

Linear reconstructions and the analysis of the stable sampling rate

Laura Terhaar

Department of Applied Mathematics and Theoretical Physics,
University of Cambridge, Wilberforce Road, Cambridge
lt420@cam.ac.uk

Anders Hansen

Department of Applied Mathematics and Theoretical Physics,
University of Cambridge, Wilberforce Road, Cambridge
ach70@cam.ac.uk

Abstract. The theory of sampling and reconstruction of data has a wide range of applications and a rich collection of techniques. For many methods, a core problem is to estimate the number of samples needed in order to secure a stable and accurate reconstruction. This can often be controlled by the *Stable Sampling Rate (SSR)*. In this paper we discuss the SSR and how it is crucial for two key linear methods in sampling theory: generalized sampling and the recently developed Parametrized Background Data Weak (PBDW) method. Both of these approaches rely on estimates of the SSR in order to be accurate. In many areas of signal and image processing binary samples are crucial and such samples, which can be modelled by Walsh functions, are the core of our analysis. As we show, the SSR is linear when considering binary sampling with Walsh functions and wavelet reconstruction. Moreover, for certain wavelets it is possible to determine the SSR exactly, allowing sharp estimates for the performance of the methods.

Key words and phrases : Sampling, Wavelets, binary measurements, Generalized sampling, linear reconstruction,

2010 AMS Mathematics Subject Classification : 94A20, 42C10, 42C40 (primary); 65R32, 94A08, 94A12 (secondary)

1. Introduction

Sampling theory is a mainstay in image and signal processing as well as mathematics of information, data science and inverse problems. Since the early results of Shannon [33, 48, 50] many techniques have been developed, and there is now a myriad of methods available. Moreover, the many applications, such as Magnetic Resonance Imaging (MRI) [27, 39], electron tomography [35, 36], lensless cameras, fluorescence microscopy [46, 49], X-ray computed tomography [16, 45], surface scattering [34] as well as parametrised PDEs [9, 12, 19], make the field well connected to different areas of the sciences.

A standard sampling model is as follows. We have an element $f \in \mathcal{H}$, where \mathcal{H} is a separable Hilbert space, and the goal is to reconstruct an approximation to f from a finite number of linear samples of the form $l_i(f)$, $i \in \mathbb{N}$. In particular, given that the l_i s are linear functionals, we measure the scalar product between f and some sampling element $s_i \in \mathcal{H}$, $i \in \mathbb{N}$, i.e. $l_i(f) = \langle f, s_i \rangle$. It is important to note that the l_i s cannot be chosen freely, but are dictated by the modality of the sampling device, say an MRI scanner providing Fourier samples or a fluorescence microscope giving binary measurements modelled by Walsh coefficients.

The different techniques for making the reconstruction can be divided into two main categories: linear and non-linear methods. In the non-linear case (infinite-dimensional) compressed sensing [5, 22, 23] and deep learning [32, 37] are two examples, whereas generalised sampling [1, 2, 4, 6, 7, 20, 21, 29, 31, 40, 51] and the PBDW method [9, 12, 19, 41] are examples of linear methods. Although the SSR is developed mainly for linear methods there is a strong connection to non-linear techniques. For example in infinite-dimensional compressed sensing a key condition is the so-called *balancing property* [5], which is very similar to the SSR. It is not known if a similar concept is needed in deep learning [37], as, so far, there is no mathematical theory describing its performance. However, it would be counterintuitive if deep learning reconstructions could be guaranteed to be stable and accurate without any understanding of how the number of samples relate to the reconstruction technique. Thus, just as the SSR has a connection to compressed sensing we conjecture that there will have to be a similar concept also for deep learning.

To define the SSR we first introduce the sampling space and the reconstruction space. We define the *sampling space* $\mathcal{S} = \overline{\text{span}}\{s_i : i \in \mathbb{N}\} \subset \mathcal{H}$, meaning the closure of the span. In practice, one can only acquire a finite amount of samples, therefore we denote by $\mathcal{S}_M = \text{span}\{s_i : i = 1, \dots, M\}$, the sampling space of the first M elements. Similarly, the reconstruction space denoted by \mathcal{R} is 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 sampling case, one has to restrict to a finite reconstruction space, which is denoted by $\mathcal{R}_N = \text{span}\{r_i : i = 1, \dots, N\}$.

The key ingredient in the definition of the SSR is the *subspace angle* ω between the subspaces \mathcal{R}_N and \mathcal{S}_M . In particular,

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

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

$$\mu(\mathcal{R}_N, \mathcal{S}_M) = 1 / \cos(\omega(\mathcal{R}_N, \mathcal{S}_M)),$$

which, as we will see below, plays a key role in all the error estimates. We can now define the *stable sampling rate*

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

In particular, the SSR determines how many samples M are needed when given N reconstruction vectors in order to bound μ by θ .

There are many applications which can be modelled with Fourier measurements. In particular, the s_j s represent complex exponentials, and the SSR when considering wavelets as reconstruction bases is well known to be linear [3]. Another important group of applications are those which have binary measurements. These measurements can be represented by inner products of the data with Walsh functions. A main application where this arises is fluorescence microscopy [43].

The purpose of this paper is to demonstrate the similarities and differences between the approach of generalized sampling [4] and the PBDW approach based on data assimilation [12]. Moreover, we show how both methods completely rely on the SSR in order to be accurate and we provide sharp results on the SSR when considering Walsh functions and Haar wavelets. This can be done by realising the common structure of the Walsh functions and the Haar wavelets. Although the SSR is linear when considering Walsh samples and wavelet reconstructions, sharp results on the SSR for arbitrary Daubechies wavelets are still open. The difficulty is that the higher order Daubechies wavelets share very little structural similarities with the Walsh functions.

2. Reconstruction Methods

In terms of reconstruction methods there are three different properties that are often desired. The most important are obviously *accuracy* and *stability*. However, *consistency*, meaning that the reconstruction will yield the same samples as the true solution, is also often considered an advantage. Below, we will see how the SSR is crucial for the two former properties.

2.1. Reconstruction and Sampling Space. Throughout the paper $\mathcal{H} = L^2([0, 1]^d)$. Due to the fact that we are dealing with the d -dimensional case, we introduce multi indices to make the notation more readable. Let $j = \{j_1, \dots, j_d\} \in \mathbb{N}^d$, $d \in \mathbb{N}$ be a multi index. A natural number n is in the context with a multi index interpreted as a multi index with the same entry, i.e. $n = \{n, \dots, n\}$. Then we define the addition of two multi indices for $j, r \in \mathbb{N}^d$ by the pointwise addition, i.e. $j + r = \{j_1 + r_1, \dots, j_d + r_d\}$ and the sum $\sum_{j=k}^r x_j := \sum_{j_1=k_1}^{r_1} \dots \sum_{j_d=k_d}^{r_d} x_{j_1, \dots, j_d}$, where $k, r \in \mathbb{N}^d$. The multiplication of a multi index with a real number is understood pointwise, as well as the division by a multi index.

To define the sampling space \mathcal{S}_M we first need to define the Walsh functions. The Walsh functions in higher dimensions can be represented by the tensor product of one dimensional Walsh functions.

