Recovering wavelet coefficients from binary samples using fast transforms*

Vegard Antun†

February 16, 2023

Abstract

Recovering a signal (function) from finitely many binary or Fourier samples is one of the core problems in modern imaging, and by now there exist a plethora of methods for recovering a signal from such samples. Examples of methods, which can utilise wavelet reconstruction, include generalised sampling, infinite-dimensional compressive sensing, the parameterised-background data-weak (PBDW) method etc. However, for any of these methods to be applied in practice, accurate and fast modelling of an $N \times M$ section of the infinite-dimensional change-of-basis matrix between the sampling basis (Fourier or Walsh-Hadamard samples) and the wavelet reconstruction basis is paramount. Building on the work of [Gataric & Poon, *SIAM J. Sci. Comput.* 38 (2016) pp. A1075-A1099] we derive an algorithm, which bypasses the NM storage requirement and the $\mathcal{O}(NM)$ computational cost of matrix-vector multiplication with this matrix and its adjoint when using Walsh-Hadamard samples and wavelet reconstruction. The proposed algorithm computes the matrix-vector multiplication in $\mathcal{O}(N \log N)$ operations and has a storage requirement of $\mathcal{O}(2^q)$, where $N = 2^{dq}M$, (usually $q \in \{1, 2\}$) and $d = 1, 2$ is the dimension. As matrix-vector multiplications is the computational bottleneck for iterative algorithms used by the mentioned reconstruction methods, the proposed algorithm speeds up the reconstruction of wavelet coefficients from Walsh-Hadamard samples considerably.

Keywords: Fast transforms, Sampling theory, Wavelets, Walsh functions, Walsh-Hadamard samples

Mathematics Subject Classification (2010) 94A20, 94A11, 42C10, 42C40, 46C05.

1 Introduction

Approximating a function from finitely many samples is one of the fundamental problems in approximation theory, and, by now, there exist myriads of conditions and algorithms for obtaining good function approximation. The problem is often motivated by the many applications in natural sciences where one is given a finite set of samples of an underlying unknown signal (function) that one wants to recover (approximate).

In this work, we consider the recovery of signals, where physical constraints dictate the type of samples one can acquire. This is a well-studied problem with numerous applications in medical imaging. Examples include Magnetic Resonance Imaging (MRI) [43, 44], surface scattering [39, 40], X-ray Computed Tomography (CT) [29] and electron microscopy [42], all of which employ Fourier sampling. Other examples, employing binary samples, include fluorescence microscopy [55], lensless imaging [16] and compressive holography [20].

*Submitted to the editors 17 June 2021.

†Department of Mathematics, University of Oslo, Norway (Email: vegarant@math.uio.no).

Given the long list of applications, there are many efficient methods for reconstructing a function from a fixed sampling modality. Examples of such methods include *generalised sampling* [3, 6, 7, 10, 38, 45], studied by Adcock, Hansen, Hrycak, Gröchenig, Kutyniok, Ma, Poon, Shadrin and others; its predecessor *consistent sampling* [26, 27, 28, 37, 60, 61], developed by Aldroubi, Eldar, Unser and others. More recently Adcock, Antun, Hansen, Kutyniok, Lim, Poon, Thesing and many others have analysed reconstruction methods based on *infinite-dimensional compressive sensing* [1, 4, 8, 41, 52, 58]. Other approaches can be found within data assimilation. A first approach here was introduced by Maday & Mula in [47], called *generalised empirical interpolation method*, this was later followed by the *Parametrized Background Data-Weak* (PBDW) method, developed by Maday, Patera, Penn & Yano in [46, 48], and later analysed by Binev, Cohen, Dahmen, DeVore, Petrova, and Wojtaszczyk in [15, 23].

This work extends the work of Gataric & Poon in [30] to binary sampling. In [30] they derive an algorithm for computing a matrix-vector multiplication with a $N \times M$ section of the change-of-basis matrix between a Fourier sampling basis and a wavelet reconstruction basis using $\mathcal{O}(N \log N)$ operations. Here we extend their approach to Walsh sampling and wavelet reconstruction. As we shall see, the analysis allowing us to do this is different, while the underlying algorithm is the same.

We model the problem as follows. Let $\mathcal{H}$ be an infinite-dimensional separable Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. Let $\{s_k : k \in \mathbb{N}\}$ and $\{r_k : k \in \mathbb{N}\}$ be two orthonormal bases for $\mathcal{H}$, called the sampling and reconstruction basis, respectively. Furthermore define the sampling space as the linear span $\mathcal{S}_N = \operatorname{span}\{s_1, \dots, s_N\}$ and the reconstruction space as $\mathcal{R}_M = \operatorname{span}\{r_1, \dots, r_M\}$.

Suppose that we can only observe the function $f \in \mathcal{H}$, using *finitely many* linear measurements $\langle f, s_k \rangle$, $k = 1, \dots, N$. Since $\{s_k : k \in \mathbb{N}\}$ is an orthonormal basis, this immediately gives the truncated series approximation

$$f_N = y_1 s_1 + \cdots + y_N s_N \in \mathcal{S}_N \tag{1}$$

where $y_k = \langle f, s_k \rangle$. In all the applications mentioned above, we have limited freedom in designing the sampling basis $\{s_k : k \in \mathbb{N}\}$ and the approximation f_N may, therefore, suffer from unpleasant reconstruction artefacts due to the characteristics of the sampling basis, slow convergence rates or the Gibbs phenomenon.

An example of such artefacts can be seen in Figure 1. Here we have chosen $\mathcal{H} = L^2([0,1])$ and consider the Fourier sampling basis $\{(2\pi)^{-1/2} e^{2\pi i n} : n \in \mathbb{Z}\}$ and the Walsh sampling basis $\{w_n : n \in \mathbb{Z}_+ \coloneqq \{0, 1, \dots, \}\}$, where the w_n's are Walsh functions (see §3.1 for more on these functions, and their relation to Hadamard matrices). In the figure, we can see how the Walsh sampling basis gives a blocky approximation to the continuous hat functions, and how the Fourier sampling basis (no matter how large we choose N) always produce the very characteristic $\mathcal{O}(1)$ Gibbs oscillations around the discontinuity. This is because f_N only converges to f in the ℓ^2-norm, rather than the stronger uniform norm.

To resolve this issue, the idea of the aforementioned reconstruction techniques is to utilize prior knowledge on f, in order to compute a better approximation in the reconstruction space $\mathcal{R}_M$, using the samples $\{y_1, \dots, y_N\}$. Two examples of such improved reconstructions can be seen in Figure 2, where we have used different wavelets to approximate the functions f and g from Figure 1.

In this work, $\mathcal{R}_M$ is spanned by orthonormal wavelets and we consider $\mathcal{H} = L^2([0,1]^d)$ for $d = 1, 2$. This reconstruction space has several advantages:

(i) Orthonormal wavelets can be computed with any desired degree of smoothness, ranging from the discontinuous Haar wavelet to higher-order Daubechies wavelets or symlets. This means that we can tailor-make the smoothness of the reconstruction space.

(ii) In one dimension, orthonormal wavelets allow for optimal non-linear approximation of functions with bounded variation [5, Ch. 10] (see also [24]) and while wavelets are not provably optimal in two dimensions, their use and applicability in imaging is ubiquitous [33, 44, 53].

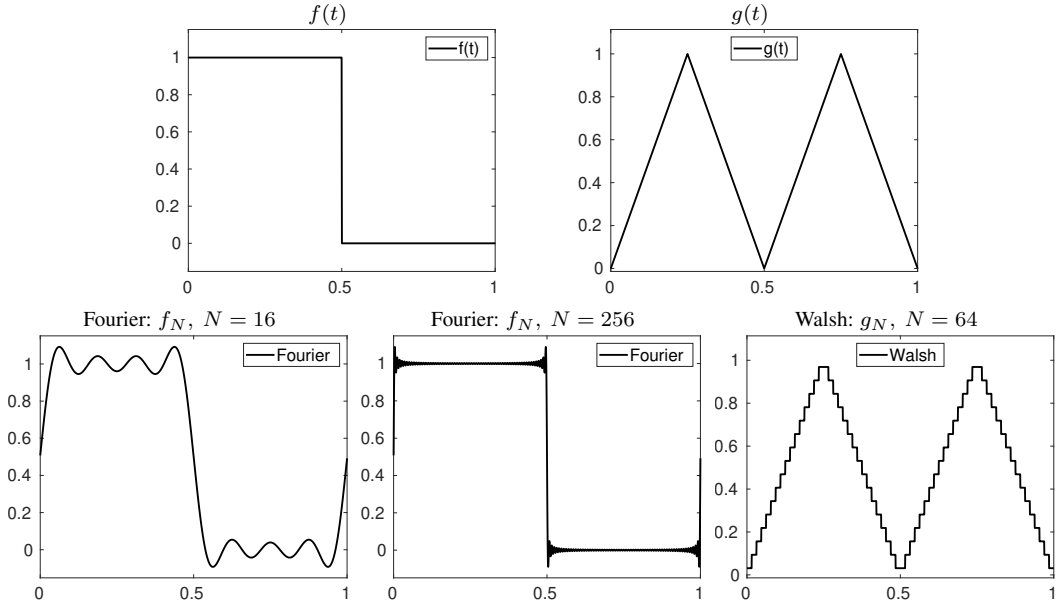

Figure 1: **(Undesirable artefacts)**. The two functions f and g (top row) are sampled using a Fourier and Walsh sampling basis, respectively. Given the acquired samples, we use the native truncated Fourier series f_N and truncated Walsh series g_N, known from (1), to approximate the functions. On the bottom row we show reconstructions f_N and g_N, for different values of N. Notice how the truncated Fourier series causes $\mathcal{O}(1)$ Gibbs oscillations around the discontinuity for every choice of N, and how the truncated Walsh series produce a reconstruction with blocky artefacts.

(iii) For Walsh sampling (considered here) and orthonormal wavelet reconstruction the so-called *stable sampling rate* (see Def. 2.2) is linear [35]. That is, to recover M wavelet coefficients using, e.g., generalised sampling, we require $N \geq CM$ Walsh samples, where $C \geq 1$ is a constant. We note that this rate is not necessarily linear for all reconstruction bases. For Fourier sampling and polynomial reconstruction, the requirement is quadratic in M, i.e., $N \geq CM^2$ samples are required [38]. For Walsh sampling and polynomial reconstruction, the stable sampling rate is not known.

1.1 Notation

Let $\ell^2(\mathbb{N})$ denote the usual set of square summable sequences, and let $\mathcal{B}(\ell^2(\mathbb{N}))$, denote the set of bounded linear operators between such sequences. For $\Omega \subseteq \{1, \dots, N\}$, we let $P_\Omega \colon \ell^2(\mathbb{N}) \to \ell^2(\mathbb{N})$ be the projection onto the coordinates indexed by Ω. That is, for $z \in \ell^2(\mathbb{N})$, $(P_\Omega z)_i = z_i$ if $i \in \Omega$, and 0 otherwise. Let $m = |\Omega|$. We sometimes abuse notation slightly and say that $P_\Omega \colon \ell^2(\mathbb{N}) \to \mathbb{C}^m$, by simply ignoring all the zero entries. Furthermore, if $\Omega = \{1, \dots, N\}$ we simply write P_N. Often we do not specify the domain and range of P_N, and let this be given by the context. Thus for an operator $U \in \mathcal{B}(\ell^2(\mathbb{N}))$, we write $P_N U P_M$ both to mean a finite dimensional $N \times M$ matrix and an operator in $\mathcal{B}(\ell^2(\mathbb{N}))$, depending on the context. When $P_M \colon \ell^2(\mathbb{N}) \to \mathbb{C}^M$, we have that $P_M^* \colon \mathbb{C}^M \to \ell^2(\mathbb{N})$, however, to unify the notation we still write $P_N U P_M$, rather than $P_N U P_M^*$.

