sensing and sparsifying matrices that are discrete versions of integral transforms, e.g. Fourier, wavelets etc., will provide asymptotic incoherence, but other orthogonal and structured matrices like Hadamard will do so too.

6 Conclusion

This work concerns the recovery of signals from measurements with binary functions, that is the measurements are inner products with functions that map to $\{0, 1\}$, in the context of linear recovery using either PBDW or generalised sampling, and non-linear recovery using infinite-dimensional compressed sensing. We considered the use of Walsh functions in the sampling domain and wavelets in the reconstruction domain. In the linear case, we showed that the methods rely on knowing the stable sampling rate, and we established its linearity and that it is sharp. Furthermore, we showed that generalised sampling keeps the solution in the reconstruction space which allows for improvements over PBDW in the case of highly sparse functions. In the non-linear case, we derived recovery guarantees and discussed the advantages of using Walsh functions (via the Hadamard transform) over incoherent sampling.

¹The STOne matrix is an orthogonal matrix that provides universality, like random matrices do. However, it was invented for many other purposes. It has a fast $\mathcal{O}(N \log N)$ transform and allows multi-scale image recovery from compressive measurements: low-resolution previews can be quickly generated by applying the fast transform on the measurements directly, and high resolution recovery is possible from the same measurements via CS solvers. In addition, it allows efficient recovery of compressive videos when sampling in a semi-random manner.

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, C. Poon, and B. Roman. Breaking the coherence barrier: A new theory for compressed sensing. *Forum Math. Sigma*, 5, 2017.
- [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] R. Baraniuk, V. Cevher, M. Duarte, and C. Hedge. Model-based compressive sensing. *IEEE T Inf Th*, 56(4), 2010.
- [12] D. Baron, S. Sarvotham, and R. Baraniuk. Bayesian compressive sensing via belief propagation. *IEEE T Sig Proc*, 58(1), 2010.
- [13] 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.

- [14] 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.
- [15] E. Candès and D. Donoho. Recovering edges in ill-posed inverse problems: optimality of curvelet frames. *Ann. Statist.*, 30(3), 2002.
- [16] E. Candès and Y. Plan. A probabilistic and RIPless theory of compressed sensing. *IEEE T Inf Th*, 57(11), 2011.
- [17] E. Candès and J. Romberg. Robust signal recovery from incomplete observations. In *IEEE Intl Conf Image Proc*, 2006.
- [18] E. Candès and J. Romberg. Sparsity and incoherence in compressive sampling. *Inverse Problems*, 23(3), 2007.
- [19] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE T Inf Th*, 52(2), 2006.
- [20] A. Cohen, I. Daubechies, and P. Vial. Wavelets on the interval and fast wavelet transforms. *Comput. Harmon. Anal.*, 1(1):54–81, 1993.
- [21] 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.*, 6(2), 2008.
- [22] R. DeVore, G. Petrova, and P. Wojtaszczyk. Data assimilation and sampling in banach spaces. *Calcolo*, 54(3):963–1007, 2017.
- [23] M. Do and M. Vetterli. The contourlet transform: An efficient directional multiresolution image representation. *IEEE T Image Proc*, 14(12), 2005.
- [24] D. Donoho. Compressed sensing. *IEEE T Inf Th*, 52(4), 2006.
- [25] D. Donoho and M. Elad. Optimally sparse representation in general (non-orthogonal) dictionaries via ℓ_1 minimization. *Proc Natl Acad Sci USA*, 100, 2003.
- [26] D. Donoho, A. Javanmard, and A. Montanari. Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing. *IEEE T Inf Th*, 59(11), 2013.
- [27] D. Donoho, A. Maleki, and A. Montanari. Message-passing algorithms for compressed sensing. *Proc Natl Acad Sci USA*, 106(45), 2009.
- [28] M. Duarte and R. Baraniuk. Kronecker compressive sensing. *IEEE T Image Proc*, 21(2), 2012.
- [29] T. Dvorkind and Y. C. Eldar. Robust and consistent sampling. *IEEE Signal Process. Letters*, 16(9):739–742, 2009.