Definition 1 (Walsh function [25]). *Let $s = \sum_{i \in \mathbb{Z}} s_i 2^{i-1}$ with $s_i \in \{0, 1\}$ be the dyadic expansion of $s \in \mathbb{R}_+$. Analogously, let $x = \sum_{i \in \mathbb{Z}} x_i 2^{i-1}$ with $x_i \in \{0, 1\}$. The generalized Walsh functions in $L^2([0, 1])$ are given by*

$$\text{Wal}(s, x) = (-1)^{\sum_{i \in \mathbb{Z}} (s_i + s_{i+1}) x_{-i-1}}.$$

We extend it to functions in $L^2([0, 1]^d)$ by the tensor product for $s = (s_k)_{k=1, \dots, d}$, $x = (x_k)_{k=1, \dots, d}$

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

These functions then span the sampling space, i.e. for $M = m^d$, $m \in \mathbb{N}$ we have

$$\mathcal{S}_M = \text{span} \{ \text{Wal}(s, \cdot), s = (s_k)_{k=1, \dots, d}, s_k = 1, \dots, m, k = 1, \dots, d \}.$$

Moreover, Walsh functions can be extended to negative inputs by $\text{Wal}(-s, x) = \text{Wal}(s, -x) = -\text{Wal}(s, x)$.

With the help of the Walsh functions we can define the continuous Walsh transform of a function $f \in L^2([0, 1]^d)$ as in [25]

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

The Walsh functions have the following very useful properties. First, they obey the scaling property, i.e. $\text{Wal}(2^j s, x) = \text{Wal}(s, 2^j x)$ for all $j \in \mathbb{N}$ and $s, x \in \mathbb{R}$. Second, the multiplicative identity holds, this means $\text{Wal}(s, x) \text{Wal}(s, y) = \text{Wal}(s, x \oplus y)$, where \oplus is the dyadic addition. These properties are also easily transferred to the Walsh transform. For further information on Walsh functions and transforms see [10, 15, 47].

Direct inversion from a finite amount of samples, both in the Fourier and Walsh case may lead to substantial artefacts such as the Gibbs phenomenon in the Fourier case or block artefacts known from Walsh functions. This can be seen in the numerical experiments in Figures 3c, 4c and 5b. Therefore, it is important to consider reconstruction spaces \mathcal{R} , which represent the data in a much better way, such that already a finite and low dimensional subspace leads to a good reconstruction. In many different applications such as image and signal processing, and also representations of solution manifolds for PDEs [42], wavelets have become highly popular alternatives. The reconstruction space is then spanned by reconstruction functions r_j , $j \in \mathbb{N}$. As it is not possible to deal numerically with an infinite amount of samples it also is not possible to reconstruct infinite amount of coefficients. Therefore, we have a look at the reconstruction space $\mathcal{R}_N = \text{span} \{ r_i : i = 1, \dots, N \}$. When \mathcal{R}_N is used to be an approximation for the solution manifold of a PDE we are also given the approximation error ϵ_N .

In the following we use wavelets as the reconstruction space, due to their good time and frequency localisation. First, we have a look at the one dimensional case to then get to higher dimensions. We use the common notation and denote the mother wavelet with ψ and the corresponding scaling function with ϕ . These functions are then scaled and translated. This results in the functions

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

where $R, j \in \mathbb{Z}$. The wavelet space at a certain level r is then given by $W_r := \text{span} \{\psi_{r,j} : j \in \mathbb{Z}\}$ and the scaling space is given by $V_r := \text{span} \{\phi_{r,j} : j \in \mathbb{Z}\}$. Often one is interested in the representation of functions in $L^2([0, 1])$ instead of $L^2(\mathbb{R})$. For this sake boundary corrected wavelets were introduced in [17]. The scaling space for boundary corrected wavelets is spanned by the original scaling function ϕ and reflections around 1 of the scaling function $\phi^\#$, i.e.

$$V_r^b = \text{span} \left\{ \phi_{r,n} : n = 0, \dots, 2^r - p - 1, \phi_{r,n}^\# : n = 2^r - p, \dots, 2^r - 1 \right\},$$

for Daubechies wavelets of order p . Also the boundary corrected Daubechies wavelets obey the multi resolution analysis. Therefore, it is possible to represent the union of the wavelet spaces up to a certain level $R - 1$ by the scaling space at this level, i.e.

$$\bigcup_{r < R} W_r^b = V_R^b.$$

Hence, it is not necessary for the analysis to have a deeper look in the ordering of the wavelets and their construction for the boundary corrected version as long as we have that the amount of coefficients equals the amount of elements in that level, i.e. if $N = 2^R$ for some $R \in \mathbb{N}$. The reconstruction space is then given by

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

The higher dimensional scaling spaces are constructed by tensor product of the one dimensional one, such that we get in d dimensions

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

for $N = 2^{dR}$. Remark, that the corresponding wavelet space is not simply the tensor product of the one dimensional wavelets. Fortunately, this is not often a problem in the analysis, since Daubechies wavelets obey the multi resolution analysis. We will have a look at the internal ordering for the one and two dimensional case for the Haar wavelets.

2.2. Reconstruction Techniques. In the following we present two different interesting reconstruction methods which are both optimal in their setting. We highlight some advantages and disadvantages. Especially, we see that the performance of both methods depends highly on the subspace angle between the sampling and the reconstruction space. This gives rise to the discussion in §3.1 about the question for which sampling and reconstruction spaces the stable sampling rate is linear.

2.2.1. PBDW-method. In [9, 12, 19] the PBDW-method from [41] is analysed. This method arises from the application with PDEs. One tries to estimate a state u of a physical system by solving a parametric family of PDEs depending on a parameter μ , which may not be known exactly. Therefore, the value of u can not be attained by just solving the PDE. Hence, other information are necessary. In most applications, one has access to linear measurements $l_i(u)$, $i = 1, \dots, M$ of the state u . This alone is not sufficient to estimate u or more ambitiously

even the parameter μ . Fortunately, one has also information about the PDE, which can be used to analyse the solution manifold \mathcal{M} . The solution manifold is usually quite complicated and not given directly. Hence, approximations are used. The method used in this context is the approximation by a sequence of nested finite subspaces

$$\mathcal{R}_0 \subset \mathcal{R}_1 \subset \dots \subset \mathcal{R}_N, \quad \dim(\mathcal{R}_j) = j,$$

where the approximation error is known to be ϵ_j for each subspace \mathcal{R}_j . There are a lot of different methods which allow to construct the spaces \mathcal{R}_j as the *reduced basis method* [11, 14, 18, 52] and the use of wavelets [42].

The given setting leads to the goal of trying to merge the data driven and model based information, which leads to the concept of *PBDW-method*. Let the measurement $s := P_{\mathcal{S}_M} u$ be given. The idea is to combine the information given by the measurements and the PDE. This means we are searching for an approximation $u^* \in \mathcal{K}_s$ where

$$\mathcal{K} = \{u \in \mathcal{H} : \text{dist}(u, \mathcal{R}_N) \leq \epsilon_N\} \quad \text{and} \quad \mathcal{H}_s = \{u \in \mathcal{H} : P_{\mathcal{S}_M} u = s\}.$$

The intersection is then the space of possible solutions, i.e. $\mathcal{K} \cap \mathcal{H}_s = \mathcal{K}_s$. The aim is to reduce the distance between the approximation u^* and the true solution u .