Finally, for some closed subspace $\mathcal{V} \subset \mathcal{H}$ we let $P_\mathcal{V} \colon \mathcal{H} \to \mathcal{H}$ denote the projection onto $\mathcal{V}$.

1.2 Computing approximations in $\mathcal{R}_M$

For $f \in \mathcal{H}$, let $x_k = \langle f, r_k \rangle$ and $y_k = \langle f, s_k \rangle$ be the coefficients of f in the reconstruction basis and sampling basis, respectively. Let $x = \{x_k\}_{k \in \mathbb{N}}$ and $y = \{y_k\}_{k \in \mathbb{N}}$ and notice that $x, y \in \ell^2(\mathbb{N})$. The

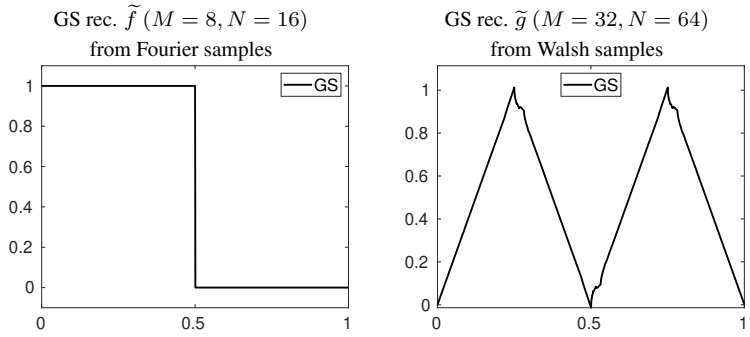

Figure 2: **(Improved reconstruction in $\mathcal{R}_M$).** Given $N = 16$ Fourier samples from the function f in Figure 1 and $N = 64$ Walsh samples from the function g in the same figure, we compute approximations $\widetilde{f}$ and $\widetilde{g}$, respectively, using generalised sampling (GS). Here we use a Haar wavelet basis with $M = 8$ functions for $\widetilde{f}$, and a Daubechies 2 (DB2) wavelet basis with $M = 32$ basis functions for the function $\widetilde{g}$. Note how increasing N in Figure 1 can not remedy the $\mathcal{O}(1)$ Gibbs oscillation for the discontinuous Haar scaling function f seen in the figure, whereas choosing a basis which spans this function enables us to capture f, using only $N = 16$ samples.

change-of-basis matrix $U \in \mathcal{B}(\ell^2(\mathbb{N}))$ between $\{r_k : k \in \mathbb{N}\}$ and $\{s_k : k \in \mathbb{N}\}$ is given by

$$U_{i,j} = \langle r_j, s_i \rangle, \quad \text{and} \quad y = Ux$$

where U is unitary, since both bases are orthonormal.

Given a finite set of (noiseless) samples, the previously mentioned reconstruction techniques compute an approximation to f by utilising the reconstruction space $\mathcal{R}_M$. We review three of the most modern approaches.

(i) (Generalised sampling). In generalised sampling [3, 7] one has access to the N samples $P_N y$ and using these we solve the least squares problem

$$\min_{z \in \mathbb{C}^M} \|P_N U P_M z - P_N y\|_{\ell^2}^2, \quad \text{where} \quad N \geq M. \tag{2}$$

Let $\widetilde{x} = \{\widetilde{x}_k\}_{k=1}^M$ be the minimiser of (2). In generalised sampling we approximate f with $\widetilde{f} = \widetilde{x}_1 r_1 + \cdots + \widetilde{x}_M r_M \in \mathcal{R}_M$. Moreover, the error committed by $\widetilde{f}$, is bounded by [7, Thm. 4.5]

$$\|f - \widetilde{f}\| \leq C_1 \|f - P_{\mathcal{R}_M} f\|, \tag{3}$$

where $C_1 > 0$ is a constant depending on the subspace angle between $\mathcal{S}_N$ and $\mathcal{R}_M$ (see §2 for details).

(ii) (PBDW-method). The PBDW-method [15, 48] is a data consistent method, which approximates f using the same N samples $P_N y$ as in generalised sampling. The approximation is computed as $\widehat{f} = P_{\mathcal{S}_N} f + P_{\mathcal{S}_N^\perp} \widetilde{f}$ where $\widetilde{f}$ is the generalised sampling approximation. As $\widehat{f} \in \mathcal{H}$ does not lie in a finite dimensional subspace, it can not be represented on a computer. We may, however, approximate $\widehat{f}$, by choosing some large $K > N$, and use the truncated sum $\widehat{f} \approx \sum_{k=1}^N y_k s_k + \sum_{k=N+1}^K (P_K U P_M \widetilde{x})_k s_k$ where $\widetilde{x}$ is the minimizer form (2). It was shown in [48] that the error committed by $\widehat{f}$ is bounded by

$$\|f - \widehat{f}\| \leq C_1 \|f - P_{\mathcal{R}_M \oplus (\mathcal{S}_N \cap \mathcal{R}_M^\perp)} f\|, \tag{4}$$

where C_1 is the same constant as in the generalised sampling error bound above.

(iii) (Infinite-dimensional compressive sensing). While the two methods above are linear reconstruction methods, compressive sensing (and more generally sparse regularization) is an example of a non-linear reconstruction method. In compressive sensing one computes an approximation in $\mathcal{R}_M$ using

$m < N$ samples. Let $\Omega \subset \{1, \dots, N\}$ have cardinality $m = |\Omega|$ and consider the measurements $P_\Omega y$. A standard way of computing a compressive sensing reconstruction is by solving the quadratically constrained basis pursuit optimisation problem

$$\min_{z \in \mathbb{C}^M} \|z\|_{\ell^1} \quad \text{subject to} \quad \|P_\Omega U P_M z - P_\Omega y\|_{\ell^2}^2 \leq \eta. \tag{5}$$

Here η is chosen so that $\eta \geq \|P_\Omega U P_M^\perp x\|_{\ell^2}^2$, to ensure that $P_M x$ is a feasible point. Given a minimizer $x^\sharp$ of (5), one approximates f with $f^\sharp = x_1^\sharp r_1 + \dots x_M^\sharp r_M \in \mathcal{R}_M$. Error bounds for compressive sensing reconstructions are probabilistic in nature and depend on the number of measurements m, and the bases $\{s_k\}_{k \in \mathbb{N}}$ and $\{r_k\}_{k \in \mathbb{N}}$ used. For concrete error bounds for Walsh sampling and wavelet reconstruction, we refer to [58] for non-uniform and [1] uniform recovery guarantees in infinite-dimensions. For a more general treatment of the subject, we refer to [5, 54].

1.3 Walsh sampling and its use in binary imaging

This work focuses on Walsh sampling (sometimes called Walsh-Hadamard, or just Hadamard sampling) due to its applicability in binary imaging. By binary imaging we mean any imaging application whose samples can be modelled with sampling functions whose range is binary (such as Walsh functions). The first practical example of such a modality was the single-pixel camera [25], which could capture images in the visible spectrum using a single pixel. Later the same principle of using a single sensor has been used successfully in fluorescence microscopy [55], lensless imaging [16] and holography [20]. For an in-depth treatment of binary imaging with a single sensor, see [31] for a recent review, or [5] for a mathematical treatment of the subject.

1.4 Contributions

In this article we let $\mathcal{H} = L^2([0,1]^d)$, $d = 1, 2$ and consider the recovery of orthonormal wavelet coefficients from Walsh samples. As outlined above, this setup has numerous applications in binary imaging. However, for any of the reconstruction methods mentioned above to work in practice, it is crucial to solve one of the optimisation problems (2) or (5). To do this, we need to form the matrix $P_N U P_M$ (potentially also $P_\Omega U P_M$), for different values of N and M. This can be computationally challenging since the entries of $P_N U P_M$ are given as the solution of MN integrals. Furthermore – ignoring the computational burden of computing these integral – using a densely stored matrix $P_N U P_M$ has several disadvantages.

(i) (Storage). In imaging applications it is not uncommon to have large dimensions, say $N = 512^2$ and $M = 256^2$. However, naively storing a dense matrix $P_N U P_M \in \mathbb{C}^{N \times M}$ with these dimensions requires approximately 137 GB of memory. This is substantially more than most workstations can handle.

(ii) (Computational complexity). When solving (2) or (5) iterative algorithms are often applied. For (2), the conjugate gradient method [36] is a popular choice, and for (5) SPGL1 [62] or Chambolle and Pock's primal-dual [18] algorithm are well-known choices. However, all of these algorithms rely on fast matrix-vector multiplications with $P_N U P_M$ or $P_\Omega U P_M$, and their adjoints. However, standard matrix-vector multiplication with a $N \times M$ matrix require $\mathcal{O}(MN)$ operations, and for large dimensions this cost can be substantial.

While some of these issues can be reduced in higher dimensions ($d > 1$) by considering tensor decompositions of the linear map $P_N U P_M$, none of these approaches can obtain a computational complexity of $\mathcal{O}(N \log N)$ and avoid storing the matrix $P_N U P_M$ altogether. In this work we do exactly this. We present

an algorithm which can compute matrix-vector multiplications with the matrix $P_N U P_M$ in $\mathcal{O}(N \log N)$[1] operations for Walsh sampling and orthonormal wavelet reconstruction in one and two dimensions without storing the matrix $P_N U P_M$. Applying the reconstruction methods outlined above allows for fast reconstruction of wavelet coefficients from Walsh samples with inconsiderable memory usage and low computational complexity.

The proposed algorithm extends the work of Gataric & Poon [30], who derives a similar algorithm with the same computational complexity for Fourier sampling and wavelet reconstruction. The two central ingredients in [30] allowing them to speed up the computation is (1) that the stable sampling rate (see §2) scales linearly for Fourier sampling and wavelet reconstruction and (2) that the translational variable m for scaling functions $\phi_{j,m}(x) = 2^{j/2}\phi(2^j x - m)$ can be disentangled from ϕ under the Fourier transform. That is, for a Fourier sampling function $e_\omega(x) = e^{2\pi \mathrm{i} \omega x}$ we have the equality $\langle \phi_{j,m}, e_\omega \rangle_{L^2[0,1]} = 2^{-j/2}\hat{\phi}(2^{-j}\omega)e_\omega(-2^{-j}m)$, where m is disentangled from the Fourier transform $\hat{\phi}$ of ϕ. As we shall see in Lemma 3.4, a similar splitting is possible under the Walsh transform. Moreover, it was proved in [35] that for Walsh sampling and wavelet reconstruction the stable sampling rate scales linearly.

In [30] the analysis is restricted to orthonormal wavelets on the interval with vanishing moments preserving wavelets at the boundaries [21]. While these boundary wavelets better approximate non-periodic function at the boundaries, their use in practice have been limited [11]. In this work we, therefore, carry out the analysis for both vanishing moments preserving boundary wavelets and periodic boundary wavelets [50, Sec. 7.5.1].

The paper is accompanied by a software implementation in MATLAB, demonstrating how this can be implemented in practice. It is a well known issue that MATLAB's implementation of the fast Walsh-Hadamard transform (FWHT), is extremely slow[2]. To mitigate this issue, the implementation also includes a MATLAB interface to the C++ library FXT (`https://www.jjj.de/fxt/`) [12], for speeding up this part of the code. Other time-critical parts of the code have also been written in C++ and interfaced with MATLAB. All accompanying code and data are accessible from `https://github.com/vegarant/cww` and `https://github.com/vegarant/fastwht`.

Remark 1.1 (Avoiding inverse crimes). Note that the proposed model avoids certain inverse crimes stemming from too early discretisation of the considered inverse problem. Indeed, by considering an infinite-dimensional model, we model measurements y_k that come from continuous integral transforms $y_k = \int_0^1 f(x)s_k(x)\,\mathrm{d}x$, rather than discrete inner-products. This model is motivated by the observation that most sensors do not compute pointwise samples of f, but rather integrate f over a short time or area [34, 40]. Discretising the problem at a too early stage using discrete inner products can result in measurement mismatch [19].