- [30] Y. C. Eldar. Sampling with arbitrary sampling and reconstruction spaces and oblique dual frame vectors. *J. Fourier Anal. Appl.*, 9(1):77–96, 2003.
- [31] Y. C. Eldar. Sampling without input constraints: Consistent reconstruction in arbitrary spaces. *Sampling, Wavelets and Tomography*, 2003.
- [32] 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.
- [33] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. Birkhäuser, 2013.
- [34] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. Springer Science+Business Media, New York, 2013.
- [35] L. Gan, T. Do, and T. Tran. Fast compressive imaging using scrambled hadamard ensemble. In *Proc Eur Signal Proc Conf*, 2008.
- [36] E. Gauss. *Walsh Funktionen für Ingenieure und Naturwissenschaftler*. Springer Fachmedien, Wiesbaden, 1994.
- [37] T. Goldstein, L. Xu, K. Kelly, and R. Baraniuk. The stone transform: Multi-resolution image enhancement and real-time compressive video. *arXiv:1311.3405*, 2013.
- [38] 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.
- [39] A. Hansen. On the approximation of spectra of linear operators on hilbert spaces. *J. Funct. Anal.*, 8(254):2092–2126, 2008.
- [40] A. Hansen and L. Thesing. On the stable sampling rate for binary measurements and wavelet reconstruction. *Applied and Computational Harmonic Analysis*, 2018.
- [41] 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.
- [42] L. He and L. Carin. Exploiting structure in wavelet-based Bayesian compressive sensing. *IEEE T Sig Proc*, 57(9), 2009.
- [43] T. Hrycak and K. Gröchenig. Pseudospectral fourier reconstruction with the modified inverse polynomial reconstruction method. *J. Comput. Phys.*, 229(3):933–946, 2010.
- [44] G. Huang, H. Jiang, K. Matthews, and P. Wilford. Lensless imaging by compressive sensing. In *IEEE Intl Conf Image Proc.*, 2013.

- [45] M. Khaejehnejad, W. Xu, A. Avestimehr, and B. Hassibi. Analyzing weighted ℓ_1 minimization for sparse recovery with nonuniform sparse models. *IEEE T Sig Proc*, 59(5), May 2011.
- [46] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová. Statistical-physics-based reconstruction in compressed sensing. *Phys Rev X*, 2, May 2012.
- [47] G. Kutyniok and W.-Q. Lim. Optimal compressive imaging of fourier data. *SIAM Journal on Imaging Sciences*, 11(1):507–546, 2018.
- [48] M. Lindner. *Infinite matrices and their finite sections: An introduction to the limit operator method*. Birkhäuser Verlag, Basel, 2006.
- [49] M. Lustig, D. Donoho, and J. Pauly. Sparse MRI: the application of compressed sensing for rapid MRI imaging. *Magn. Reson. Imaging*, 58(6), 2007.
- [50] J. Ma. Generalized sampling reconstruction from fourier measurements using compactly supported shearlets. *Appl. Comput. Harmon. Anal.*, 2015.
- [51] 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.
- [52] S. Mallat. *A wavelet tour of signal processing*. Academic Press, San Diego, 1998.
- [53] S. Mallat. *A Wavelet Tour of Signal Processing: The Sparse Way*. Academic Press, 3rd edition, 2009.
- [54] D. Needell and J. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(3), 2009.
- [55] C. Poon. Structure dependent sampling in compressed sensing: theoretical guarantees for tight frames. *Applied and Computational Harmonic Analysis*, 42(3):402–451, 2017.
- [56] J. Romberg. Imaging via compressive sampling. *IEEE Sig Proc Mag*, 25(2), 2008.
- [57] S. Som and P. Schniter. Compressive imaging using approximate message passing and a markov-tree prior. *IEEE T Sig Proc*, 60(7), 2012.
- [58] 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.

- [59] L. Thesing and A. Hansen. Non uniform recovery guarantees for binary measurements and wavelet reconstruction. *to appear*.
- [60] L. Thesing and A. Hansen. Linear reconstructions and the analysis of the stable sampling rate. *Sampling theory in image processing*, 17(1):103–126, 2018.
- [61] J. Tropp. Greed is good: algorithmic results for sparse approximation. *IEEE T Inf Th*, 50(10), 2004.
- [62] J. Tropp. Just relax: convex programming methods for identifying sparse signals in noise. *IEEE T Inf Th*, 52(3), 2006.
- [63] Y. Tsaig and D. Donoho. Extensions of compressed sensing. *Signal Process.*, 86(3), 2006.
- [64] M. Unser and J. Zerubia. A generalized sampling theory without band-limiting constraints. *IEEE Trans. Circuits Syst. II.*, 45(8):959–969, 1998.
- [65] Y. Wu and S. Verdu. Rényi information dimension: Fundamental limits of almost lossless analog compression. *IEEE T Inf Th*, 56(8), 2010.