It was shown in [12] that the following approach introduced in [41] is optimal for this task. First, the minimizing problem

$$v^* = \operatorname{argmin}_{v \in \mathcal{R}_N} \|s - P_{\mathcal{S}_M} v\|^2, \quad (1)$$

is solved. The solution to the reconstruction problem is then given by the mapping A^* defined by

$$A^*(P_{\mathcal{S}_M} u) := u^* = s + P_{\mathcal{S}_M^\perp} v^*.$$

For the performance analysis, we first have a look at the *instance optimality*. This means we analyse the error for a given measurement s , i.e.

$$\|u - A(s)\| \leq C_A(s) \text{dist}(u, \mathcal{R}_N), \quad u \in \mathcal{K}_s.$$

The algorithm which leads to the smallest constant $C_A(s)$ is called *instance optimal*. It is clear that the error scales with the distance of the element u from the reconstruction space \mathcal{R}_N . Due to the fact that we normally do not know s a priori, this estimate is not very helpful. Hence, one is interested in the performance for any kind of input $s \in \mathcal{S}_M$. Therefore, the *performance of a recovery algorithm* A on any subset $W \subset \mathcal{H}$ is given by

$$E_A(W) := \sup_{u \in W} \|u - A(P_{\mathcal{S}_M} u)\|.$$

Taking the infimum over the whole class gives the *class optimal performance* on a given set W defined by

$$E(W) := \inf_A E_A(W).$$

It is shown in [12] that the presented algorithm is both instance and class optimal and gives the estimate

$$\|u - A^*(P_{\mathcal{S}_M}u)\| \leq \mu(\mathcal{R}_N, \mathcal{S}_M) \text{dist}(u, \mathcal{R}_N).$$

Hence, with this approach we do not get reasonable estimates, if $\mathcal{R}_N \cap \mathcal{S}_M^\perp \neq \{0\}$. And moreover, a detailed knowledge about the stable sampling rate is necessary to get a useful method. In addition, it was shown in [41] that this estimate can even be improved to

$$\|u - A^*(P_{\mathcal{S}_M}u)\| \leq \mu(\mathcal{R}_N, \mathcal{S}_M) \text{dist}(u, \mathcal{R}_N \oplus (\mathcal{S}_M \cap \mathcal{R}_N^\perp)). \quad (2)$$

Note that (2) demonstrates how the PBDW-method is dependent on the SSR. Moreover, it was shown in [12] that the constant $\mu(\mathcal{R}_N, \mathcal{S}_M)$ cannot be improved. Thus, the estimate is sharp, which demonstrates the importance of the SSR even stronger. Remark that (2) also shows that the error gets smaller with larger M even though we keep the reconstruction space \mathcal{R}_N the same. Hence, with increasing \mathcal{S}_M we are leaving the reconstruction space \mathcal{R}_N and get further away as M increases. This is an important observation if one is particularly interested in having the solution staying in the reconstruction space \mathcal{R}_N . If, for example, \mathcal{R}_N yields a sparse representation of the solution, it may be desirable to make sure that the solution stays in \mathcal{R}_N . In particular, if one would expand the method further and include potential subsampling, a sparse solution would be desirable. Note that this is possible for generalized sampling [44] as the recovered solution always stays in \mathcal{R}_N .

2.2.2. Generalized Sampling. After discussing the concept of the PBDW-method, we want to have a look at the different reconstruction technique *generalized sampling*. Here the approach is a stable improvement of concepts as finite section methods [13, 26, 28, 38]. The main difference is that generalized sampling allows for different dimensions on \mathcal{S}_M and \mathcal{R}_N , whereas in the finite section method they are always the same. However, the finite section method becomes a special case of generalized sampling when the dimensions of the sampling space and reconstruction space are equal. The method is defined the following way.

Definition 2 ([1]). *For $f \in \mathcal{H}$ and $N, M \in \mathbb{N}$, we define the reconstruction method of generalized sampling $G_{N,M} : \mathcal{H} \rightarrow \mathcal{R}_N$ by*

$$\langle P_{\mathcal{S}_M} G_{N,M}(f), r_j \rangle = \langle P_{\mathcal{S}_M} f, r_j \rangle, \quad j = 1, \dots, N, \quad (3)$$

where $P_{\mathcal{S}_M}$ denotes the orthogonal projection on the subspace \mathcal{S}_M . We also refer to $G_{N,M}(f)$ as the generalized sampling reconstruction of f .

We stress at this point that generalized sampling is also a linear reconstruction scheme. In particular, one solves the following linear equation for $\alpha_N \in \mathbb{R}^N$.

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

where

$$U^{[N,M]} = \begin{pmatrix} u_{11} & \dots & u_{1N} \\ \vdots & \ddots & \vdots \\ u_{M1} & \dots & u_{MN} \end{pmatrix} \quad (4)$$

and $u_{ij} = \langle r_j, s_i \rangle$, $l(f)^{[M]} = (l_1(f), \dots, l_M(f)) \in \mathbb{R}^M$. The matrix can be seen in Figure 1 for different sampling and reconstruction spaces. The reconstruction is given by $G_{N,M}(f) = \sum_{i=1}^N \alpha_i r_i$. An interesting part about this is, that the matrix $U^{[N,N]}$ is very ill-conditioned for most cases. As pointed out in [3], in the case of Fourier samples and wavelet reconstructions, the condition number grows exponentially in N . This means that this approach is only feasible due to the above mentioned allowance of a different amount of samples and reconstructed coefficients. It can be shown, that there exist a certain amount of samples, such that (3) obeys a solution.

Theorem 1 ([2]). *Let $N \in \mathbb{N}$. Then, there exists $M_0 \in \mathbb{N}$, such that, for every $f \in \mathcal{H}$, (3) has an unique solution $G_{N,M}(f)$ for all $M \geq M_0$. Moreover, the smallest M_0 is the least number such that $\cos(\omega(\mathcal{R}_N, \mathcal{S}_{M_0})) > 0$.*

Just as for the PBDW-method, performance bounds of generalized sampling were studied. Here we observe again that the reconstruction quality highly depends on the subspace angle and on the relation between the data and the reconstruction space.

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

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

and

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

In particular, these bounds are sharp.

Considering only mappings which map into the reconstruction space \mathcal{R}_N , we get that generalized sampling is also optimal in terms of achieving low condition numbers and accuracy.

In [44] a consistent approach of generalized sampling was introduced. Let $f \in \mathcal{R}$ then it can be represented as $f = \sum_{j=1}^{\infty} x_j r_j$. The measurements can be written as $f_M = P_{[M]} U x$, where $P_{[M]}$ is the orthogonal projection onto space spanned by the first M elements and U is defined as in (4). The introduced method solves the non linear minimization problem

$$\inf_{\mu \in \ell_1} \|\mu\|_{\ell_1} \text{ with } P_{[M]} U \mu = P_{[M]} U x. \quad (5)$$

The solution is then given by

$$g_M = \sum_{j=1}^{\infty} \mu_j r_j.$$

The measurements of g_M and f are naturally equal. Hence, this approach is consistent and maps into the reconstruction space. In contrast to generalized sampling it is not necessary to decide the amount of reconstructed coefficients a priori.

In the setting of arbitrary sampling and reconstruction spaces, it was shown that for every amount of reconstructed coefficients N there exist some M_0 such that the method reconstructs the data correct up to its first N coefficients. Hence, this approach is convergent as the error is given by $\mathcal{O}(\|P_{\mathcal{R}_N}^\perp f\|)$ which goes to zero for $M \rightarrow \infty$. The speed of convergence is then depending on the sampling and reconstruction space. This speed was analysed in [44] for the case of Fourier measurements and wavelet reconstruction. Let the reconstruction space \mathcal{R} be given by wavelets and the sampling space \mathcal{S} be the space representing Fourier measurements. Then, for $x \in \ell_1(\mathbb{N})$ and $N \in \mathbb{N}$, the following holds:

- (1) If for some $A > 0$ and $\alpha \geq 1$, the Fourier transform of the scaling function ϕ decays with

$$|\hat{\phi}(\xi)| \leq \frac{A}{(1 + |\xi|)^\alpha}, \quad \xi \in \mathbb{R}$$

then there exists some constant C independent of N (but dependent on α and ϵ) such that for $M = CN^{1+1/(2\alpha-1)}$, any solution ξ to (5) satisfies

$$\|\xi - x\|_{\ell_1} \leq 6\|P_{[N]}^\perp x\|_{\ell_1}.$$

- (2) If for $k = 0, 1, 2$, for some $A > 0$ and $\alpha \geq 1.5$, the Fourier decay of the scaling function ϕ , the wavelet ψ and their first two derivatives decays with

$$|\hat{\phi}^{(k)}(\xi)| \leq \frac{A}{(1 + |\xi|)^\alpha}, \quad |\hat{\psi}^{(k)}(\xi)| \leq \frac{A}{(1 + |\xi|)^\alpha}, \quad \xi \in \mathbb{R},$$

then there exists some constant C independent of N (but dependent on ϕ , ψ and ϵ) such that for $M = CN$, any solution ξ to (5) satisfies

$$\|\xi - x\|_{\ell_1} \leq 6\|P_{[N]}^\perp x\|_{\ell_1}.$$

The assumptions of this result are natural for example for Daubechies wavelets. The first assumptions is fulfilled by all Daubechies wavelets and the second one is fulfilled for Daubechies wavelets of 7 or more vanishing moments. Numerical experiments suggest that maybe even Daubechies wavelets with less vanishing moments might have a linear relationship [44].