Remark 1.2 (Measurement noise). Above, we have focused on noiseless measurements to make the mathematical model clear. However, any realistic measurement model should also incorporate noisy measurements. Our overall goal in this manuscript is to develop an algorithm that can compute matrix-vector multiplications with the matrix $P_N U P_M$ in $\mathcal{O}(N \log N)$ operations. We will, therefore, not discuss noisy measurements in any detail. We refer to the literature on each of the specific reconstruction methods for further discussions on how the methods handle noisy measurements.

[1] Note that our bound here, is independent of M, but due to the stable sampling rate (see §2), we can take $N = 2^{dq}M$ for small values of q, usually $q \in \{1, 2, 3, 4\}$ (see Rem. 4.1).

[2] See https://ch.mathworks.com/matlabcentral/answers/395334-why-does-the-fwht-function-calculate-slower-than-the-fft-function-even-though-the-documentation-say

1.5 Outline of the paper

In §2 we define the subspace angle and the stable sampling rate, and we explain how these quantities dictate how we must choose N in relation to M to achieve stable and accurate reconstruction. This is followed by the definitions of the Walsh and wavelet sampling bases in §3, along with a key lemma used extensively in the derivation of the algorithm. We then describe the algorithm in one and two dimensions in §4 and §5, respectively, followed numerical examples in §6 and some concluding remarks in §7.

2 The subspace angle and the stable sampling rate

It is important to realize that stable and accurate recovery in $\mathcal{R}_M$, from samples $y_k = \langle f, s_k \rangle$, $k = 1, \dots N$, is not possible for arbitrary choices of bases $\{s_k : k \in \mathbb{N}\}$ and $\{r_k : k \in \mathbb{N}\}$. What is crucial for accurate and stable recovery in $\mathcal{R}_M$, is that the subspace angle between $\mathcal{S}_N$ and $\mathcal{R}_M$ is sufficiently small.

Definition 2.1 (Subspace angle)**.** Let $\mathcal{R}_M = \operatorname{span}\{r_1, \dots r_M\}$ and $\mathcal{S}_N = \operatorname{span}\{s_1, \dots, s_N\}$. The *subspace angle* $\omega \in [0, \pi/2]$ between $\mathcal{R}_M$ and $\mathcal{S}_N$ is

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

We set the reciprocal value as $\mu(\mathcal{R}_M, \mathcal{S}_N) \coloneqq \frac{1}{\cos(\omega(\mathcal{R}_M, \mathcal{S}_N))}$, and if $\cos(\omega(\mathcal{R}_M, \mathcal{S}_N)) = 0$, we set $\mu(\mathcal{R}_M, \mathcal{S}_N) = \infty$.

We note that a necessary condition for $\mu(\mathcal{R}_M, \mathcal{S}_N) < \infty$ is that $N \geq M$ (see e.g. [56, Thm. 2.1]). Furthermore, we have that $\mu(\mathcal{R}_M, \mathcal{S}_N)$ is related to the condition number of the matrix $P_M U^* P_N U P_M$, used for solving the normal equations in generalised sampling and the PBDW-method. Indeed, let $\sigma_1(A) \geq \cdots \geq \sigma_M(A)$ denote the ordered singular values of a matrix $A \in \mathbb{C}^{N \times M}$, with $N > M$. Then, using Parseval's identity, we have that

$$\cos(\omega(\mathcal{R}_M, \mathcal{S}_N)) \coloneqq \inf_{h \in \mathcal{R}_M, \|h\|=1} \|P_{\mathcal{S}_N} h\| = \inf_{z \in \mathbb{C}^M, \|z\|=1} \|P_N U P_M z\| = \sigma_M(P_N U P_M).$$

We also have that $\sigma_1(P_N U P_M) = \sup_{z \in \mathbb{C}^M, \|z\|=1} \|P_N U P_M z\| \leq 1$, since U is unitary, and hence the condition number

$$\operatorname{cond}(P_M U^* P_N U P_M) = \frac{\sigma_1^2(P_N U P_M)}{\sigma_M^2(P_N U P_M)} \leq \mu^2(\mathcal{R}_M, \mathcal{S}_N).$$

This directly relates to the numerical stability of the normal equations, used to solve the least-squares problem (2), and compute the generalised sampling and the PBDW-method's solution.

Furthermore, the accuracy of these two methods is also related to the subspace angle. Indeed, the constant C_1 fund in the error bounds (3) and (4) equals $C_1 = \mu(\mathcal{R}_M, \mathcal{S}_N)$. See [7, Thm. 4.5] and [15, Eq. (1.7)] (and [48] for earlier work). Thus both the numerical stability and accuracy of these two methods hinges on choosing $\mathcal{R}_M$ in relation to the samples one can acquire.

The situation is the same in infinite-dimensional compressive sensing, but the quantity $\mu(\mathcal{R}_M, \mathcal{S}_N)$, is camouflaged via the so-called *balancing property*, introduced in [4]. In infinite-dimensional compressive sensing, the balancing property typically governs the required number of samples needed to satisfy the restricted isometry property (RIP) [54], and its generalisations [1, 13, 59] for certain constants. These constants will again affect the constants found in the error bound for the minimiser $x^\sharp$ of (5), see, e.g., [54] for details. To see the relation between the subspace angle and the balancing property, we refer to the proof of Proposition 4.4 in [1].

From the above discussion, it is evident that the subspace angle between $\mathcal{R}_M$ and $\mathcal{S}_N$, affects both the accuracy and the stability of all the reconstruction methods. Thus, an important question is, therefore, how we should choose N in relation to M, to ensure that $\mu(\mathcal{R}_M, \mathcal{S}_N) \leq \gamma$ stays bounded. This relates to the so-called *stable sampling rate* [4, 7].

Definition 2.2 (Stable sampling rate)**.** Let $\mathcal{R}_M = \operatorname{span}\{r_1, \dots r_M\}$ and $\mathcal{S}_N = \operatorname{span}\{s_1, \dots, s_N\}$. The *stable sampling rate* for $M \in \mathbb{N}$ and $\gamma > 1$ is

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

For Walsh sampling and orthonormal wavelet reconstruction in $\mathcal{H} = L^2([0,1]^d)$, $d \geq 1$, it was shown by Hansen & Thesing [35] that the stable sampling rate scales linearly in M. That is, for a fixed $\gamma > 1$ there exist a constant $q_\gamma \geq 0$ such that whenever $N = 2^{d(r+q_\gamma)} \geq 2^{dr} = M$ for $r \in \mathbb{N}$, we have $\mu(\mathcal{R}_M, \mathcal{S}_N) \leq \gamma$. Hence for a fixed $q_\gamma > 0$, we get a fixed bound on $\mu(\mathcal{R}_M, \mathcal{S}_N)$, for all M and N on the form above.

This is important, since it tells us that for a fixed number of reconstruction coefficients M, we need no more than $N = CM$ samples, where $C = 2^{dq_\gamma}$ is a constant, to ensure that $\mu(\mathcal{R}_M, \mathcal{S}_N) \leq \gamma$. In Table 1 we have computed $1/\sigma_M(P_N U P_M) = \mu(\mathcal{R}_M, \mathcal{S}_N)$, for $N = 2^{dq}M$, for $d = 1, 2$ and $q = 1, 2, 3, 4$, for Walsh sampling and different wavelet reconstruction bases. From the table, we see that in all cases the choice $q = 1$ or $q = 2$ is sufficient to ensure that $1 < \gamma < 2$, indicating that the constant C is not necessarily very large for these bases.

3 The sampling and reconstruction spaces

This section introduces the necessary notation and background on the Walsh sampling basis and the orthonormal wavelet reconstruction bases. We also present a few useful results, needed to derive the final algorithm in later sections.

3.1 Walsh functions

Walsh functions (see [14] or [32] for an introduction) are closely related to dyadic representations of numbers. For an integer $n \in \mathbb{Z}_+ = \{0, 1, 2, \dots\}$ its dyadic series is $n = n^{(1)}2^0 + n^{(2)}2^1 + n^{(3)}2^2 + \cdots$, where the $n^{(j)}$'s are 0 or 1. Similarly for $x \in [0, 1)$ we can express its dyadic series as $x = x^{(1)}2^{-1} + x^{(2)}2^{-2} + x^{(3)}2^{-3} + \cdots$, for $x^{(j)} \in \{0, 1\}$. For rational numbers x, this expansion is not unique and in such cases we consider the expansion not ending with infinitely many repeating 1's.

There exist different orderings of Walsh functions, all of which leads to slightly different definitions. In this manuscript, we use the *sequency* ordered Walsh functions. This ordering has the advantage that the n'th Walsh function w_n has n sign changes.

Definition 3.1. Let $n \in \mathbb{Z}_+$ and $x \in [0, 1)$. The *Walsh function* $w_n \colon [0, 1) \to \{+1, -1\}$ is given by $w_n(x) \coloneqq (-1)^{\sum_{j=1}^{\infty}(n^{(j)}+n^{(j+1)})x^{(j)}}$

We note that $\{w_n : n \in \mathbb{Z}_+\}$ is an orthonormal basis for $L^2([0, 1])$, and we let

$$\mathcal{W}f(n) = \int_0^1 f(x) w_n(x) \, \mathrm{d}x$$

denote the Walsh transform of a function $f \in L^2([0, 1])$.

When working with Walsh functions, the XOR operation applied to binary sequences has many uses. We denote it by $\oplus$ and define it as follows.

Definition 3.2. Let $x = \{x^{(j)}\}_{j=1}^{\infty} \in \{0, 1\}^{\mathbb{N}}$ and $y = \{y^{(j)}\}_{j=1}^{\infty} \in \{0, 1\}^{\mathbb{N}}$ be binary sequences. The operation $\oplus$ applied to these sequences is given by $x \oplus y \coloneqq \{|x^{(j)} - y^{(j)}|\}_{j=1}^{\infty}$. For $x, y \in \mathbb{Z}_+$ or $x, y \in [0, 1)$, the operation $x \oplus y$ is understood in the sense of x and y's representation as binary sequences.

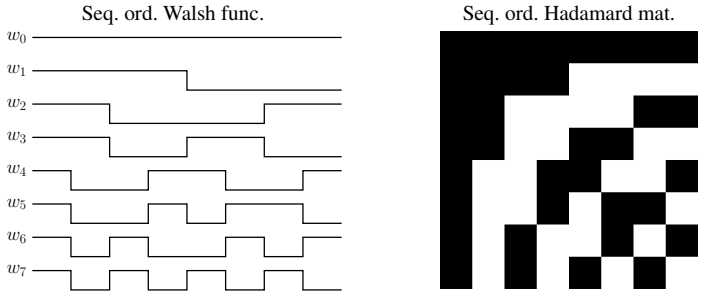

Figure 3: (**Relation between Walsh functions and Hadamard matrices**). Left: The eight first sequency ordered Walsh functions. Right: A 8×8 sequency ordered Hadamard matrix, where black corresponds to 1 and white to -1. We can see that the Walsh functions' sign changes correspond to the sign changes in the matrix.

Lemma 3.3. *For $x, y \in [0, 1)$, $n, j, l \in \mathbb{Z}_+$, the following three equalities hold*

$$w_n(x \oplus y) = w_n(x)w_n(y), \tag{6}$$

$$w_n(2^{-j}l) = w_l(2^{-j}n) \qquad \text{if } n, l < 2^j, \tag{7}$$

$$w_n(2^{-j}x) = w_{\lfloor n/2^j \rfloor}(x). \tag{8}$$

Proof. The two first equalities can be found in [32, Eq. 1.2.15 & §1.3]. The last equality follows from direct computations, see e.g., [1, Prop. 6.4] or [5, Lem. F.3].