Finally, we would like to mention the stability of this approach. Stability is considered in the ℓ_1 setting and the used definition is equal to the existence of the condition number of the method. This means that the problem is well-conditioned in terms of solving an ℓ_1 minimization problem rather than the reconstruction of f from M samples. Especially, this means that the problem is not robust against noise.

2.2.3. Comparison. Both methods are optimal in their setting. The PBDW-method is optimal in contrast to other consistent methods and generalized sampling is optimal in the class of reconstruction methods, which map to the reconstruction space \mathcal{R}_N . Also generalized sampling can be changed to a consistent method, however, this method is then not robust. Moreover, both methods have in common that their accuracy and stability is depending on the reciprocal of the subspace angle $\mu(\mathcal{R}_N, \mathcal{S}_M)$. Therefore, we devote the following §3 to the analysis of the relation between the dimension of the sampling and the reconstruction space to bound the subspace angle. This similarity in the dependence is not to surprising. When we compare the two methods. We see that the intermediate step v^* in (1) is the same as the generalized sampling solution. Hence, one is solving the same least square problem. However, in the PBDW-method this solution is tweaked to make the solution consistent. This is done by adding the measurements and removing the measurements of the calculated solution, i.e.

$$u^* = v^* + P_{\mathcal{S}_M}u - P_{\mathcal{S}_M}v^*.$$

3. Stable Sampling Rate

3.1. Linearity of the Stable Sampling Rate. In §2.2 we saw that the subspace angle between the sampling and the reconstruction space controls the reconstruction accuracy. Therefore, one is interested in the relation between the amount of samples and the amount of reconstructed coefficients, such that the subspace angle is bounded. In detail we are interested in the stable sampling rate:

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

The stable sampling rate has been analysed for important cases which appear frequently in practice, i.e. for the Fourier-Wavelet [3], Fourier-Polynomial bases [31] and Walsh-Wavelet cases [30]. For the reconstruction with wavelets we get for both Fourier and Walsh sampling that the stable sampling rate is linear. This is the best relation one can wish for and means that the methods discussed above are, up to a constant, as good as if one could access the wavelet coefficients directly. In particular, for the Walsh case we have the following theorem.

Theorem 3 ([30]). *Let \mathcal{S} and \mathcal{R} be the sampling and reconstruction spaces of Walsh functions and boundary wavelets in $L^2([0, 1]^d)$. If $N = 2^{dR}$ where $R \in \mathbb{N}$, then for every $\theta \in (1, \infty)$ there exist a constant S_θ depending on θ such that $\Theta(N, \theta) \leq S_\theta N = \mathcal{O}(N)$ for any $\theta \in (1, \infty)$, i.e. the stable sampling rate is linear.*

A natural question which arises from this theorem is, if it is possible to give sharp bounds on the constant S_θ . In Figure 1 we can see the stable sampling rate for different Wavelets and the bound $\theta = 2$. The slope S_θ is unknown in most cases and very difficult to find. This comes from the fact that for the majority of wavelets the reconstruction matrix is not perfectly block diagonal as

in 1d and 1f. Hence, one has to take the off diagonals into consideration. The numerics suggest that the slope is higher the further away the reconstruction matrix gets from block diagonal. Only for the case of Haar wavelets and Walsh functions we get that the reconstruction matrix is perfectly block diagonal, as it can be seen in 1b. Note that from the numerical example one may deduce that $S_\theta = 1$. This is indeed the case, and the analysis detailed below establishes that $S_\theta = 1$ for all $\theta \in (1, \infty)$.

3.2. Sharpness for the Haar wavelet - Walsh case. The sharp bound on S_θ can be summarised in the following theorem.

Theorem 4. *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. $\Theta(N, \theta) = N$.*

For the proof we first have a look at the behaviour of Haar wavelets in one dimension under the Walsh transform. This way we also get a theoretical argument for the block structure that can be seen in the numerical implementation.

Lemma 1. *Let $\psi = \mathcal{X}_{[0,1/2]} - \mathcal{X}_{(1/2,1]}$ be the Haar wavelet. Then we have that*

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

Proof. For the scalarproduct we have

$$\begin{aligned} \langle \psi_{R,j}, \text{Wal}(n, \cdot) \rangle &= \int_0^1 2^{R/2} (\mathcal{X}_{[0,1/2]}(2^R x - j) - \mathcal{X}_{(1/2,1]}(2^R x - j)) \text{Wal}(n, x) dx \\ &= 2^{R/2} \left(\int_{\Delta_{2^j}^{R+1}} \text{Wal}(n, x) dx - \int_{\Delta_{2^{j+1}}^{R+1}} \text{Wal}(n, x) dx \right) \end{aligned}$$

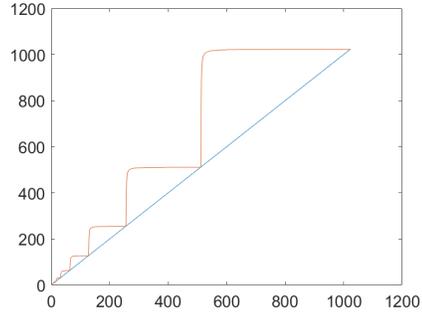
Let $\Delta_k^p = [2^{-p}k, 2^{-p}(k+1))$. We know from [8] that the function $\text{Wal}(n, x)$ for $2^p \leq n < 2^{p+1}$ takes the value +1 on the interval $\Delta_{2^k}^{p+1}$ or $\Delta_{2^{k+1}}^{p+1}$ and -1 on the other one for $k = 0, \dots, 2^p - 1$. Now, we have a look at three different cases.