□

We also note that Walsh functions and Hadamard matrices are closely related, since the (n, k)'th entry of a sequency ordered Hadamard matrix $H \in \mathbb{R}^{2^j \times 2^j}$ is given by $w_{n-1}(2^{-j}(k-1))$. See Figure 3 for an illustration of this relationship. Furthermore, for $N = 2^j$ we note that a matrix-vector product with H can be computed in $\mathcal{O}(N \log N)$ operations using the fast Walsh-Hadamard transform (FWHT) [14]. That is, for $x = \{x_k\}_{k=1}^N$, the N sums

$$\left\{ \sum_{k=1}^{N} w_n((k-1)/N)x_k \right\}_{n=0}^{N-1}$$

can utilize the FWHT algorithm to compute the result with $\mathcal{O}(N \log N)$ operations, and without storing the matrix H in memory. It is also worth noting that the FWHT algorithm can utilize parallel computations, something which reduces the time needed to execute the $\mathcal{O}(N \log N)$ operations.

3.2 Wavelets

Let $\phi\colon \mathbb{R} \to \mathbb{R}$ and $\psi\colon \mathbb{R} \to \mathbb{R}$ be a compactly supported orthonormal scaling function and wavelet [22], respectively, corresponding to an multiresolution analysis (MRA). We say that the wavelet ψ has ν vanishing moments if it is orthogonal to all polynomials of degree $\nu - 1$. That is, if $\langle x^k, \psi \rangle_{L^2(\mathbb{R})} = 0$ for $k = 0, \ldots, \nu - 1$. For simplicity, we work with wavelets with minimal support. Thus, for $\nu = 1$ the above wavelet is the Haar wavelet, but for $\nu \geq 2$ there are different choices, ranging from the classical *Daubechies wavelet* (which has minimum-phase) to *symlets* which are close to being symmetric, but with a larger phase [50, p. 294].

If ψ generates a system of orthonormal wavelets with ν vanishing moments and minimal support, then the support of ψ and ϕ is an interval of size $2\nu - 1$. For convenience, we use the convention that $\operatorname{supp}(\phi) = \operatorname{supp}(\psi) = [-\nu + 1, \nu]$.

Let $\phi_{j,m}(x) \coloneqq 2^{j/2}\phi(2^j x - m)$ and $\psi_{j,m}(x) \coloneqq 2^{j/2}\psi(2^j x - m)$ denote the dilated and translated versions of ϕ and ψ. To work on the interval $[0, 1]$, we need to construct bases on this interval consisting of

functions $\phi_{j,m}$ and $\psi_{j,m}$, with $j \geq J_0$ for some J_0, chosen so that $\text{supp}(\phi_{j,m}) = \text{supp}(\psi_{j,m}) \subset [0,1]$ for at least one choice of m. It is readily seen that if $J_0 \geq \lceil \log_2(2\nu) \rceil$ for $\nu \geq 2$ and $J_0 \geq 0$ for $\nu = 1$, then this holds for at least one m.

Constructing an orthonormal wavelet basis on the interval requires special care at the boundaries, and it is common to replace all wavelets and scaling functions intersecting the boundary with certain "replacement" functions. Hence for $j \geq J_0$ we define the set of functions

$$B_{\phi,j} = \{\phi_{j,m}^{\text{rep}}\}_{m=0}^{\nu-1} \bigcup \{\phi_{j,m}\}_{m=\nu}^{2^j-\nu-1} \bigcup \{\phi_{j,m}^{\text{rep}}\}_{m=2^j-\nu}^{2^j-1},$$
$$B_{\psi,j} = \{\psi_{j,m}^{\text{rep}}\}_{m=0}^{\nu-1} \bigcup \{\psi_{j,m}\}_{m=\nu}^{2^j-\nu-1} \bigcup \{\psi_{j,m}^{\text{rep}}\}_{m=2^j-\nu}^{2^j-1}$$

where $\psi_{j,m}^{\text{rep}}$ and $\phi_{j,m}^{\text{rep}}$ are replacement wavelets and scaling functions supported on $[0,1]$. There are several ways to construct these replacement functions so that they retain the orthonormality condition, and we consider both a periodic boundary extension and the vanishing moments preserving (VMP) boundary wavelets introduced by Cohen, Daubechies & Vial in [21].

The advantage of the former is that it is both easy to define and implement. Indeed, to compute a discrete wavelet transform (DWT) using a periodic boundary extension, one simply use a periodic convolutions between between the filters and the signal. The disadvantage of the periodic wavelets basis is that we lose the vanishing moments property at the boundaries. This may result in a few high amplitude coefficients at each scale. Another issue with these wavelets is that any ℓ^2-approximation of a non-periodic function on $[0,1]$ will have certain artefacts at the boundaries due to the underlying assumption of periodicity, see e.g., [50, Fig. 9.1] for typical artefacts at discontinuities.

The vanishing moments preserving boundary extension introduced in [21] circumvents this issue by designing special wavelets at the boundaries, which retain both orthonormality, vanishing moments and avoids any assumptions about periodicity. However, as pointed out by Antun & Ryan in [11], most wavelet libraries do not support these wavelets. In [30] Gataric & Poon extended the WaveLab library [17] to also include (Daubechies) vanishing moments preserving boundary wavelets with $\nu = 4, \ldots, 8$ vanishing moments. In this work, we use the implementation from [11], to also include orthonormal wavelets such as symlets.

For the periodic wavelet basis, we extend the wavelets and scaling functions at the boundaries periodically. That is, we let

$$\phi_{j,m}^{\text{per}} = \phi_{j,m}|_{[0,1]} + \phi_{j,2^j+m}|_{[0,1]} \qquad \text{for } m = 0, \ldots, \nu-1,$$
$$\phi_{j,m}^{\text{per}} = \phi_{j,m}|_{[0,1]} + \phi_{j,m-2^j}|_{[0,1]} \qquad \text{for } m = 2^j-\nu, \ldots, 2^j-1,$$

and similar for $\psi_{j,m}^{\text{per}}$. Here $|_{[a,b]}$ means the restriction to the interval $[a,b]$. Strictly speaking, we could have omitted the definition of $\phi_{j,\nu}^{\text{per}}$, $\psi_{j,\nu}^{\text{per}}$ and $\phi_{j,2^j-\nu}^{\text{per}}$, $\psi_{j,2^j-\nu}^{\text{per}}$, as these function are pure interior functions, but we define these functions to unify the notation with the vanishing moments preserving boundary wavelets.

In [21] one constructs special boundary wavelets and scaling functions ϕ_m^{left}, ψ_m^{left}, ϕ_m^{right}, and ψ_m^{rigth}, for $m = 0, \ldots, \nu-1$. These functions are created using finite linear combinations of the interior functions, and their supports are staggered. That is $\text{supp}(\phi_m^{\text{left}}) = [0, \nu+m]$ and $\text{supp}(\phi_m^{\text{right}}) = [-m-\nu, 0]$ and similar for ψ_m^{left} and ψ_m^{right}. The corresponding boundary functions (similar for the wavelets) are defined as

$$\phi_{j,m}^{\text{bd}}(x) = 2^{j/2}\phi_m^{\text{left}}(2^j x) \qquad \text{for } m = 0, \ldots, \nu-1,$$
$$\phi_{j,m}^{\text{bd}}(x) = 2^{j/2}\phi_{2^j-1-m}^{\text{right}}(2^j(x-1)) \qquad \text{for } m = 2^j-\nu, \ldots, 2^j-1.$$

With these functions well defined, we let "rep", mean either "per" or "bd".

Let $\mathcal{V}_j = \text{span}\{B_{\phi,j}\}$ and $\mathcal{U}_j = \text{span}\{B_{\psi,j}\}$, and note that by construction these satisfy $\mathcal{V}_j \oplus \mathcal{U}_j = \mathcal{V}_{j+1}$. Now, let $C_{\psi,j} = B_{\phi,J_0} \cup B_{\psi,J_0} \cup \cdots \cup B_{\psi,j-1}$. It should be clear from the previous discussion that

$B_{\phi,j}$ and $C_{\phi,j}$ span the same space. We can perform a change-of-basis between the two bases using a DWT matrix $W \in \mathbb{R}^{2^j \times 2^j}$.

Finally, note that no closed-form formula exists for the compactly supported orthonormal wavelets considered (except for the Haar wavelet). We can, however, compute approximations to $\phi(2^j k)$ and $\psi(2^j k)$, at dyadic grid points using the cascade algorithm [22].

3.3 A useful lemma

Before we proceed, we prove a lemma that lays the foundation for the fast computations derived in the following sections. We note that the lemma is a generalisation of what is used in the proof of Lemma 6.6 in [1].

Lemma 3.4. *Let $h \in L^2(\mathbb{R})$ with $\mathrm{supp}(h) \subset [a, b]$ for integers $a \leq 0 < b$. Denote by $h_{j,m}(x) = 2^{j/2} h(2^j x - m)$ a translated and dilated version of h. Suppose that $j, m \in \mathbb{Z}$ are chosen so that $\mathrm{supp}(h_{j,m}) \subset [0, 1]$. Then*

$$\langle h_{j,m}, w_n \rangle = 2^{-j/2} \sum_{l=a}^{b-1} w_n \left(\frac{l+m}{2^j} \right) \mathcal{W} h_{0,-l}|_{[0,1)} \left(\left\lfloor 2^{-j} n \right\rfloor \right).$$

Proof. First notice that by assumption we have that $\mathrm{supp}(h_{j,m}) \subset [2^{-j}(a+m), 2^{-j}(b+m)] \subset [0, 1]$. This implies that $b - a \leq 2^j$, and that $m \in \{-a, -a+1, \ldots, 2^j - b\} \subset \{0, \ldots, 2^j - 1\}$, where have used the assumption $a \leq 0 < b$, in the final inclusion. Next notice that for any $x \in [0, 1)$ we have the following equality

$$\begin{aligned} \frac{x}{2^j} + \frac{m}{2^j} &= \sum_{i=j}^{\infty} x^{(i-j+1)} 2^{-i-1} + \sum_{i=1}^{j} m^{(i)} 2^{-j-1+i} \\ &= \sum_{i=j}^{\infty} x^{(i-j+1)} 2^{-i-1} \oplus \sum_{i=1}^{j} m^{(i)} 2^{-j-1+i} = \frac{x}{2^j} \oplus \frac{m}{2^j}, \end{aligned} \tag{9}$$

where the second equality holds, since the support of the two numbers does not intersect when they are represented as binary sequences. Utilising (9) and Lemma 3.3 now give

$$\begin{aligned} \langle h_{j,m}, w_n \rangle &= \int_0^1 2^{j/2} h(2^j x - m) w_n(x) \, \mathrm{d}x \\ &= \sum_{l=a}^{b-1} \int_{2^{-j}(l+m)}^{2^{-j}(l+1+m)} 2^{j/2} h(2^j x - m) w_n(x) \, \mathrm{d}x \\ &= \sum_{l=a}^{b-1} \int_{l+m}^{l+1+m} 2^{-j/2} h\left(x - m\right) w_n\left(\frac{x}{2^j}\right) \, \mathrm{d}x \\ &= \sum_{l=a}^{b-1} \int_0^1 2^{-j/2} h\left(x + l\right) w_n\left(\frac{x+l+m}{2^j}\right) \, \mathrm{d}x \\ &= \sum_{l=a}^{b-1} \int_0^1 2^{-j/2} h\left(x + l\right) w_n\left(\frac{x}{2^j} \oplus \frac{l+m}{2^j}\right) \, \mathrm{d}x \\ &= 2^{-j/2} \sum_{l=a}^{b-1} w_n \left(\frac{l+m}{2^j} \right) \mathcal{W} h_{0,-l}|_{[0,1)} \left(\left\lfloor 2^{-j} n \right\rfloor \right). \end{aligned}$$

□

4 The one dimensional algorithm

Next, we describe an algorithm for computing a matrix-vector multiplication with the matrix

$$P_N U P_M = \begin{bmatrix} \langle \phi_{j,0}^{\text{rep}}, w_0 \rangle & \cdots & \langle \phi_{j,M-1}^{\text{rep}}, w_0 \rangle \\ \vdots & \ddots & \vdots \\ \langle \phi_{j,0}^{\text{rep}}, w_{N-1} \rangle & \cdots & \langle \phi_{j,M-1}^{\text{rep}}, w_{N-1} \rangle \end{bmatrix} \tag{10}$$

and its adjoint, using $\mathcal{O}(N \log N)$ operations and without explicitly storing the matrix (10) in memory. Throughout, we let $M = 2^j$ and $N = 2^{j+q}$ where $j \geq J_0$ and $q > 0$ are integers. Other values of M and N can be considered by ultilizing appropriate zero padding. Below, we describe the algorithm stepwise by defining different operators, which we combine to achieve the desired matrix-vector multiplication. The complete algorithm is summarised in Algorithm 1.