Case 1: $n < 2^R$. There exist $r < R$ such that $2^r \leq n < 2^{r+1}$. Then, the function $\text{Wal}(n, x)$ is constant on the interval Δ_k^r for any $k = 0, \dots, 2^r - 1$. Remark, that for $j = 0, \dots, 2^R - 1$ we have that the rounding error is bounded as follows

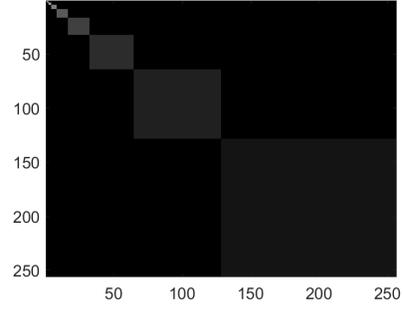
$$\lfloor 2^{r-R}j \rfloor \geq 2^{r-R}j - (1 - 2^{r-R}).$$

Then we have the interval inclusion

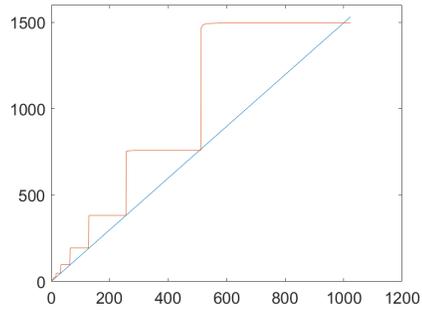
$$\begin{aligned} \Delta_{2^j}^{R+1} &= [2^{-R-1}(2j), 2^{-R-1}(2j+1)) \\ &= [2^{-r}2^{-R-1+r}2j, 2^{-r}2^{-R-1+r}(2j+1)) \\ &\subset [2^{-r}\lfloor 2^{r-R}j \rfloor, 2^{-r}(\lfloor 2^{r-R}j \rfloor + 1)) = \Delta_{\lfloor 2^{r-R}j \rfloor}^r. \end{aligned}$$



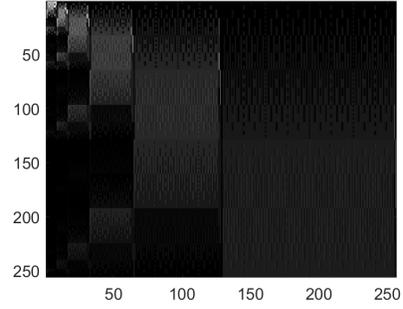
(a) Haar Wavelet where
 $\Theta(N, 2) = N$



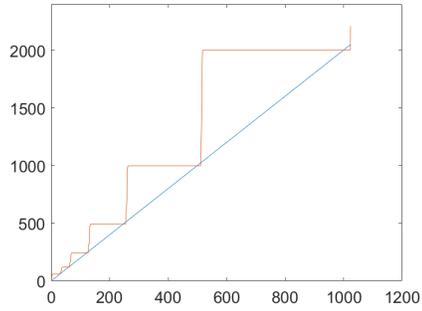
(b) Haar-Walsh



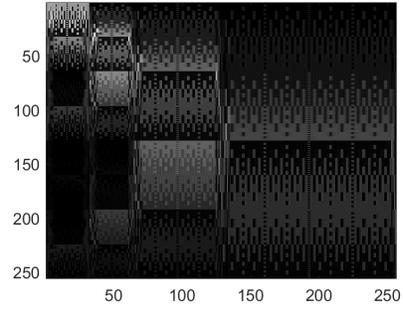
(c) Daubechies
 2 Wavelet where
 $\Theta(N, 2) = 1.49N$



(d) db2 - Walsh



(e) Daubechies
 8 Wavelet where
 $\Theta(N, 2) = 2N$



(f) db8 - Walsh

Figure 1. Stable sampling rate for $\theta = 2$ and reconstruction matrix

Moreover,

$$\Delta_{2^{j+1}}^{R+1} \subset \Delta_{[2^{r-R}j]}^r.$$

Hence, $\text{Wal}(n, x)$ takes the same value on $\Delta_{2^j}^{R+1}$ and $\Delta_{2^{j+1}}^{R+1}$. Therefore, the two integrals are equal and the scalarproduct vanishes.

Case 2: $2^R \leq n < 2^{R+1}$. We have for $j = 0, \dots, 2^R - 1$ that $\text{Wal}(n, x)$ is equal to $+1$ on $\Delta_{2^j}^{R+1}$ or $\Delta_{2^{j+1}}^{R+1}$ and -1 on the other. Therefore, we have either

$$\begin{aligned} \langle \psi_{R,j}, \text{Wal}(n, x) \rangle &= 2^{R/2} \left(\int_{\Delta_{2^j}^{R+1}} \text{Wal}(n, x) dx - \int_{\Delta_{2^{j+1}}^{R+1}} \text{Wal}(n, x) dx \right) \\ &= 2^{R/2} \left(\int_{\Delta_{2^j}^{R+1}} 1 dx - \int_{\Delta_{2^{j+1}}^{R+1}} -1 dx \right) = 2^{R/2} (2^{-R}) = 2^{-R/2}. \end{aligned}$$

Or in the other case analogously

$$\langle \psi_{R,j}, \text{Wal}(n, x) \rangle = 2^{R/2} \left(\int_{\Delta_{2^j}^{R+1}} -1 dx - \int_{\Delta_{2^{j+1}}^{R+1}} 1 dx \right) = 2^{R/2} (-2^{-R}) = -2^{-R/2}.$$

Now, we are left with the last case.

Case 3: $n \geq 2^{R+1}$. There exists an integer $r \geq R + 1$ such that $2^r \leq n < 2^{r+1}$. Similar to the first case. Moreover, we have for $j = 0, \dots, 2^R - 1$ that

$$\Delta_{2^j}^{R+1} = \bigcup_{l=2^{r-R}j}^{2^{r-R}j+2^{r-R}-1} \Delta_{2^l}^{r+1} \cup \Delta_{2^{l+1}}^{r+1}.$$

and with the fact that $\text{Wal}(n, x)$ takes the value $+1$ on one of the intervals $\Delta_{2^l}^r$ and $\Delta_{2^{l+1}}^r$ and -1 on the other, we have that the function $\text{Wal}(n, x)$ takes the values $+1$ and -1 on half of the interval of $\Delta_{2^j}^{R+1}$. Therefore, the integral vanishes. The same holds true for $\Delta_{2^{j+1}}^{R+1}$, such that we get the desired result. \square

Before we prove Theorem 4 we analyse the reconstruction matrix for the Haar wavelet - Walsh case in two dimensions. The amount of functions which span the wavelet space in d dimensions growth exponentially with d . Therefore, we restrict ourselves to two dimensions to underline the main idea. In 2 we can see that the reconstruction matrix in 2 dimensions has additional structure in each level. Similar to the one dimensional case we have perfect block structure for the Haar case and nearly block structure for the higher order wavelets. For the analysis of this phenomena in the Haar wavelet - Walsh case we have a look at the definition and order of the two dimensional Haar wavelet. Remark, that this is necessary for the analysis of the reconstruction matrix, instead of the SSR. In two dimensions the wavelets are constructed by the tensor product of

the wavelet and the scaling function and the tensor product only between the wavelets. Especially we have for $R \in \mathbb{N}, 0 \leq j_1, j_2 \leq 2^R - 1$ that

$$\psi_{R,j_1,j_2,l}(x_1, x_2) = \begin{cases} \phi_{R,j_1}(x_1)\psi_{R,j_2}(x_2) & l = 1 \\ \psi_{R,j_1}(x_1)\phi_{R,j_2}(x_2) & l = 2 \\ \psi_{R,j_1}(x_1)\psi_{R,j_2}(x_2) & l = 3 \end{cases} \quad (6)$$

and for the first level scaling function

$$\phi(x_1, x_2) = \phi(x_1)\phi(x_2).$$

For the order of the reconstruction matrix we first take the first level scaling function ϕ . Then, we increase by levels, in each level R we let first j_1 go from $0, \dots, 2^R - 1$ and then $j_2 = 0, \dots, 2^R - 1$. Finally, we let $l = 1, \dots, 3$. Such that we get for the order of the wavelets: $\phi, \psi_{R,0,0,1}, \dots, \psi_{R,2^R-1,0,1}, \psi_{R,0,1,1}, \dots, \psi_{R,2^R-1,1,1}, \dots, \psi_{R,2^R-1,2^R-1,1}, \psi_{R,0,0,2}, \dots, \psi_{R,2^R-1,2^R-1,3}$.

Due to the fact that the higher dimensional wavelets are constructed also by means of the scaling functions, it is necessary to analyse the decay rate for the scaling function as well. Moreover, this is also a main ingredient for the proof of Theorem 4 as we represent the union of the wavelet spaces by the scaling space.

Lemma 2. *Let $\phi = \mathcal{X}_{[0,1]}$ be the Haar scaling function. Then we have that the Walsh transform obeys the following block and decay structure*

$$|\langle \phi_{R,j}, \text{Wal}(n, \cdot) \rangle| = \begin{cases} 2^{-R/2} & n < 2^R, 0 \leq j \leq 2^R - 1 \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

Proof. The scalarproduct can be expressed as integral over the interval Δ_j^R .

$$\begin{aligned} \langle \phi_{R,j}, \text{Wal}(n, \cdot) \rangle &= \int_0^1 2^{R/2} \mathcal{X}_{[0,1]}(2^R x - j) \text{Wal}(n, x) dx \\ &= 2^{R/2} \int_{\Delta_j^R} \text{Wal}(n, x) dx. \end{aligned}$$

We look at the two different cases

Case 1: $n < 2^R$ Remember from before that $\text{Wal}(n, x)$ is constant to $+1$ or -1 on the interval Δ_j^R for $j = 0, \dots, 2^R - 1$. Hence, we get that

$$|\langle \phi_{R,j}, \text{Wal}(n, \cdot) \rangle| = |2^{R/2} \int_{\Delta_j^R} \text{Wal}(n, x) dx| = 2^{-R/2}.$$

Case 2: $n \geq 2^R$ This follows as in Case 3 of Theorem 1. With the difference that we are looking at the integral over the interval Δ_j^R instead of the two integrals Δ_{2j}^{R+1} and Δ_{2j+1}^{R+1} . Nevertheless, they vanish for the same reason. \square

With this in hand we can now state the structure of the reconstruction matrix in two dimensions.

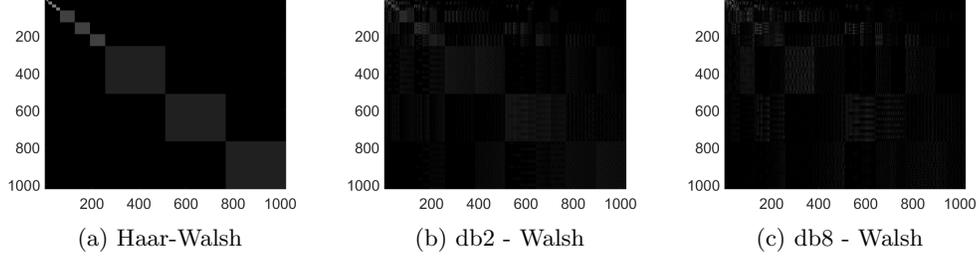


Figure 2. Reconstruction matrix in 2 dimensions

Corollary 1. Let $\psi_{R,j_1,j_2,l}$ be the Haar wavelet defined as in (6). Then, the Walsh transform has the following property for $0 \leq j_1, j_2 \leq 2^R - 1$

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

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

and for the third version

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

Proof. The proof follows directly from the tensor product structure and Theorem 1 and Lemma 2. \square

After this sorrow analysis of the behaviour of the wavelet and the scaling function under the Walsh transform. We are able to proof Theorem 4.

Proof of Theorem 4. We want to analyse the subspace angle $\mu(\mathcal{R}_N, \mathcal{S}_M)$ for $N = M$. In detail, we are interested in bounding $\mu(\mathcal{R}_M, \mathcal{S}_M) < \theta$ for all $\theta \in (1, \infty)$. Hence, we try to show that $\mu(\mathcal{R}_M, \mathcal{S}_M) = 1$ or equally $1/\mu(\mathcal{R}_M, \mathcal{S}_M) = 1$ for $M = 2^{dR}$. Due to the fact that the circle $\{r \in \mathcal{R}_M, \|r\| = 1\}$ is compact and the orthogonal projection is continuous there exist $\varphi \in \mathcal{R}_M, \|\varphi\| = 1$ such that we have

$$\frac{1}{\mu(\mathcal{R}_M, \mathcal{S}_M)} = \inf_{r \in \mathcal{R}_M, \|r\|=1} \|P_{\mathcal{S}_M} r\| = \|P_{\mathcal{S}_M} \varphi\| = 1 - \|P_{\mathcal{S}_M}^\perp \varphi\|.$$

The minimal element φ can be represented by

$$\varphi = \sum_{j=0}^{2^R-1} \bigotimes_{i=1}^d \alpha_j \phi_{R,j} \quad \text{with} \quad \sum_{j=0}^{2^R-1} |\alpha_j|^2 = 1,$$

where the multi index notation is used. Then we have that

$$\begin{aligned} P_{\mathcal{S}_M}^\perp \varphi &= \sum_{n=2^{R+1}}^{\infty} \left\langle \sum_{j=0}^{2^R-1} \bigotimes_{i=1}^d \alpha_j \phi_{R,j}, \text{Wal}(n, \cdot) \right\rangle \\ &= \sum_{n=2^{R+1}}^{\infty} \sum_{j=0}^{2^R-1} \sum_{i=1}^d \alpha_{j_i} \langle \phi_{R,j_i}, \text{Wal}(n_i, \cdot) \rangle. \end{aligned}$$

With (7) we get that this sum vanishes. Hence,

$$\mu(\mathcal{R}_M, \mathcal{S}_M) = 1$$

as desired. \square

4. Numerical Experiments

In this section we want to underline the differences between generalized sampling and the PBDW technique with examples using both Fourier and Walsh samples. We emphasise that some of the examples are chosen particularly to highlight the differences between the two methods. Hence, the examples may not reflect typical practical scenarios. In Figure 3 we look at an extreme case and consider the task of recovering the Haar wavelet from Fourier coefficient. Generalized sampling will obviously recover the function perfectly given that we choose the Haar basis for the reconstruction space. Although in this very special case, since one gets perfect recovery, the recovered solution is consistent with the samples, this is not the case in general. In particular, generalized sampling will usually provide non-consistent solution. The PBDW-method on the other hand is always consistent. The effect of this is that, even when the reconstruction space is fixed, the solution changes with the number of samples. Moreover, in this example it becomes closer, as the number of samples increase, to the solution provided by simply truncating the Fourier series.

The next example in Figure 4 considers Walsh samples. The original function displayed in Figure 4a is continuous with two jump discontinuities. Moreover, the continuous part is very well represented with Daubechies 8 wavelets, as demonstrated by the generalized sampling reconstruction in Figure 4b. Nevertheless, there are some artefacts at the jumps. Walsh functions represent the jumps better but lead to heavy artefacts along the continuous part of the function 3c. Hence, both spaces have pros and cons and can represent the signal well in different areas. The PBDW-method allows us to take advantages from both spaces, as it does not force the solution to stay in the reconstruction space. This leads to a different reconstruction quality as in 4d and 4f. In this case we also get better results with more samples even though we reconstruct the same amount of coefficients.