Remark 4.1 (On the scaling between N and M). In Table 1 we have computed the ratio between the largest and smallest singular value of the matrix $P_N U P_M$, for different wavelets and choices for q, both in one and two dimensions. We observe that in all cases, the matrix is well-conditioned for the simplest choice of $q = 1$. This corresponds to $N = 2M$ in one dimension and $N = 4M$ in two dimensions. Moreover, since we know that the stable sampling rate for Walsh sampling and wavelet reconstruction is linear, we have that $N = \mathcal{O}(M)$, with a reasonable constant.

Remark 4.2 (Applications to compressive sensing). Note that a sparse representation of f is needed for compressive sensing to achieve successful recovery. For this method it is, therefore, better to represent an approximation to f in the basis $C_{\psi,j}$, than the $B_{\phi,j}$ basis used above. Changing the basis can easily be achieved by using the matrix $P_N U P_M W^{-1}$, where $P_N U P_M$ is as above, and $W^{-1} \in \mathbb{C}^{M \times M}$ is the inverse discrete wavelet transform (IDWT). As W^{-1} is a change of basis matrix from $C_{\psi,j}$ to $B_{\phi,j}$, this matrix will simulate the desired matrix if $C_{\psi,j}$ is the reconstruction basis. Furthermore, the cost of applying W^{-1} is $\mathcal{O}(M)$ using the cascade algorithm. This means that the overall cost of the matrix-vector multiplication does not grow by applying this change-of-basis.

Remark 4.3 (Haar wavelet reconstruciton). For $N = M = 2^j$, the Haar wavelet basis and Walsh sampling basis, span the same space. For the Haar reconstruction basis there is, therefore, no benefit of applying generalised sampling or the PBDW-method for reconstruction. Compressive sensing, on the other hand, can be applied since it allows for reconstruction of M wavelet coefficients from $m < M$ samples, under the assumption of sparsity. Since many natural images are sparse in the Haar wavelet basis, this approach is widely studied, see e.g. [9, 51, 57]. For Walsh sampling and Haar wavelet reconstruction using the basis $C_{\psi,j}$, the truncated change-of-basis matrix $P_M U P_M = H W^{-1}$, where W^{-1} is the Haar IDWT matrix, and H is the Hadamard matrix. This matrix can be computed using fast transforms with the FWHT and DWT. Below we do, therefore, not consider Haar wavelet reconstruction.

4.1 The forward operator

The wavelet basis $B_{\phi,j}$ with $\nu > 1$ vanishing moments consists of three types of wavelets, the left boundary corrected wavelets, interior wavelets and the right boundary corrected wavelets. The matrix-vector multiplication $P_N U P_M \xi$ for $\xi \in \mathbb{C}^M$ is, therefore, naturally divided into the three sums

$$\sum_{m=0}^{\nu-1} \langle \phi_{j,m}^{\text{rep}}, w_n \rangle \xi_m + \sum_{m=\nu}^{M-\nu-1} \langle \phi_{j,m}, w_n \rangle \xi_m + \sum_{m=M-\nu}^{M-1} \langle \phi_{j,m}^{\text{rep}}, w_n \rangle \xi_m \tag{11}$$

for each $0 \leq n < N$. In this subsection we foucs on how to speed up the computations of the middle summand. Throughout we take ν to be some small fixed number, usually in the range $\{2, \ldots, 8\}$, and we

The value of $\mu(\mathcal{R}_M, \mathcal{S}_N) = 1/\cos(\omega(\mathcal{R}_M, \mathcal{S}_N)) = 1/\sigma_M(P_N U P_M)$

One dimension

$M = 2^7$, $N = 2^{7+q}$, $L^2([0,1])$

Wavelet	$q=1$	$q=2$	$q=3$	$q=4$
DB2	1.200	1.050	1.014	1.004
DB3	2.610	1.135	1.028	1.006
DB4	1.251	1.068	1.023	1.007
DB5	1.392	1.109	1.025	1.011
DB6	6.499	1.137	1.033	1.016
DB7	2.642	1.104	1.046	1.022
sym2	1.200	1.050	1.014	1.004
sym3	2.610	1.135	1.028	1.006
sym4	1.188	1.037	1.008	1.003
sym5	1.179	1.042	1.013	1.005
sym6	1.300	1.059	1.015	1.005
sym7	1.549	1.095	1.021	1.008

Two dimensions

$M = 2^{2\cdot 5}$, $N = 2^{2(5+q)}$, $L^2([0,1]^2)$

Wavelet	$q=1$	$q=2$	$q=3$	$q=4$
DB2	1.439	1.102	1.028	1.008
DB3	6.814	1.289	1.057	1.013
DB4	1.565	1.141	1.047	1.013
DB5	1.937	1.230	1.050	1.022
DB6	42.233	1.292	1.068	1.032
DB7	6.980	1.220	1.094	1.044
sym2	1.439	1.102	1.028	1.008
sym3	6.814	1.289	1.057	1.013
sym4	1.412	1.075	1.016	1.006
sym5	1.389	1.085	1.026	1.009
sym6	1.690	1.121	1.029	1.009
sym7	2.400	1.199	1.043	1.015

Table 1: We compute the fraction $1/\sigma_M(P_N U P_M) = \mu(\mathcal{R}_M, \mathcal{S}_N)$, for the matrix $P_N U P_M$, where U is the change-of-basis matrix between a Walsh sampling basis and an orthonormal wavelet basis with vanishing moments preserving boundary wavelets. We consider both one and two-dimensional bases. We see that in all the considered cases, the smallest singular value is not close to zero, indicating good conditioning of the matrix. We also see that the two-dimensional values equals the square of the one-dimensional values. This is expected since the two-dimensional matrix can be formed as the Kronecker product of two one-dimensional matrices. Here DBX and sym$X$, refer to a Daubechies or symlet wavelet, respectively, with X vanishing moments.

omit the dependence on ν whenever we summarize the computational cost of the algorithm. The first and third summand require $\mathcal{O}(N\nu) = \mathcal{O}(N)$ operations each, and their dependence is therefore independent of M. We consider the edge scaling functions in §4.3.

We start by applying Lemma 3.4 to the middle summand in (11). This gives

$$\begin{aligned}\sum_{m=\nu}^{M-\nu-1} \langle \phi_{j,m}, w_n \rangle \xi_m &= 2^{-j/2} \sum_{m=\nu}^{M-\nu-1} \sum_{l=-\nu+1}^{\nu-1} \mathcal{W}\phi_{0,-l}|_{[0,1)}(\lfloor 2^{-j} n \rfloor) w_n\left(\tfrac{l+m}{2^j}\right) \xi_m \\ &= 2^{-j/2} \sum_{l=-\nu+1}^{\nu-1} \mathcal{W}\phi_{0,-l}|_{[0,1)}(\lfloor 2^{-j} n \rfloor) \sum_{m=\nu}^{M-\nu-1} w_n\left(\tfrac{l+m}{2^j}\right) \xi_m.\end{aligned}$$

Recall that $M = 2^j$ and $N = 2^{j+q}$, and define the linear operator $H_l \colon \mathbb{R}^M \to \mathbb{R}^N$ by

$$H_l(\xi) = \left[\sum_{m=\nu}^{M-\nu-1} w_n\left(\frac{2^q(l+m)}{N}\right) \xi_m \right]_{n=0}^{N-1}, \quad \xi \in \mathbb{R}^M,$$

and the linear operator $D_l \colon \mathbb{R}^N \to \mathbb{R}^N$ by

$$D_l(\alpha) = \left[2^{-j/2} \mathcal{W}\phi_{0,-l}|_{[0,1)}(\lfloor 2^{-j} n \rfloor) \alpha_n \right]_{n=0}^{N-1}, \quad \alpha \in \mathbb{R}^N.$$

Combining these operators, we can write the middle sum in (11) as

$$\left[\sum_{m=\nu}^{M-\nu-1} \langle \phi_{j,m}, w_n \rangle \xi_m \right]_{n=0}^{N-1} = \sum_{l=-\nu+1}^{\nu-1} D_l(H_l(\xi)). \tag{12}$$

Note that H_l can be implemented by embedding $\xi \in \mathbb{R}^M$ in a zero-padded vector of length N, and apply an $N \times N$ fast Walsh-Hadamard transform. Thus, evaluatning H_l can be done in $\mathcal{O}(N \log N)$ operations.

Also notice that the coefficients $\mathcal{W}\phi_{0,-l}|_{[0,1)}(\lfloor 2^{-j}n \rfloor)$ are independent of the input, and can be computed a priori. This reduces the cost of evaluating D_l to at most $\mathcal{O}(N)$ operations. The cost of computing (12) is, therefore, $\mathcal{O}(N \log N)$.

Also note that $n < N = 2^{j+q}$, which implies that $\lfloor 2^{-j}n \rfloor \leq 2^q - 1$. This means that for each l we only need compute $\mathcal{W}\phi_{0,-l}(s)|_{[0,1)}$ for $s = 0, \ldots, 2^q - 1$. Furthermore, from §2 we know that for a fixed $\gamma > 1$ the stable sampling rate scales linearly. Hence for fixed q we may vary j without affecting the stable sampling rate. This means that we only need to precompute these coefficients for some $q > q_{\gamma'}$ where $\gamma' > 1$ is the smallest stable sampling rate of interest. Moreover, from Table 1 we see that even the simplest choice of $q = 1$, results in $1 < \gamma < 2$ in many cases.

4.2 The adjoint operator

Next we consider the matrix-vector multiplication $P_M U^* P_N \alpha$ for $\alpha \in \mathbb{C}^N$. Since the computational burden is on the $M - 2\nu$ middle columns, we once more foucs on these and prostpone the edge wavelet functions until §4.3. That is, for $m = \nu, \ldots, M - \nu - 1$ we can write the matrix-vector product as

$$\begin{aligned} \sum_{n=0}^{N-1} \langle w_n, \phi_{j,m} \rangle \, \alpha_n &= \frac{1}{\sqrt{2^j}} \sum_{n=0}^{N-1} \sum_{l=-\nu+1}^{\nu-1} w_n \left(\frac{l+m}{2^j} \right) \mathcal{W}\phi_{0,-l}|_{[0,1)} \left(\left\lfloor \frac{n}{2^j} \right\rfloor \right) \alpha_n \\ &= \frac{1}{\sqrt{2^j}} \sum_{l=-\nu+1}^{\nu-1} \sum_{n=0}^{N-1} w_{2^q(l+m)} \left(\frac{n}{N} \right) \mathcal{W}\phi_{0,-l}|_{[0,1)} \left(\left\lfloor \frac{n}{2^j} \right\rfloor \right) \alpha_n \end{aligned} \tag{13}$$

by utilizing Lemma 3.4 and (7). Next define the operator $B_l \colon \mathbb{R}^N \to \mathbb{R}^M$ as

$$B_l(s) = \begin{cases} \sum_{n=0}^{N-1} w_{2^q(l+m)}(n/N) s_n & \text{for } m = \nu, \ldots, M - \nu - 1 \\ 0 & \text{for } m \in \{0, \ldots, \nu - 1\} \cup \{M - \nu, \ldots, M - 1\} \end{cases}$$

for $s \in \mathbb{R}^N$, and observe that $B_l = H_l^*$. Thus, from Equation (13) we now have

$$\left[\sum_{n=0}^{N_1} \langle w_n, \phi_{j,m} \rangle \, \alpha_n \right]_{m=\nu}^{M-\nu-1} = \sum_{l=-\nu+1}^{\nu-1} B_l(D_l(\alpha)).$$

Finally, observe that B_l can be computed in $\mathcal{O}(N \log N)$ operations by applying a fast Walsh Hadamard transform of dimension N and selecting the appropriate output from this transform. Since the cost of applying D_l is $\mathcal{O}(N)$, the total cost of computing the output from the middle rows are of order $\mathcal{O}(N \log N)$.