In the last example we used the code from [24] and consider reconstruction of images from Fourier samples. In Figure 5 we see that the PBDW-method provides good results in the 2D setting and with Fourier measurements. The function is very smooth outside the discontinuous part. Therefore it can be nicely

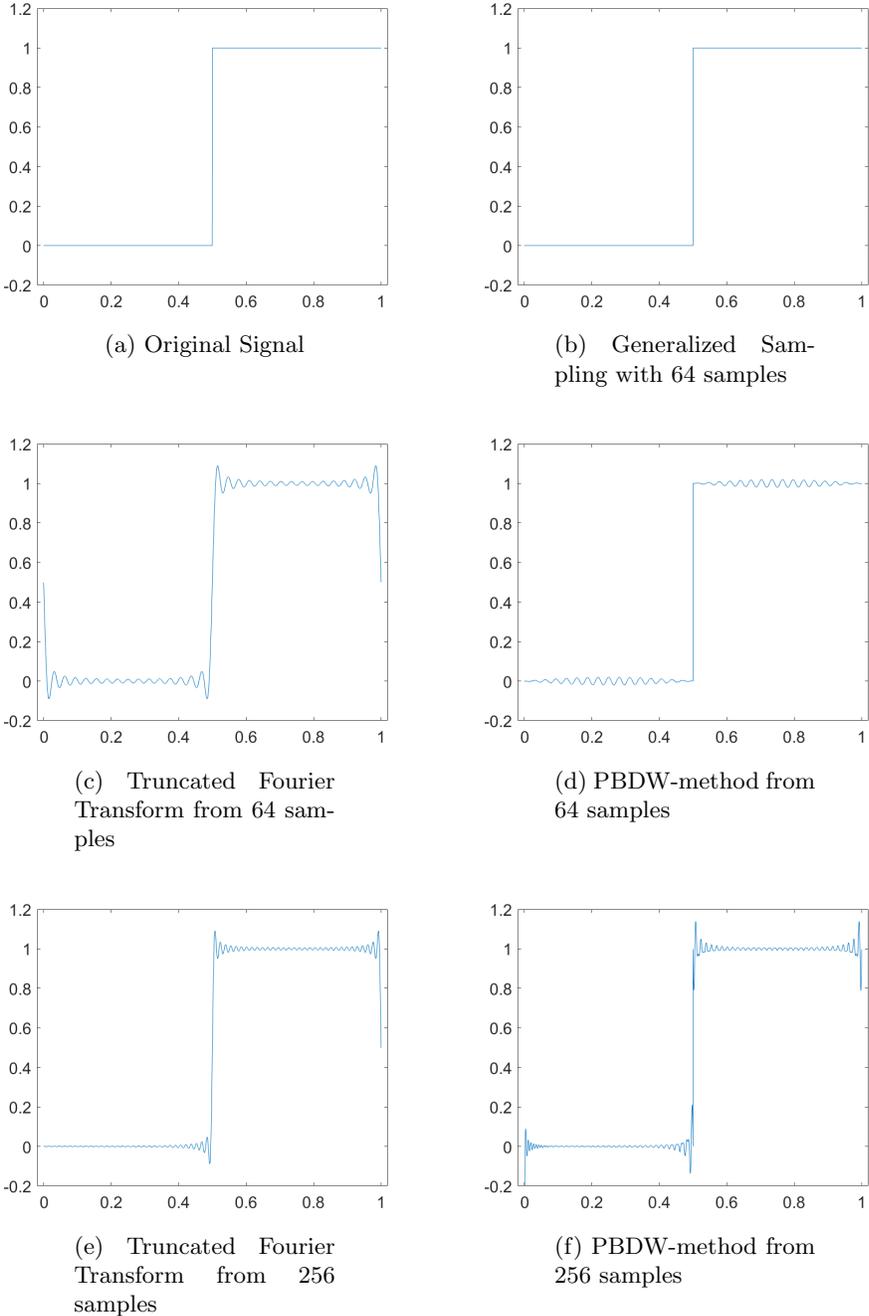


Figure 3. Reconstruction from Fourier measurements with Haar wavelets and $\dim \mathcal{R}_N = 32$

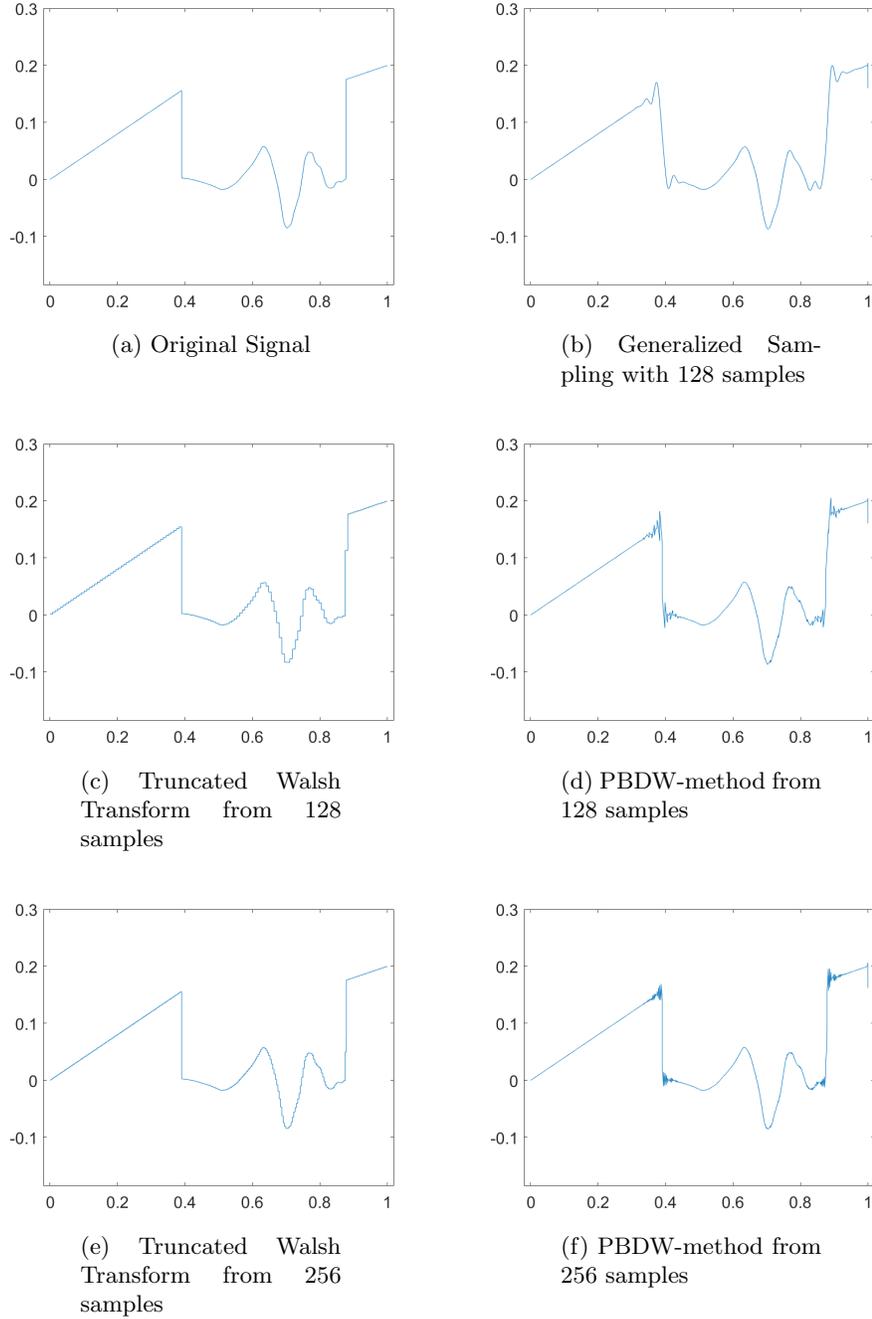
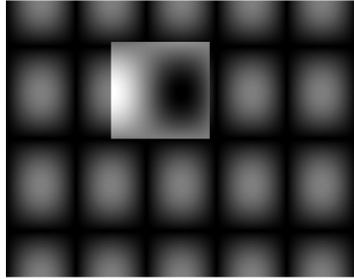


Figure 4. Reconstruction from binary measurements with Daubechies wavelets 8 and $\dim \mathcal{R}_N = 64$



(a) Original Signal

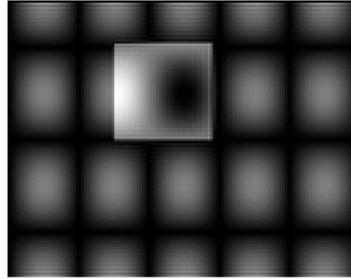
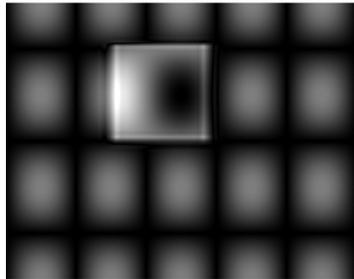
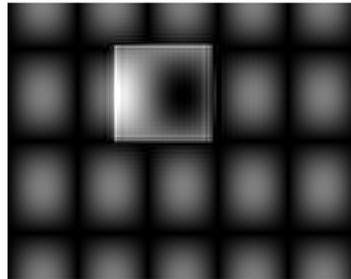
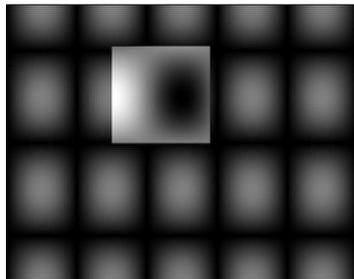
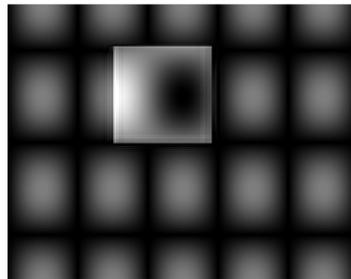
(b) Truncated Fourier transform with 128^2 samples(c) Generalized sampling from 128^2 samples(d) PBDW-method from 128^2 samples(e) Truncated Fourier Transform from 256^2 samples(f) PBDW-method from 256^2 samples

Figure 5. Reconstruction from Fourier measurements with Daubechies wavelets 4 and $\dim \mathcal{R}_N = 64^2$

represented with wavelets in this area. Nevertheless, there are some obvious artefacts around the discontinuity in Figure 5c. The artefacts that arise from the truncated Fourier transform are also easy to spot in Figure 5b and less clear in Figure 5e due to the increased amount of samples. With the PBDW-method it is possible to merge the advantages of both systems and decrease the error, as seen in Figures 5d and 5f. We can also see that an increased amount of samples leads to better performance of the PBDW-method even if the amount of reconstructed coefficients stays the same.

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, Oct 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] A. Aldroubi and M. Unser. A general sampling theory for nonideal acquisition devices. *IEEE Trans. Signal Process.*, 42(11):2915–2925, 1994.
- [8] V. Antun. Coherence estimates between hadamard matrices and daubechies wavelets. Master’s thesis, University of Oslo, 2016.
- [9] 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.
- [10] K. Beauchamp. *Walsh Functions and their Applications*. Academic Press, London, 1975.
- [11] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk. Convergence rates for greedy algorithms in reduced basis methods. *SIAM journal on mathematical analysis*, 43(3):1457–1472, 2011.
- [12] 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.
- [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] A. Buffa, Y. Maday, A. T. Patera, C. Prudhomme, and G. Turinici. A priori convergence of the greedy algorithm for the parametrized reduced basis method. *ESAIM: Mathematical Modelling and Numerical Analysis*, 46(3):595–603, 2012.

- [15] P. Butzer and H. Wagner. On the dyadic analysis based on pointwise dyadic derivative. *Anal. Math.*, (1):171–196, 1975.
- [16] 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.
- [17] A. Cohen, I. Daubechies, and P. Vial. Wavelets on the interval and fast wavelet transforms. *Comput. Harmon. Anal.*, 1(1):54–81, 1993.
- [18] R. DeVore, G. Petrova, and P. Wojtaszczyk. Greedy algorithms for reduced bases in banach spaces. *Constructive Approximation*, 37(3):455–466, 2013.
- [19] R. DeVore, G. Petrova, and P. Wojtaszczyk. Data assimilation and sampling in banach spaces. *Calcolo*, 54(3):963–1007, 2017.
- [20] T. Dvorkind and Y. C. Eldar. Robust and consistent sampling. *IEEE Signal Process. Letters*, 16(9):739–742, 2009.
- [21] Y. C. Eldar. Sampling with arbitrary sampling and reconstruction spaces and oblique dual frame vectors. *J. Fourier Anal. Appl.*, 9(1):77–96, 2003.
- [22] Y. C. Eldar and G. Kutyniok. *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012.
- [23] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. Springer Science+Business Media, New York, 2013.
- [24] M. Gataric and C. Poon. A practical guide to the recovery of wavelet coefficients from fourier measurements. *SIAM J. Sci. Comput.*, 38(2):A1075–A1099.
- [25] E. Gauss. *Walsh Funktionen für Ingenieure und Naturwissenschaftler*. Springer Fachmedien, Wiesbaden, 1994.
- [26] K. Gröchenig, Z. Rzeszutnik, 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.
- [27] 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.
- [28] A. C. Hansen. On the approximation of spectra of linear operators on hilbert spaces. *J. Funct. Anal.*, 8(254):2092–2126, 2008.
- [29] 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.
- [30] A. C. Hansen and L. Terhaar. On the stable sampling rate for binary measurements and wavelet reconstruction. preprint, 2017.
- [31] T. Hrycak and K. Gröchenig. Pseudospectral fourier reconstruction with the modified inverse polynomial reconstruction method. *J. Comput. Phys.*, 229(3):933–946, 2010.
- [32] K. Hwan Jin, M. McCann, E. Froustey, and M. Unser. Deep convolutional neural network for inverse problems in imaging. PP, 11 2016.
- [33] A. J. Jerri. The shannon sampling theorem its various extensions and applications: A tutorial review. *Proc. IEEE*, (65):1565–1596, 1977.
- [34] 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.
- [35] 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.
- [36] R. Leary, Z. Saghi, P. A. Midgley, and D. J. Holland. Compressed sensing electron tomography. *Ultramicroscopy*, 131(0):70–91, 2013.
- [37] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

- [38] M. Lindner. *Infinite matrices and their finite sections: An introduction to the limit operator method*. Birkhäuser Verlag, Basel, 2006.
- [39] M. Lustig, D. Donoho, and J. M. Pauly. Sparse mri: the application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine*, 2007.
- [40] J. Ma. Generalized sampling reconstruction from fourier measurements using compactly supported shearlets. *Appl. Comput. Harmon. Anal.*, 2015.
- [41] 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.
- [42] N. Mahadevan and K. A. Hoo. Wavelet-based model reduction of distributed parameter systems. 55:4271–4290, 10 2000.
- [43] M. Müller. *Introduction to Confocal Fluorescence Microscopy*. SPIE, Bellingham, Washington, 2006.
- [44] C. Poon. A consistent and stable approach to generalized sampling. *J. Fourier Anal. Appl.*, (20):985–1019, 2014.
- [45] 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.
- [46] B. Roman, B. Adcock, and A. C. Hansen. On asymptotic structure in compressed sensing. *arXiv:1406.4178*, 2014.
- [47] F. Schipp, P. Simon, and W. Wade. *Walsh Series an Introduction to dyadic harmonic Analysis*. Adam Hilger, Bristol and New York, 1990.
- [48] C. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, (27):379–423, 623–656, 1948.
- [49] 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.
- [50] M. Unser. Sampling - 50 years after shannon. *Proc. IEEE*, 4(88):569–587, 2000.
- [51] M. Unser and J. Zerubia. A generalized sampling theory without band-limiting constraints. *IEEE Trans. Circuits Syst. II.*, 45(8):959–969, 1998.
- [52] P. Wojtaszczyk. On greedy algorithm approximating kolmogorov widths in banach spaces. *Journal of Mathematical Analysis and Applications*, 424(1):685–695, 2015.