4.3 The edge operations

We now turn to the edge functions and consider the two boundary extensions given by the periodic and vanishing moments preserving boundary wavelets. This gives us four different edge inner products, one for each edge and boundary extension.

Consider the forward operator. According to (11) we can write the sum of the ν first and ν last columns as

$$\left[\sum_{m=0}^{\nu-1} \langle \phi_{j,m}^{\text{rep}}, w_n \rangle \xi_m \right]_{n=0}^{N-1} \quad \text{and} \quad \left[\sum_{m=M-\nu}^{M-1} \langle \phi_{j,m}^{\text{rep}}, w_n \rangle \xi_m \right]_{n=0}^{N-1}, \tag{14}$$

respectively, for $\xi \in \mathbb{R}^M$. Likewise for the adjoint operator we can write the ν first and ν last rows as

$$\left[\sum_{n=0}^{N-1} \langle w_n, \phi_{j,m}^{\text{rep}} \rangle \alpha_n \right]_{m=0}^{\nu-1} \quad \text{and} \quad \left[\sum_{n=0}^{N-1} \langle w_n, \phi_{j,m}^{\text{rep}} \rangle \alpha_n \right]_{m=M-\nu}^{M-1}, \tag{15}$$

respectively, for $\alpha \in \mathbb{R}^N$.

At the edges, we could, – potentially – compute the inner products $\langle \phi^{\text{rep}}_{j,m}, w_n \rangle$ a priori and store the result as dense matrices. A challenge with this approach, is that we need to compute and store the inner products for every possible combination of j, m and n. This is infeasible in general, and would only allows us to do computations for certain dimensions. However, by applying Lemma 3.4 once more, we can disentangle j, m, n from the integral computation so that we only need to compute $\mathcal{W}\phi^{\text{rep}}_{0,-l}|_{[0,1]}(s)$ for $s \in \{0, \dots, 2^q - 1\}$. In the next proposition we do just this. Note that we use the convention that if $b < a$ and we write $\sum_{l=a}^{b}(\cdots)$, then this should be interpreted as zero.

Proposition 4.4. *Let ϕ be a scaling function, whose wavelet has $\nu > 1$ vanishing moments. Let $M = 2^j$ and $N = 2^{j+q}$ for positive integers $j \geq \lceil \log_2(2\nu) \rceil$ and $q > 0$. Let $n \in \{0, \dots, N-1\}$. Then for $m = 0, \dots, \nu - 1$,*

$$\begin{aligned}\langle \phi^{per}_{j,m}, w_n \rangle = &\sum_{l=-\nu+1}^{-m-1} 2^{-j/2} w_n \left(\frac{2^j + m + l}{2^j} \right) \mathcal{W}\phi_{0,-l}|_{[0,1)}\Big(\Big\lfloor \frac{n}{2^j} \Big\rfloor \Big) \\ &+ \sum_{l=-m}^{\nu-1} 2^{-j/2} w_n \left(\frac{m + l}{2^j} \right) \mathcal{W}\phi_{0,-l}|_{[0,1)}\Big(\Big\lfloor \frac{n}{2^j} \Big\rfloor \Big)\end{aligned}$$

and

$$\langle \phi^{bd}_{j,m}, w_n \rangle = \sum_{l=0}^{\nu-1+m} 2^{-j/2} w_n \left(\frac{l}{2^j} \right) \mathcal{W}\phi^{left}_{m}(\cdot + l)|_{[0,1)} \left(\Big\lfloor \frac{n}{2^j} \Big\rfloor \right).$$

Furthermore, for $m = 2^j - \nu, \dots, 2^j - 1$,

$$\begin{aligned}\langle \phi^{per}_{j,m}, w_n \rangle = &\sum_{l=-\nu+1}^{2^j-m-1} 2^{-j/2} w_n \left(\frac{l + m}{2^j} \right) \mathcal{W}\phi_{0,-l}|_{[0,1)}\Big(\Big\lfloor \frac{n}{2^j} \Big\rfloor \Big) \\ &+ \sum_{l=2^j-m}^{\nu-1} 2^{-j/2} w_n \left(\frac{l + m - 2^j}{2^j} \right) \mathcal{W}\phi_{0,-l}|_{[0,1)}\Big(\Big\lfloor \frac{n}{2^j} \Big\rfloor \Big),\end{aligned}$$

and

$$\langle \phi^{bd}_{j,m}, w_n \rangle = \sum_{l=m-2^j-\nu+1}^{-1} 2^{j/2} w_n \left(\frac{l + 2^j}{2^j} \right) \mathcal{W}\phi^{right}_{2^j-1-m}(\cdot + l)|_{[0,1)} \left(\Big\lfloor \frac{n}{2^j} \Big\rfloor \right).$$

Proof. For an interval $I \subset \mathbb{R}$, let χ_I denote the characteristic funciton on I. The result follows by using Lemma 3.4 on all the considered inner products. For all functions intersecting the left edge this is trivial, the result follows by recalling that $\text{supp}(\phi) = [-\nu + 1, \nu]$ and $\text{supp}(\phi^{\text{left}}_m) = [0, \nu + m]$. The same can be said, about the functions $\phi^{\text{per}}_{j,\nu-1} = \phi_{j,\nu-1}$, and $\phi^{\text{per}}_{j,2^j-\nu} = \phi_{j,2^j-\nu}$, since these are interior functions. Applying Lemma 3.4 to the right edges require slighly more care, since it is assumed that the function under consideration is supported on an interval $[a, b]$, with $a, b \in \mathbb{N}$, and $b > 0$. On the right edges this can be achived by using the change of variable $y = 2^j x - (2^j - 1)$. We do not write out the details for all the considered functions, but demonstrate the idea on $\phi_{j,2^j+m}|_{[0,1]}$, for $m = 0, \dots, \nu - 2$ (used in $\phi^{\text{per}}_{j,m}$). We start by noticing that $\chi_{[0,1]}(x) = \chi_{[-(2^j-1),1]}(2^j x - (2^j - 1))$. This means that

$$\phi_{j,2^j+m}|_{[0,1]}(x) = 2^{j/2}\phi(2^j x - (2^j - 1) - (m + 1))\chi_{[-(2^j-1),1]}(2^j x - (2^j - 1)),$$

where the function $\phi(\cdot - (m + 1))\chi_{[-(2^j-1),1]}$ has support $[-\nu + 1 + (m + 1), 1]$. Applying Lemma 3.4, and using that $\mathcal{W}\phi_{0,-l}(\cdot - (m + 1))|_{[0,1)}(s) = \mathcal{W}\phi_{0,-l+m+1}|_{[0,1)}(s)$ gives the result. □

Given the inner products $\langle \phi^{\text{rep}}_{j,m}, w_n \rangle$ and $\langle w_n, \phi^{\text{rep}}_{j,m} \rangle$, the computational cost of (14) and (15), is $\mathcal{O}(N)$. Furthermore, to compute these inner products we may use Proposition 4.4 for each $n \in \{0, \dots, N-1\}$.

However, this can can be challenging since evaluating the above sums, require the computation of $w_n(2^{-j}(m+l))$ for many different choices of m, l and n, and – to the best of the author's knowledge – no software package implements the pointwise evaluation of Walsh functions. Moreover, a naive implementation in C++ using Definition 3.1 is rather slow. To speed up this part of the code we use the relation between Walsh functions and Hadamard matrices, and use the FWHT algorithm to evaluate $w_n(2^{-j}(m+l))$ for all the relevant values of m, l and n, simultaneously. However, this raises the computational cost of the edge computations to $\mathcal{O}(N \log N)$.

Algorithm 1 The one-dimensional forward and adjoint operator.

procedure THE FORWARD OPERATOR
- **Input:** $j, q \in \mathbb{N}$. $M = 2^j,\ N = 2^{j+q}$ and $\xi \in \mathbb{R}^M$.
- **Output:** $\alpha = P_N U P_M \xi$ where $P_N U P_M$ is given by (10).
- Compute vectors $\beta^{\text{left}} \in \mathbb{R}^N$ and $\beta^{\text{right}} \in \mathbb{R}^N$ from (14), using Prop. 4.4.
- Compute the vector $\beta^{\text{mid}} = \sum_{l=-\nu+1}^{\nu-1} D_l H_l(\xi)$.
- **return** $\alpha = \beta^{\text{left}} + \beta^{\text{mid}} + \beta^{\text{right}}$.

procedure THE ADJOINT OPERATOR
- **Input:** $j, q \in \mathbb{N}$. $M = 2^j,\ N = 2^{j+q}$ and $\alpha \in \mathbb{R}^N$.
- **Output:** $\xi = P_M U^* P_N \alpha$ where $P_N U P_M$ is given by (10).
- Compute $\xi_0, \dots, \xi_{\nu-1}$ and $\xi_{M-\nu}, \dots \xi_{M-1}$ from (15), using Prop. 4.4.
- Compute $\xi'_\nu, \dots, \xi'_{M-\nu-1}$ extracting elements from $\xi' = \sum_{l=-\nu+1}^{\nu-1} B_l(D_l(\alpha))$.
- **return** $\xi = [\xi_0, \dots, \xi_{\nu-1}, \xi'_\nu, \dots, \xi'_{M-\nu-1}, \xi_{M-\nu}, \dots, \xi_{M-1}]$

5 Extension to two dimensions

We restrict our attention to $d = 2$ dimensions since it applies to any kind of imaging application. It is certainly possible to extend the algorithm to any d-dimensional tensor product space, though its practical relevance seems limited. We let $\mathcal{H} = L^2([0,1]^2)$ and consider samples from the tensor product basis $\{w_{n_1} \otimes w_{n_2} : (n_1, n_2) \in \mathbb{Z}_+^2\}$. As for the one dimensional algorithm, we consider the case where the sampling and reconstruction spaces are dyadic cubes. That is, for $N = 2^{j+q}$ and $M = 2^j$ we let the sampling space $\mathcal{S}_{N^2} = \{w_{n_1} \otimes w_{n_2} : 0 \le n_1, n_2 < N\}$ and the reconstruction space $\mathcal{R}_{M^2} = \{\phi_{j,m_1} \otimes \phi_{j,m_2} : 0 \le m_1, m_2 < M\}$.

For a tensor $\xi \in \mathbb{R}^{M \times M}$ we can split the change-of-basis computation as

$$\begin{aligned} \alpha_{n_1,n_2} &= \sum_{m_1=0}^{M-1} \sum_{m_2=0}^{M-1} \xi_{m_1,m_2} \langle \phi_{j,m_1} \otimes \phi_{j,m_2}, w_{n_1} \otimes w_{n_2} \rangle \\ &= \sum_{m_1=0}^{M-1} \langle \phi_{j,m_1}, w_{n_1} \rangle \sum_{m_2=0}^{M-1} \xi_{m_1,m_2} \langle \phi_{j,m_2}, w_{n_2} \rangle \end{aligned} \tag{16}$$

for each $n_1, n_2 \in \{0, \dots, N-1\}$, so that it is a double sum of one-dimensional inner products. Thus, letting $G \in \mathbb{R}^{N \times M}$ denote the forward operator we derived for the one dimensional case, and letting $\eta \in \mathbb{C}^{M \times N}$ have components

$$\eta_{m_1,n_2} = \sum_{m_2=0}^{M-1} \xi_{m_1,m_2} \langle \phi_{j,m_2}, w_{n_2} \rangle = G\left([\xi_{m_1,m_2}]_{m_2=0}^{M-1} \right)$$

we see that the above computation simplifies to

$$\alpha_{n_1,n_2} = \left(G\left([\eta_{m_1,n_2}]_{m_1=0}^{M-1} \right) \right)_{n_1}, \quad \text{for } n_1, n_2 \in \{0, \dots, N-1\},$$

or simply $\alpha = G\xi G^*$ for $\xi \in \mathbb{R}^{M\times M}$.

Now, since G can be evaluated in $\mathcal{O}(N \log N)$ operations, we can compute $\alpha = G\xi G^*$ in $\mathcal{O}(2MN \log N)$ operations. Furthermore, since $N = 2^{2q}M$, and $q \in \{1, 2\}$ is a resonable choice, this is reduces to $\mathcal{O}(N^2 \log N^2)$, where N^2 is the dimension of the sampling space.

By considering (16), it should be clear that we can do the same type of splitting also for the adjoint operator. We do not do the full derivation, but notice that as an intermediate step one would need to compute

$$\beta_{n_1,m_2} = \sum_{n_2=0}^{N-1} \alpha_{n_1,n_2} \langle w_{n_2}, \phi_{j,m_2} \rangle = G^* \left([\alpha_{n_1,n_2}]_{n_2=0}^{N-1} \right)$$

for $0 \leq m_2 < M$. Applying the same transform in the row direction, leads to a transform which can be computed in $\mathcal{O}(N^2 \log N^2)$ operations. The complete algorithm is summarized in Algorithm 2.

Algorithm 2 The two dimensional forward and adjoint operator.

Let $G \in \mathbb{R}^{N\times M}$ be the one dimensional truncated change-of-basis matrix (10).
procedure THE FORWARD OPERATOR
    Input: $j, q \in \mathbb{N}$. $M = 2^j,\ N = 2^{j+q}$ and $\xi \in \mathbb{R}^{M\times M}$.
    Compute $\alpha = G\xi G^*$.
    return α.
procedure THE ADJOINT OPERATOR
    Input: $j, q \in \mathbb{N}$. $M = 2^j,\ N = 2^{j+q}$ and $\alpha \in \mathbb{R}^{N\times N}$.
    Compute $\xi = G^*\alpha G$.
    return ξ.

6 Numerical examples

We conclude with two numerical examples.

Example 1 (Computational cost) The proposed algorithm requires $\mathcal{O}(N \log N)$ floating-point operations to compute a matrix-vector multiplication with the $N \times M$ matrix in (10). This is in contrast to the $\mathcal{O}(NM)$ operations normally required for standard dense matrix-vector multiplication. In two dimensions with tensor product bases, as considered here, it is customary to compute the change of basis as $\xi \mapsto G\xi G^*$ for $\xi \in \mathbb{R}^{m\times m}$ and $G \in \mathbb{R}^{n\times m}$, rather than forming the Kronecker product matrix $G \otimes G$. In this case the computational cost, both for the forward and adjoint operator, reduces to $\mathcal{O}(N^{3/2})$, where $N = n \times n$ and $n = \mathcal{O}(m)$.

In this example, we want to investigate if the savings offered by this new algorithm can be seen in practice. To this end, we measured the computational time of both the forward and adjoint operator in one and two dimensions, using the proposed algorithm and dense matrices. To get an accurate estimate of the computational complexity of the algorithms, all computations were executed on a single core on a central processing unit (CPU), to prevent the use of parallel computations.

The results can be seen in Figure 4. In all experiments we choose $N = 2^{d(R+k+1)}$ and $M = 2^{d(R+k)}$ for different values of R, k and d. According to our estimates above, the computational time of the proposed algorithm should scale roughly (ignoring the log factor) like $2^{dk} \times \text{Constant}$, when we increase k for these choices of N and M. On the other hand, for dense matrix-vector multiplication we expect that the time scales like $4^k \times \text{Constant}$ in one dimension and $8^k \times \text{Constant}$ in two dimensions, for these choices of N and M. As we can see from the figure, these estimates agree well for all of the algorithms. It is, however, necessary to choose a large R to observe the asymptotic behaviour for the proposed algorithm. This is most

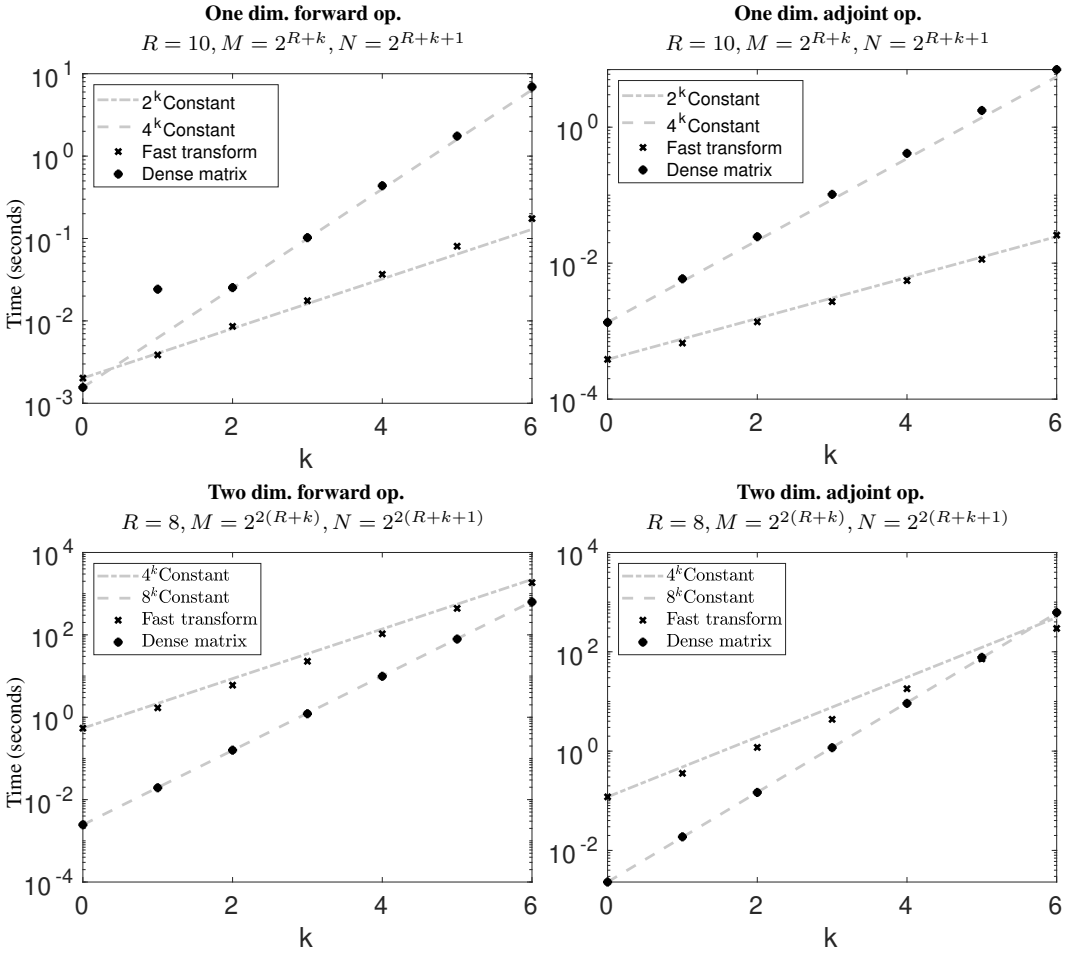

Figure 4: (**Comparison of runtime**). Given input of increasing size, we compare the proposed algorithm in one and two dimensions with dense matrix-vector multiplication. In all experiments we use a periodic DB4 wavelet reconstruction basis, and all experiments are computed on a single CPU core. The two-dimensional dense matrix-vector multiplication is computed using a tensor product decomposition, e.g., for the forward operator $\xi \mapsto G\xi G$ for $\xi \in \mathbb{R}^{m\times m}$ and $G \in \mathbb{R}^{n\times m}$, with $N = n \times n$ and $M = m \times m$.

likely related to the fact that MATLAB's dense matrix-vector multiplication is a highly optimized procedure, written in a compiled language, whereas the proposed algorithm is implemented in MATLAB, something which gives it a substantial constant overhead.

Example 2 (Two-dimensional reconstruction) We conclude with an example showing how the proposed algorithm can be used for two-dimensional problems. In this example we let $f(t_1, t_2) = \cos\left(\frac{3}{2}\pi t_1\right)\sin\left(3\pi t_2\right)$, and we acquire the first 32×32 Walsh samples form f. Using these samples, we compute a truncated Walsh series approximation to f as well as a generalised sampling reconstruction with different wavelet smoothness. Note that the smoothness of the wavelet basis increases with ν and since f is smooth we expect that the reconstruction improves with increasing values of ν. As expected, we see this effect in Figure 5, where the reconstruction error decreases with increasing values of ν.

7 Concluding remarks

Reconstruction of functions from finitely many samples is a central problem in mathematics, which has be tackled by numerous disciplines including approximation theory, signal processing, data science and inverse

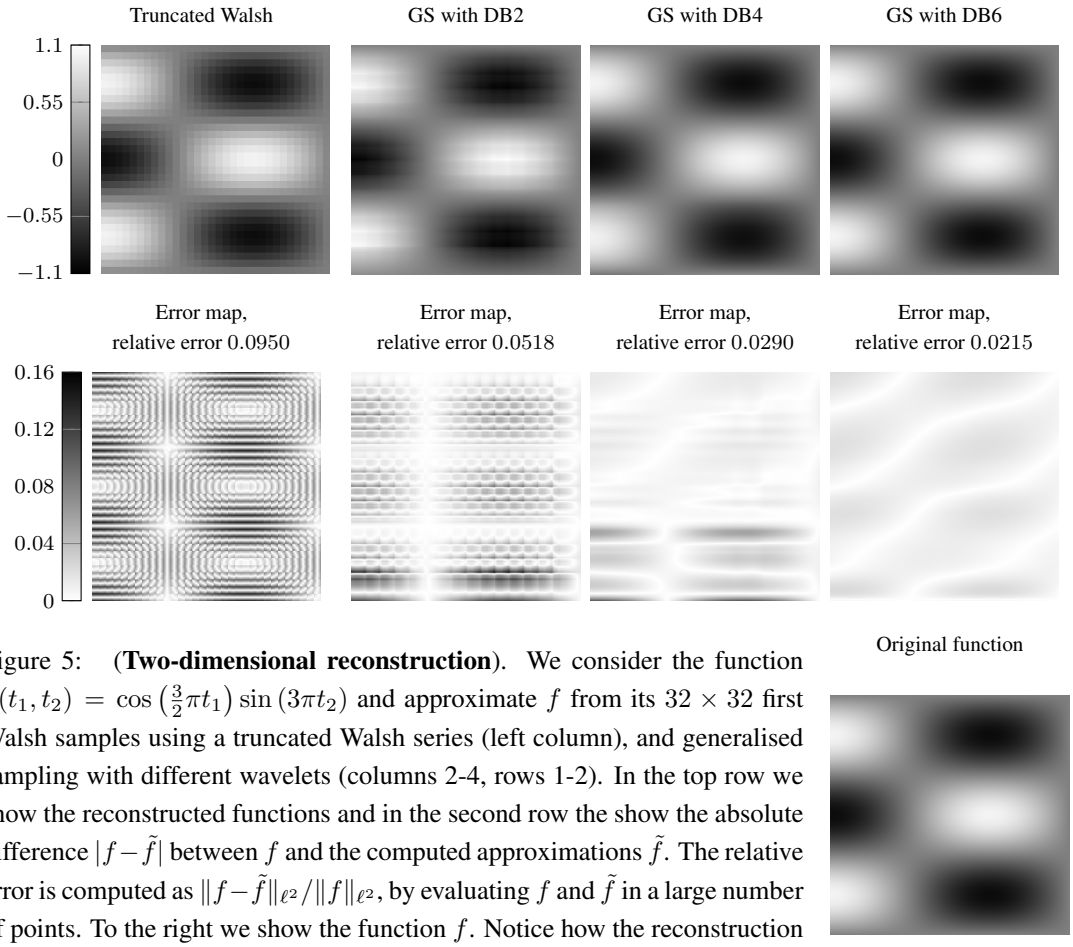

Figure 5: (**Two-dimensional reconstruction**). We consider the function $f(t_1, t_2) = \cos\left(\frac{3}{2}\pi t_1\right)\sin\left(3\pi t_2\right)$ and approximate f from its 32×32 first Walsh samples using a truncated Walsh series (left column), and generalised sampling with different wavelets (columns 2-4, rows 1-2). In the top row we show the reconstructed functions and in the second row the show the absolute difference $|f - \tilde{f}|$ between f and the computed approximations $\tilde{f}$. The relative error is computed as $\|f - \tilde{f}\|_{\ell^2} / \|f\|_{\ell^2}$, by evaluating f and $\tilde{f}$ in a large number of points. To the right we show the function f. Notice how the reconstruction error decreases with increasing values of ν.

problems. In this work, we have focused on the recovery of wavelet coefficients from binary measurements and derived an algorithm for computing the matrix-vector product with an $N \times M$ section of the change-of-basis matrix between a Walsh and a wavelet basis using $\mathcal{O}(N \log N)$ operations. As we have seen, this matrix is used by numerous classical reconstruction algorithms and can model many real-life sampling modalities. These days, a popular approach is to build reconstruction algorithms that approximate f using neural networks. While the practical performance of these algorithms is not yet mature [2], the inclusion of known operators into the network architectures can reduce the error bounds [49] of the learning algorithms. A promising new avenue is, therefore, to make further investigations into how known operators, such as the one considered in this work, can be utilized by modern machine learning. Future work could also consider whether it is possible to derive fast matrix-vector multiplications for other classical reconstruction bases, such as polynomials, splines, or bi-orthogonal wavelets etc. Such results would increase the use of these reconstruction bases when approximating functions from Walsh samples.

Acknowledgments

The author would like to thank the anonymous reviewers, Anders C. Hansen and Laura Thesing for their comments.

References

[1] B. Adcock, V. Antun, and A. C. Hansen. Uniform recovery in infinite-dimensional compressed sensing and applications to structured binary sampling. *Appl. Comput. Harmon. Anal.*, 55(7):1–40, 2021.

[2] B. Adcock and N. Dexter. The gap between theory and practice in function approximation with deep neural networks. *SIAM J. Math. Data Sci.*, 3(2):624–655, 2021.

[3] B. Adcock and A. C. Hansen. A generalized sampling theorem for stable reconstructions in arbitrary bases. *J. Fourier Anal. Appl.*, 18(4):685–716, 2012.

[4] B. Adcock and A. C. Hansen. Generalized sampling and infinite-dimensional compressed sensing. *Found. Comput. Math.*, 16(5):1263–1323, 2016.

[5] B. Adcock and A. C. Hansen. *Compressive imaging: Structure, Sampling, Learning*. Cambridge University Press, 2021.

[6] B. Adcock, A. C. 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.

[7] B. Adcock, A. C. 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.

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

[9] B. Adcock, A. C. Hansen, and B. Roman. A note on compressed sensing of structured sparse wavelet coefficients from subsampled Fourier measurements. *IEEE Signal Process. Lett.*, 23(5):732–736, 2016.

[10] B. Adcock, A. C. Hansen, and A. Shadrin. A stability barrier for reconstructions from Fourier samples. *SIAM J. Numer. Anal.*, 52(1):125–139, 2014.

[11] V. Antun and Ø. Ryan. On the unification of schemes and software for wavelets on the interval. *Acta Appl. Math.*, 173(7), 2021.

[12] J. Arndt. *Matters Computational: ideas, algorithms, source code*. Springer Science & Business Media, 2010.

[13] A. Bastounis and A. C. Hansen. On the absence of uniform recovery in many real-world applications of compressed sensing and the restricted isometry property and nullspace property in levels. *SIAM J. Imaging Sci.*, 10(1):335–371, 2017.

[14] K. G. Beauchamp. *Walsh functions and their applications*. Academic press, 1975.

[15] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk. Data assimilation in reduced modeling. *SIAM/ASA J. Uncertain. Quantif.*, 5(1):1–29, 2017.

[16] V. Boominathan, J. K. Adams, M. S. Asif, B. W. Avants, J. T. Robinson, R. G. Baraniuk, A. C. Sankaranarayanan, and A. Veeraraghavan. Lensless imaging: A computational renaissance. *IEEE Signal Proc. Mag.*, 33(5):23–35, 2016.

[17] J. Buckheit, S. Chen, D. L. Donoho, I. Johnstone, and J. Scargle. About WaveLab, 1995.

[18] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vision*, 40(1):120–145, 2011.

[19] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank. Sensitivity to basis mismatch in compressed sensing. *IEEE Trans. Signal Proces.*, 59(5):2182–2195, 2011.

[20] P. Clemente, V. Durán, E. Tajahuerce, P. Andrés, V. Climent, and J. Lancis. Compressive holography with a single-pixel detector. *Opt. Lett.*, 38(14):2524–2527, Jul 2013.

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

[22] I. Daubechies. *Ten lectures on wavelets*. SIAM, 1992.

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

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

[25] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk. Single-pixel imaging via compressive sampling. *IEEE Signal Proc. Mag.*, 25(2):83–91, 2008.

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

[27] Y. C. Eldar. Sampling without input constraints: Consistent reconstruction in arbitrary spaces. In *Sampling, wavelets, and tomography*, pages 33–60. Springer, 2004.

[28] Y. C. Eldar and T. Werther. General framework for consistent sampling in Hilbert spaces. *Int. J. Wavelets. Multi.*, 3(04):497–509, 2005.

[29] C. L. Epstein. *Introduction to the mathematics of medical imaging*. SIAM, 2007.

[30] 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, 2016.

[31] G. M. Gibson, S. D. Johnson, and M. J. Padgett. Single-pixel imaging 12 years on: a review. *Opt. Express*, 28(19):28190–28208, 2020.

[32] B. Golubov, A. Efimov, and V. Skvortsov. *Walsh series and transforms: theory and applications*, volume 64. Springer Science & Business Media, 1991.

[33] M. Guerquin-Kern, M. Haberlin, K. P. Pruessmann, and M. Unser. A fast wavelet-based reconstruction method for magnetic resonance imaging. *IEEE Trans. Med. Imaging*, 30(9):1649–1660, 2011.

[34] M. Guerquin-Kern, L. Lejeune, K. P. Pruessmann, and M. Unser. Realistic analytical phantoms for parallel magnetic resonance imaging. *IEEE Trans. Med. Imaging*, 31(3):626–636, 2012.

[35] A. C. Hansen and L. Thesing. On the stable sampling rate for binary measurements and wavelet reconstruction. *Appl. Comput. Harmon. Anal.*, 48(2):630–654, 2020.

[36] M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *J. Res. Natl. Bur. Stand.*, 49(6), 1952.

[37] A. Hirabayashi and M. Unser. Consistent sampling and signal recovery. *IEEE Trans. Signal Proces.*, 55(8):4104–4115, 2007.

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

[39] A. Jardine, H. Hedgeland, G. Alexandrowicz, W. Allison, and J. Ellis. Helium-3 spin-echo: Principles and application to dynamics at surfaces. *Prog. Surf. Sci.*, 84(11-12):323–379, 2009.

[40] A. Jones, A. Tamtögl, I. Calvo-Almazán, and A. Hansen. Continuous compressed sensing for surface dynamical processes with helium atom scattering. *Sci. rep.*, 6(1):1–11, 2016.

[41] G. Kutyniok and W.-Q. Lim. Optimal compressive imaging of Fourier data. *SIAM J. Imaging Sci.*, 11(1):507–546, 2018.

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

[43] Z.-P. Liang and P. C. Lauterbur. *Principles of magnetic resonance imaging: a signal processing perspective*. SPIE Optical Eng. Press, 2000.

[44] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly. Compressed sensing MRI. *IEEE Signal Proc. Mag.*, 25(2):72–82, 2008.

[45] J. Ma. Generalized sampling reconstruction from Fourier measurements using compactly supported shearlets. *Appl. Comput. Harmon. Anal.*, 42(2):294–318, 2017.

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

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

[48] 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. *Int. J. Numer. Meth. Eng.*, 102(5):933–965, 2015.

[49] A. K. Maier, C. Syben, B. Stimpel, T. Würfl, M. Hoffmann, F. Schebesch, W. Fu, L. Mill, L. Kling, and S. Christiansen. Learning with known operators reduces maximum error bounds. *Nat. Mach. Intell.*, 1(8):373–380, 2019.

[50] S. Mallat. *A wavelet tour of signal processing: The sparse way*. Academic Press, 3rd edition, 2008.

[51] A. Moshtaghpour, J. M. Bioucas-Dias, and L. Jacques. Close encounters of the binary kind: Signal reconstruction guarantees for compressive Hadamard sampling with Haar wavelet basis. *IEEE Trans. Inf. Theory*, 66(11):7253–7273, 2020.

[52] C. Poon. A consistent and stable approach to generalized sampling. *J. Fourier Anal. Appl.*, 20(5):985–1019, 2014.

[53] S. Ravishankar, J. C. Ye, and J. A. Fessler. Image reconstruction: From sparsity to data-adaptive methods and machine learning. *Proc. of the IEEE*, 108(1):86–109, 2019.

[54] simon foucart and holger rauhut. *A mathematical introduction to compressive sensing*. springer - birkäuser, 1th edition, 2013.

[55] V. Studer, J. Bobin, M. Chahid, H. S. Mousavi, E. Candes, and M. Dahan. Compressive fluorescence microscopy for biological and hyperspectral imaging. *Proc. Natl. Acad. Sci. USA*, 109(26):E1679–E1687, 2012.

[56] W.-S. Tang. Oblique projections, biorthogonal Riesz bases and multiwavelets in Hilbert spaces. *P. Amer. Math. Soc.*, 128(2):463–473, 2000.

[57] L. Thesing and A. C. Hansen. Linear reconstructions and the analysis of the stable sampling rate. *Sampling Theory in Signal and Image Processing*, 2018.

[58] L. Thesing and A. C. Hansen. Non-uniform recovery guarantees for binary measurements and infinite-dimensional compressed sensing. *J. Fourier Anal. Appl.*, 27(2):1–44, 2021.

[59] Y. Traonmilin and R. Gribonval. Stable recovery of low-dimensional cones in Hilbert spaces: One RIP to rule them all. *Appl. Comput. Harmon. Anal.*, 45(1):170–205, 2018.

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

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

[62] E. van den Berg and M. P. Friedlander. Probing the Pareto frontier for basis pursuit solutions. *SIAM J. Sci. Comput.*, 31(2):890–912, 2008.